FZU IPv6 testbed updates

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

HEPiX Fall 2012 Workshop, 15–19 October, 2012

Institute of Physics AS CR, v. v. i. (FZU)

Outline:

1. IPv6 address configuration: SLAAC vs. DHCPv6
2. Monitoring of HEPiX IPv6 testbed: Nagios and Smokeping
3. Network installation in IPv6
4. Hardware support survey
5. Middleware tests

Address configuration again

Fyzikální ústav
Akademie věd ČR, v. v. i.

Scenario and its general requirements

- Address configuration on production servers and workernodes
- Every server need a fixed IPv6 address
- We want small configuration overload in case of:
  - Change NIC in a server machine
  - Reinstallation of a server machine OS

Stateless address autoconfiguration (SLAAC)

- Router sends a prefix and a next-hop address in a routing advertisement (RA)
- Clients use this information to setup IP and routing
- IPv6 address chosen from prefix Deterministically (MAC based) or undeterministically
- Routing setup according to next-hop option in RA

Problems with SLAAC

- NIC is replaced → different MAC → different IPv6 address
- Change in DNS to reflect IPv6 address change → long TTL
- Domain name still cannot be used sometimes:
  - Firewall ACLs
  - What about a single node exceptions in external Firewall?
  - small devices using NTP or syslog logging without DNS
- With IPv4 you need only to update the MAC in dhcpd.conf
- With IPv6 and SLAAC you need to consider all mentioned cases

How does it work

- Client is identified by DUID
  - DUID-LL derived only from MAC address
  - DUID-LLT contains time of its generation
    → can't be predicted e.g. when installing a machine
- DUID is stored in a lease file (somewhere in /var/lib)
- If a lease file exists, dhclient always uses DUID stored there
- In SL6, separate instances of dhclient per interface
  → separate lease files per interface
  → different DUIDs per interface

Replacing NICs

- RFC[1] says: "a device's DUID should not change as a result of a change in the device's network hardware."
- This is also the case in the SL6 implementation (Unfortunately?)
- When a NIC is replaced, old DUID-LL is still stored in /var/lib
- . . . until reinstallation of the machine
- When you reinstall whole rack of workernodes, machines with replaced NICs wont get an IPv6 address

[1] RFC 3315: Dynamic Host Configuration Protocol for IPv6 (DHCPv6)

Problems with DHCPv6

Multiple NICs

- RFC[2] says: "DHCP client and server has exactly one DUID."
- In SL6 this is not the case
- DUIDs are different even when running a single instance of dhclient with a single lease file (dhclient 4.1.1-P1)
- But what DUID-LL should a system with multiple NICs use?
- DUID does not identify the whole system; same case in IPv4

[2] RFC 3315: Dynamic Host Configuration Protocol for IPv6 (DHCPv6)

What to do now?

- Choose one and accept the pain

Choice at FZU:

- DHCPv6
- When replacing a NIC one need to delete the lease file
  ... and update dhcpd.conf as in IPv4

Nagios

What do we monitor:

- FZU IPv6 testbed: mostly IPv6-only services
- HEPiX IPv6 testbed: only dual-stack services

Nagios Checks in IPv6

- General checks work fine (PING, SSH, LOAD etc) work fine
- SNMPv6 checks: usually use snmpwalk and snmpget
  These commands need IPv6 addresses in a special form:
  `snmpget "ipv6:[fec0::dead:beef]"`
  → custom check command is needed
- NRPE and NSCA not tested yet

Nagios

Special checks for HEPiX IPv6 testbed

- DNS check: checks resolvability by IPv6-only resolver
- GridFTP upload and download

Monitoring of dual-stack services

- Dual-stack services need to be accessible from both IPv4-only and IPv6-only hosts
- Nagios on dual-stack: very difficult to ensure that a check uses only IPv4 or only IPv6, check command can connect implicitly
  - to another service specified by config file
  - to another service specified by environment
  - to another service returned by protocol

  and we probably can't force the IP version to be used on nagios side
  $\rightarrow$ IPv6-only and IPv4-only checking instances

Nagios

- Klinec & Elwell (CERN): nagios probes for dual-stack testing[3]
- We need results to be accessible through both IPv4 and IPv6

Deployed solution

- Three instances of Nagios
- IPv6-only and IPv4-only checking instances with livestatus
- checking instances use dual-stack DNS resolvers;
  IPv4/IPv6-only resolvers are used for DNS check
- Dual-stack agregating instance[4] with check_mk multisite
- No dual-stack checking instance: Do we need one?
- check_mk multisite can not connect to livestatus through IPv6
  $\rightarrow$ a special hack with xinetd and netcat deployed

[3] Presented on F2F meeting of HEPiX IPv6 WG, 4–5 Oct 2012

[4] `http://monitor.ipv6.farm.particle.cz/check_mk/`
(Ask me for the password)

