IPv6 deployment in Tier2 site at FZU in Prague

Marek Eliáš, Lukáš Fiala, Jiří Chudoba, Tomáš Kouba
elias@fzu.cz

The International Symposium on Grids and Clouds (ISGC), February 26 - March 2, 2012

Institute of Physics AS CR, v. v. i. (FZU)
About Computing Centre at FZU

- Computing for solid state and high energy physics (HEP)
- Experiments: Atlas and Alice (CERN), D0 (FNAL)
- Participating in WLCG (Tier2 site), 4200 cores, 2PB of storage
- Only a computing centre — all computers centrally managed

Outline

1. Our current problems with IPv4
2. Hardware issues
3. IPv6 testbed and monitoring
4. Automatic installation
5. Interoperability with IPv4
6. GRID services and availability of GRID CRLs
Motivation

Lack of IPv4 addresses

- We have one /24 subnet, we currently use 3 times more addresses
- Workernodes are in a private address space

Routing problems

- DPM does not support multihoming (workernodes have to use the same address to access a DPM pool node as the outside world)
- Traffic between workernodes and DPM pool nodes must be routed through the central router
- New 10 Gbps infrastructure in FZU last year ⇒ routing has become unsustainable
Currently deployed solution

- Suggested by Maarten Litmaath
- DPM pool nodes are connected directly to both, private and public networks
- Workernodes from private 172.16/16 network have a static route to access public IP of each DPM pool node directly without routing
Firewall

- We are using a Cisco C6500 with FWSM in transparent mode
- FWSM does not support filtering of IPv6 traffic in transparent mode
- Possible solutions:
  - Switch whole FWSM to routed mode
  - Enable multiple context mode and filter IPv4 in transparent and IPv6 in routed mode
- Possible temporary workaround:
  - Use the ethertype ACL rules to have the FWSM pass all IPv6 traffic unfiltered
Cisco IOS vulnerability

- Security advisory published in September 2011\(^1\)
- An attacker can cause a router to reload by sending malformed IPv6 packets to the right interface of the router
- Nearly all versions of IOS were vulnerable
- Fixes are available from Cisco

Switches, management interfaces etc

- Our switches support switching IPv6 traffic, no performance issues noticed
- Only two of our switches can setup IPv6 connectivity on a management interface
- None of our servers can setup IPv6 on their management interface
- Maybe things are getting better:
  - tested IBM System x3550 M4 — pre-production sample
  - Supports manual IPv6 configuration, Autoconfiguration and stateful DHCPv6 on a management interface
  - Not able to setup a route if only DHCPv6 chosen
  - Web interface works fine in IPv6, remote console not tested
IPv6 testbed

Testbed mission

- Test our current tools and computing centre administration processes in IPv6 environment
- Find alternatives for IPv6 incompatible tools and processes
- We try to setup small IPv6 "computing site" with middleware services, workernodes etc
- Maybe an experimental way of accessing production services

Testbed parameters

- Two physical servers for running virtual machines
- Using infrastructure and connectivity of production network
- Several separate VLANs
- Running mostly Scientific Linux 6.1 (SL6.1)
Availability over both IPv4 and IPv6

- One of our authoritative nameservers has IPv6 connectivity
- Parent zones have an AAAA glue record of the IPv6 enabled nameserver
- We have one IPv6 only zone with a master without IPv4 connectivity
- IPv6 only resolver is not a good idea for now (not everybody trying IPv6 has IPv6 enabled nameservers)
- Also production services have resolvability problems, i.e. EPEL repositories of Fedora are not resolvable through IPv6²

²https://bugzilla.redhat.com/show_bug.cgi?id=758317
DHCPv6

What has been tested

- Autoconfiguration (SLAC) using RA from our Cisco 6500
- Stateless DHCP from Cisco 6500
- ISC DHCP and Dibbler
- worked fine

What to choose

- SLAC + stateless DHCP
  - SLAC do not advertise DNS resolvers → stateless DHCP is needed
  - Client can choose IPv6 address deterministically from MAC
  - Problems with failover or when NIC is replaced
- Stateful DHCP + RA
  - DHCPv6 do not advertise routes → RA is needed
  - Client is identified by DUID, not by MAC
DHCPv6

Deployed solution

- Stateful DHCP, ISC implementation version 4.1 from SL6.1
- DUID-LL (based on MAC address)
- Routes configured using RA

