

Enabling IPv6 at FZU - WLCG Tier2 in Prague

Tomáš Kouba, Jiří Chudoba, Marek Eliáš

Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic

Introduction

The Tier2 Computing Center in Prague (aka FZU) is preparing for the new generation of Internet Protocol (IPv6) by testing crucial parts of the computing infrastructure (ranging from network setup and hardware management to software setup and configuration) on a dedicated IPv6 testbed. This testbed is used for mimicking production infrastructure of Prague Tier2. The fundamental parts are presented on this poster: network setup, monitoring, node installation and batch system. Our testbed also participates in HEPiX IPv6 working group testbed.

SLAAC problems

Stateless address autoconfiguration (SLAAC) uses MAC address as a source for computing IP address. But we want every server to have fixed IP address even after OS reinstallation or NIC replacement. Why?

- ▶ Special devices use NTP or syslog protocol but they do not support DNS
- ▶ Firewall ACLs do not support domain names
- ▶ Single node exceptions in *external* firewall
- ▶ TTL issues when configuration of DNS is changed every time NIC is replaced

DHCPv6 issues

DHCPv6 works in a more controlled way:

- ▶ Client is identified by DUID
 - ▶ DUID-LL derived only from MAC address
 - ▶ DUID-LLT contains time of its generation
- ▶ DUID is stored in a lease file (somewhere in `/var/lib`)
- ▶ If a lease file exists, dhclient always uses DUID stored there
- ▶ In SL6 there are separate instances of dhclient per each interface
- ▶ When a NIC is replaced, old DUID-LL is still used
 - ▶ Until reinstallation of the machine
 - ▶ This means machines with changed NICs get inaccessible after reinstallation
 - ▶ Solution at FZU: delete the lease file when replacing NIC (and update DHCPv6 configuration)
- ▶ Another problem with SL6 implementation is not following RFC 3315:
 - ▶ DHCP client and server has exactly one DUID
 - ▶ DUIDs are different per each NIC

Monitoring - Nagios and Smokeping

We run three instances of Nagios (with `livestatus+check_mk` extension) on three machines:

- ▶ IPv6-only (livestatus extension needs a workaround with `xinetd` to work over IPv6 only connections)
- ▶ IPv4-only
- ▶ dual-stack (aggregating instance)

All the instances monitor FZU IPv6 testbed and also the HEPiX IPv6 testbed. The basic connectivity tests (PING, SSH) work fine over IPv6. SNMPv6 checks using `snmpwalk` and `snmpget` need a special format of the address `snmpget "ipv6:[fec0::aabb:cdd]"` and so it is necessary to define custom check command in Nagios.

Network installation tests in IPv6

Nodes installation over network does not work very well in IPv6-only networks now:

- ▶ PXE in IPv6 is described in RFC 5970
 - ▶ No next-server option in DHCPv6, image is directly specified by `boot-file-url` option
 - ▶ Neither of tested DHCPv6 implementations (dibbler and ISC DHCP 4.2.2) support RFC 5970
- ▶ No hardware at FZU supports PXE over IPv6
- ▶ Opensource implementation gPXE tested on an old hardware
 - ▶ Some NICs were bricked during testing several versions
 - ▶ gPXE was very unstable, but we were able to boot an installation image few times

The working solution at FZU IPv6 testbed for installation of nodes looks like this:

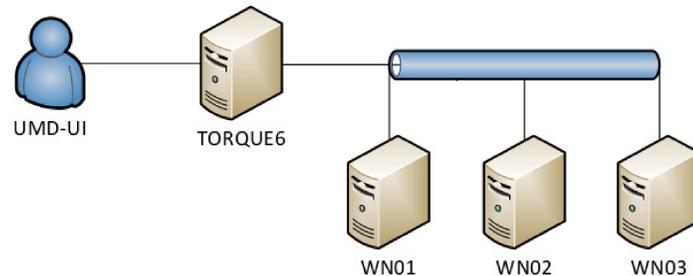
- ▶ Unrouted private IPv4 network inside an IPv6 VLAN
- ▶ Whole installation proceeds via IPv4
- ▶ After the installation the IPv4 setup is discarded and IPv6 is introduced

Such configuration required the following services: DHCPv4 and DHCPv6 server, IPv4 PXE install server, HTTP proxy connected also to IPv4 Internet, Puppet (works fine with IPv6).

Batch system - Torque

One of the critical services of FZU site is the batch system. We use Torque resource manager with Maui scheduler. The following schema shows a typical workflow of a user job:

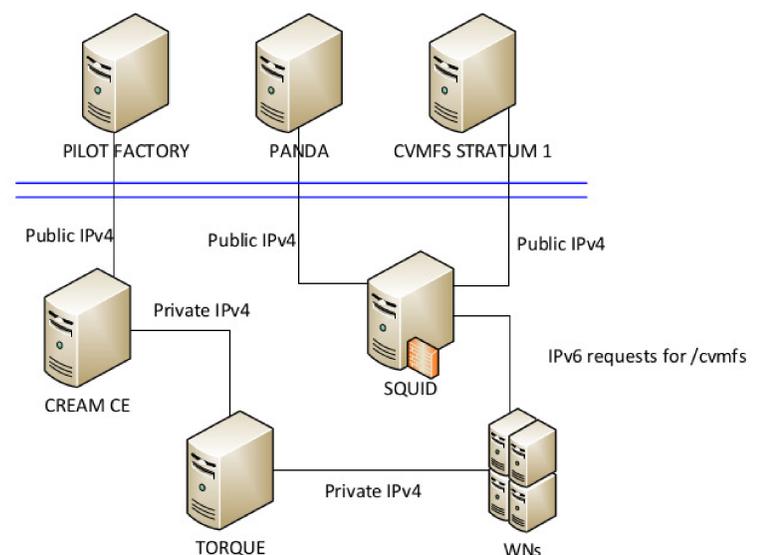
User job



We have successfully tested submitting a job as a user from user interface to torque system. The job was executed and input and output files were transferred. The main issues encountered were:

- ▶ Daemon `trqauthd` used for authentication does not support IPv6
- ▶ Torque server listens only on IPv4 addresses
- ▶ Mom daemons require to communicate via IPv4 (we had to configure the system to return IPv4 address of torque server first)
- ▶ File transfers (job input+output) use scp over IPv6

ATLAS grid job



- ▶ We have installed CVMFS on worker nodes. The SQUID proxy used for CVMFS is the same instance we use for packages installation.
- ▶ The CREAM CE has been configured:
 - ▶ With public IPv4 address for submission from external world
 - ▶ With private IPv4 address for communication with torque
 - ▶ IPv6 address was configured for other protocols

Myproxy

Myproxy was recently fixed to support IPv6 protocol in version 5.8. We have installed the service from EMI3 on a dual-stack machine `myproxy.ipv6.farm.particle.cz` and tested proxy registration and renewal from IPv6-only client.

Conclusion and Future work

We plan to test full job execution chain for the two WLCG experiments supported at FZU (ATLAS and ALICE). The building blocks are tested and work fine (CREAM CE gateway, batch jobs, central services access via HTTP proxy). The next step is to put our testbed in the WLCG infrastructure and run real jobs.

Acknowledgement



This work is partially supported by CESNET project 482/2013 MYSE project LG13009.