Smokeping

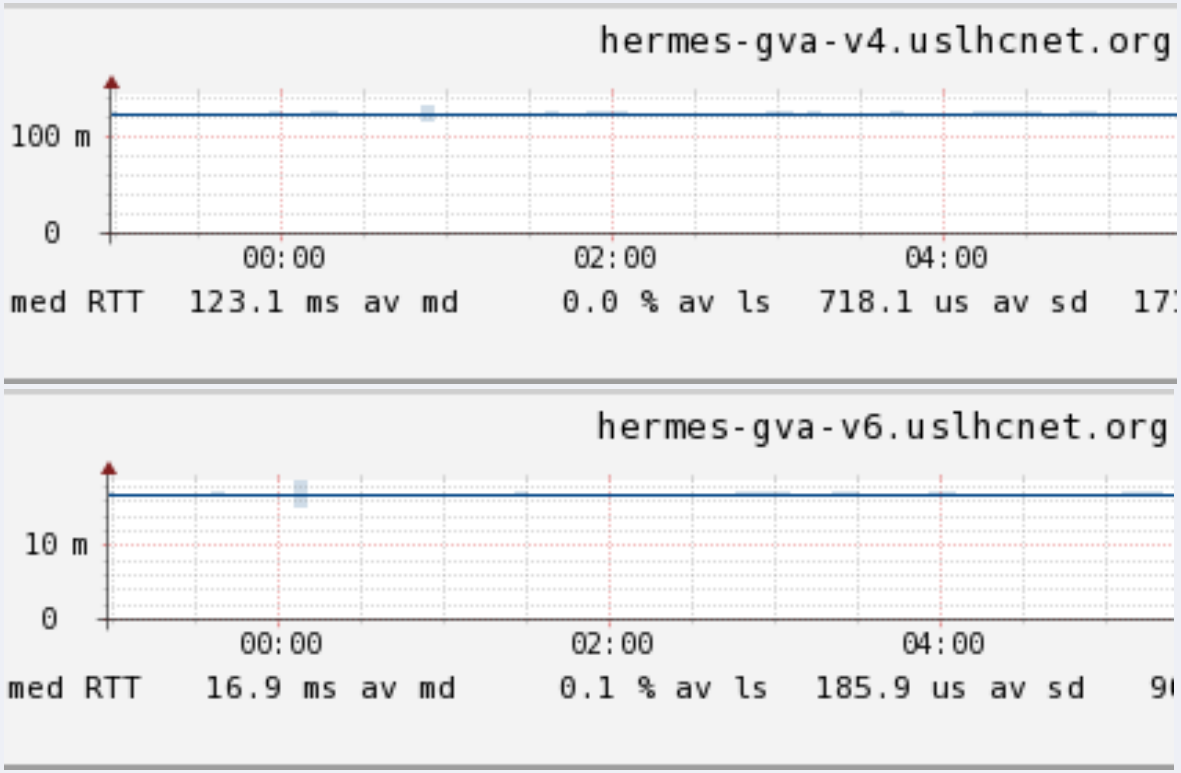
- Latency monitoring between FZU and nodes of HEPiX IPv6 testbed
- IPv6 equivalents of some utils should be used (fping6 instead fping etc)
- No IPv6 related configuration issues
- Roundtrip measurement
- GridFTP upload and download latency

`http://monitor.ipv6.farm.particle.cz/smokeping/sm.cgi`

Fyzikální ústav
Akademie věd ČR, v. v. i.

Results

- RTT in IPv6 is similar to IPv4 and sometimes even better
- Latency of GridFTP is very similar in IPv6 and IPv4
- Sometimes very different results (probably different routes)



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 $\rightarrow$ no automatic installation

Network 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)

IPv6 support in hardware

Fyzikální ústav
Akademie věd ČR, v. v. i.

IPv6 support in server hardware at FZU:

Hardware name	Mgmt	PXE
HP BL 35p	No	No
HP BL 460c	No	No
HP DL360 G3 – G6	No	No
IBM x3650 M2	No	No
IBM iDataPlex dx340	No	No
IBM iDataPlex dx360 M2	No	No
IBM iDataPlex dx360 M3	No	No
SGI Altix XE 310	No	No
SGI Altix XE 340	No	No
SGI C1001-G13	No	No
Supermicro X8DTU	No	No
IBM System x3550 M4[5]	Yes	No

[5] Preproduction sample, license for a remote console was not included

IPv6 support in hardware

IPv6 support in networking hardware at FZU:

Hardware name	switching	Mgmt	SNMPv6
Cisco Catalyst C6500	Yes	Yes	Yes
SMC TigerStack II 10/100/1000	Yes	Yes	Yes
HP ProCurve J4904A Switch 2848	Yes	No	No
HP ProLiant BL p-Class C-GbE2	Yes	No	No
HP GbE2c Switch c-Class Blade	Yes	No	No
BNT RackSwitch G8000	Yes	No	No
Force10 S2410-01-10GE-24P	Yes	No	No
BNT RackSwitch G8124	Yes	No	No

Middleware

What works:

- GridFTP, EMI UI, DPM, LB

Batchsystem

- Torque was tested
- Torque from SVN (version 4.1.3) can't communicate with workernodes through IPv6
- IPv6 development branch (based on 2.5.0): compilation problems

Problems

- YAIM relies on `hostname -f`, which is broken in SL5
  $\rightarrow$ wrapper script deployed
- CRLs of 21 out of 94 CAs from lcg-CA bundle are accessible through IPv6 (only 6 were by Vancouver meeting)
  `http://www.particle.cz/farm/admin/IPv6EuGridPMACrlChecker/`

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