Problems and remarks

- dhclient uses LLT by default (includes time) → dhclient.conf should be setup before connecting to network
- ip6tables drop DHCPv6 responses by default in RHEL6\(^3\)
- Since 4.2 DHCP can match both LL and LLT using one hardware ethernet statement as in IPv4
- Route option in DHCPv6 maybe coming soon\(^4\)

---

\(^3\)https://bugzilla.redhat.com/show_bug.cgi?id=591630
\(^4\)http://tools.ietf.org/html/draft-dec-dhcpv6-route-option-02
Nagios

- Smooth installation, standard checks work fine
- Problem with SNMP checks — `snmpget` needs IPv6 host name/address written in a special manner

MRTG

- Smooth installation, no problems with getting SNMP data through IPv6

syslog

- OpenBSD 5.0 and syslog-ng 3.1.4
- No problems encountered

netflow

- We use `flow-tools` in production IPv4 environment
- `nfdump` seems to be a better alternative in IPv6
PXE in IPv6

- Changed since IPv4
- Described in RFC 5970 from September 2010
- No `next-server` option in DHCPv6, server directly specifies url of image to be loaded through option `boot-file-url`
- RFC 5970 is missing on dibbler’s list of implemented RFCs; ISC DHCP 4.2.2 doesn’t seem to support it

Support of PXE through IPv6 in hardware

- In our current hardware there is `none`
- Opensource implementation gPXE tested on an old hardware
  - Some NICs were burned out
  - Premature IPv6 support, but SLAC seems to work fine
  - No support of RFC 5970 → no automatic installation
Working solution

- Unrouted private IPv4 network inside an IPv6 VLAN
- Whole installation proceeds via IPv4
- After the installation the IPv4 setup is discarded and IPv6 is introduced

Required services

- Both DHCPv4 and DHCPv6 server
- IPv4 PXE install server
- HTTP proxy connected also to IPv4 Internet
- DNS (optional, installer connects only to proxy)
- Puppet, if configuration management required (works in IPv6)
Question:
I am now in IPv4-only network, how do I access our IPv6 testbed?

6to4

▶ No control of data flow (performance/security/stability issues)
▶ We did not deploy own 6to4 gateway

6in4

▶ Point to point tunnel
▶ Any dual stack node can be setup as a gateway
▶ Works well inside IPSec/VPN tunnel
▶ Tunnel interfaces are created on both ends:
  host-A:~# /sbin/ip tunnel add sit1 mode sit ttl <ttldefault> \ remote <host-B-ipv4-addr> local <host-A-ipv4-addr>
▶ Tunnel interfaces are then configured with IPv6 addresses
Question 2:
I am in IPv6-only network, how do I access IPv4 network?

ipv6_tunnel

- Point to point tunnel
- Tunnel interfaces are created on both ends:
  ```
  host-A:~# /sbin/ip -6 tunnel add mytun mode ipip6 remote \
  remote <host-B-ipv6-addr> local <host-A-ipv6-addr>
  ```
- Tunnel interfaces are then configured with IPv4 addresses
- Can be used to interconnect (possibly private) IPv4 networks through IPv6

SSH tunnel

- Client connects through IPv6 with `ssh -w` to a remote gateway
- SSH creates tunnel interfaces on both ends
- These interfaces can be then configured with IPv4 addresses
Our experiments with middleware deployment are only at the beginning.

**gLite user interface**
- Running SL5.7
- Not a full featured UI
- Only a gridFTP client and server
- Participating in HEPiX IPv6 testbed

**Certificate revocation lists**
- Most components of gLite need access to CRLs
- Our tool\(^5\) checks availability of CRLs from lcg-CA bundle
- 13 of 96 CRLs downloadable through IPv6 (6 were in October)
- Should be improved before any serious middleware deployment in IPv6 environments

\(^5\)http://www.particle.cz/farm/admin/IPv6EuGridPMACrlChecker/
## Availability of Grid CRLs

### IPv6 EUGridPMACrlChecker

Check date: 19.10.2011  [*text report (take a while)*]

<table>
<thead>
<tr>
<th>CA Subject</th>
<th>IPv6 DNS</th>
<th>IPv6 GET</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN=CERN Trusted Certification Authority, DC=cern, DC=ch</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://ca.cern.ch/ca/CRL/CERN">http://ca.cern.ch/ca/CRL/CERN</a> Trusted Certification Authority.crl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CESNET CA 3, O=CESNET CA, DC=cesnet-ca, DC=cz</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td><a href="http://crl.cesnet-ca.cz/CESNET_CA_3.crl">http://crl.cesnet-ca.cz/CESNET_CA_3.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CESNET CA Root, O=CESNET CA, DC=cesnet-ca, DC=cz</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td><a href="http://crl.cesnet-ca.cz/CESNET_CA_ROOT.crl">http://crl.cesnet-ca.cz/CESNET_CA_ROOT.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CESNET CA, DC=cesnet-ca, DC=cz</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td><a href="http://www.cesnet.cz/pki/crl/cn/CESNET">http://www.cesnet.cz/pki/crl/cn/CESNET</a> CA,dc=cesnet-ca,dc=cz.crl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=GRID2-FR, O=CNRS, C=FR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://crls.services.cnrs.fr/GRID2-FR/getpem.crl">http://crls.services.cnrs.fr/GRID2-FR/getpem.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CNRS2-Projects, O=CNRS, C=FR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://crls.services.cnrs.fr/CNRS2-Projects/getpem.crl">http://crls.services.cnrs.fr/CNRS2-Projects/getpem.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CNRS2, O=CNRS, C=FR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://crls.services.cnrs.fr/CNRS2/getpem.crl">http://crls.services.cnrs.fr/CNRS2/getpem.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=CyGridCA, O=HPCL, O=CyGrid, C=CY</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://cygrid.org.cy/CyGridCA/afe55e6e6.r0">http://cygrid.org.cy/CyGridCA/afe55e6e6.r0</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=Grid-Ireland Certification Authority, O=Grid-Ireland, C=IE</td>
<td>YES</td>
<td>ERR</td>
</tr>
<tr>
<td><a href="http://www.cs.tcd.ie/Grid-Ireland/gi-ca/1e43b9cc.crl">http://www.cs.tcd.ie/Grid-Ireland/gi-ca/1e43b9cc.crl</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=Grid Canada Certificate Authority, O=Grid, C=CA</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.gridcanada.ca/ca/bffbd7d0.r0">http://www.gridcanada.ca/ca/bffbd7d0.r0</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=HellasGrid CA 2006, OU=Certification Authorities, O=HellasGrid, C=GR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://crl.grid.auth.gr/hellasgrid-ca-2006/62b36fca.pem">http://crl.grid.auth.gr/hellasgrid-ca-2006/62b36fca.pem</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN=HellasGrid Root CA 2006, OU=Certification Authorities, O=HellasGrid, C=GR</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><a href="http://crl.grid.auth.gr/hellasgrid-root-ca-2006/28a58577.pem">http://crl.grid.auth.gr/hellasgrid-root-ca-2006/28a58577.pem</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There is a long way before us...
Standards still changing or completing
There are bugs and security issues in software, routers etc
There is a long way before us . . .
Standards still changing or completing
There are bugs and security issues in software, routers etc
Migration to IPv6 is necessary
Standards / software / hardware won’t be fixed unless somebody use them
and . . .
Conclusion

- There is a long way before us ...
- Standards still changing or completing
- There are bugs and security issues in software, routers etc
- Migration to IPv6 is necessary
- Standards / software / hardware won’t be fixed unless somebody use them
- and ...

**IPv6 itself is usable now**

- at least in a testbed,
Conclusion

- There is a long way before us . . .
- Standards still changing or completing
- There are bugs and security issues in software, routers etc
- Migration to IPv6 is necessary
- Standards / software / hardware won’t be fixed unless somebody use them
- and . . .

IPv6 itself is usable now

- at least in a testbed, but also in ChinaGrid
- but many applications are still not IPv6-ready
Thank You

Marek Eliáš

elias@fzu.cz

http://www.farm.particle.cz

Work partially supported by CESNET, z. s. p. o.
project number 416R1/2011
### IPv6 support in networking hardware:

<table>
<thead>
<tr>
<th>Hardware name</th>
<th>switching</th>
<th>Mgmt</th>
<th>SNMPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco Catalyst C6500</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SMC TigerStack II 10/100/1000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HP ProCurve J4904A Switch 2848</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HP ProLiant BL p-Class C-GbE2</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HP GbE2c Switch c-Class Blade</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BNT RackSwitch G8000</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Force10 S2410-01-10GE-24P</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BNT RackSwitch G8124</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
**IPv6 support in server hardware:**

<table>
<thead>
<tr>
<th>Hardware name</th>
<th>Mgmt</th>
<th>PXE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP BL 35p</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HP BL 460c</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HP DL360 G3 – G6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>IBM x3650 M2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>IBM iDataPlex dx340</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>IBM iDataPlex dx360 M2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>IBM iDataPlex dx360 M3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SGI Altix XE 310</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SGI Altix XE 340</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SGI C1001-G13</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Supermicro X8DTU</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>