

CESNET Technical Report 5/2008

IPv6 in METACenter

DAVID ANTOŠ, JIŘÍ SITERA, DANIEL KOUŘIL

Received 6.5.2008

Abstract

The report describes IPv6 address allocation architecture and policy in METACenter. Status of IPv6 support in applications (Grid middleware, network filesystems) is discussed.

Keywords: IPv6, METACenter, addressing scheme, IPv6 application support

1 Introduction

With the growth of number of network nodes and provided services, the need of unique IP addresses increases. Techniques like Network Address Translation disable full end-to-end addressing, natural routing, and introduce throughput problems without offering a compensating added value. The way to get sufficient address space and restore full end-to-end addressing without obstacles is the IPv6 protocol.

One of the main reasons for deploying IPv6 in METACenter is growing usage of virtual machines. The number of virtual machines running on a single physical node is theoretically unlimited and a-priori unpredictable. In such an infrastructure, physical computation nodes as well as virtual machines running on them have distinct IP addresses. The IPv4 address space is tight even for physical nodes themselves. Moreover, it may also be interesting to use IP addresses to uniquely identify virtual machine images, not just running virtual machines, allocating large number of addresses (semi-)permanently.

This report covers the initial steps taken to enable IPv6 in METACenter. We describe the addressing and naming plan as well as necessary link layer infrastructure supporting the chosen way of addressing. We study IPv6 readiness of middleware used in METACenter, summarise experience gained from testing, and finally give a list of services currently available over IPv6.

2 Addressing Architecture

In this section, we describe METACenter IPv6 addressing and DNS naming architecture.

2.1 Preamble

The main problem of current IPv4 METACenter infrastructure is lack of uniqueness of addressing and domain name assignment: local clusters use local addressing and naming policies. Because of tight IPv4 address space, no practical options are available to solve this. Current IPv4 infrastructure will be therefore left intact, it

will be used during the transition phase to IPv6. No end-of-service time is set for the IPv4 infrastructure, it is supposed to be supported as-is for legacy applications for a long time. The IPv4 network is not further discussed in this section.

METACenter addressing infrastructure is planned completely independent on current IPv4 addressing, IPv6 addressing of local sites and names of the machines in local sites of METACenter partners.

2.2 Addressing

All METACenter sites share a common prefix `2001:718:ff01::/48`. Link layer support for interconnecting flat addressing infrastructure will be described in Section 3.

IPv6 addressing in METACenter is not strictly internally structured, the structure introduced serves just making the administration easier. From routing point of view, the whole `/48` prefix is a single virtual local network in order to allow virtual machines to migrate freely among physical hosts. The METACenter prefix is divided into `/56` networks for physical and virtual machines.

Physical nodes are assigned the `2001:718:ff01::/56` prefix. The prefix for physical nodes is internally divided into three `/64` prefixes based on the location of the cluster clouds (see Table 1). This serves only for ease of administration, the network of physical machines is routed as a whole based on the `/56` prefix.

Table 1. Prefix allocation for physical machines

Brno	<code>2001:718:ff01:1::/64</code>
Plzeň	<code>2001:718:ff01:2::/64</code>
Praha	<code>2001:718:ff01:3::/64</code>

The addressing plan (including tools for DNS record generation) is kept in CVS~[1].

2.3 DNS

DNS AAAA records for physical machines are created in `meta6.cesnet.cz` name space. There are three reasons for providing a separate name space for IPv6:

1. It unifies naming of machines. The IPv4 names used so far belong to name spaces of hosting organisations.
2. If a machine runs a service on IPv4 only, trying to access the service on IPv6 first could cause an unpleasant timeout before IPv4 access takes place. Moreover, despite the fact that trying all addresses obtained from DNS is the recommended behaviour, many clients do not conform and do not attempt to use more than one received address.
3. Assigning names to the `meta6.cesnet.cz` space does not require cooperation with local administrators of hosting organisations. Besides the fact that distributing the administration requires complex communication even for simple tasks and corrections, holding the records in local nameservers often requires major upgrades of nameserver software which is not often accepted by local administrators who correctly understand DNS as a critical service.

Names of the machines in the `meta6` name space are identical to currently used host names in IPv4, for example `skirit58.ics.muni.cz` will be available over IPv6 as `skirit58.meta6.muni.cz`.

2.4 Local Addressing and Naming Policies

This addressing scheme and name assignment is independent on IPv6 addressing and naming in the organisation physically hosting the machines. For example, the `skirit` cluster is available also under names `skirit<n>.ipv6.ics.muni.cz` with the addresses of the Institute of Computer Science. METACenter does not define local addressing and naming policies nor requires the hosts to be available over local IPv6 addresses.

For internal METACenter operation, names from the `meta6.cesnet.cz` name space are generally preferred (e.g., for Kerberos, issuing authentication tickets, etc.), especially for cluster machines.

METACenter hosts can be connected to a network that is either technically or economically difficult to integrate to the uniform addressing infrastructure (this may be common for single hosts providing special services, e.g., LDAP server). In that case, local addresses of the machines may be also used for METACenter services. Such machines may obtain METACenter domain names even for “non-METACenter” addresses, but this is not strictly required. Cases like this have to be judged individually.

3 Underlying L2 Network Architecture

All the cluster hosts are connected into a single virtual LAN. The VLAN is propagated over the academic backbone using Virtual Private LAN Service (VPLS) over IP/MPLS and/or over a dedicated line using Xponder cards as described in [5].

Using a single LAN is mainly advantageous for allowing virtual machines to migrate among physical hosts. On a single LAN, the network connections can be kept as the address of the migrated machine doesn't change.

4 IPv6 Readiness of Applications

IPv6 support in applications requires using larger data structures for addresses, moreover, additional information is attached to addresses, for example validity range and/or interface in case of link local addresses. Applications must take lists of addresses returned by DNS into account. The outgoing interface can have several addresses to choose, too. It is also necessary to parse IPv6 addresses from configuration files.

If the application uses library functions and structures to manipulate IP addresses, the modifications are likely to be small. Otherwise, the effort for porting the application is difficult to estimate. Using hard-coded IPv4 addresses in the code and misusing the addresses as internal application layer identifiers (as we can find, e.g., in AFS code) can enforce complete re-programming from scratch.

In general, mainstream widely used Internet servers have been supporting IPv6

for several years. Unfortunately, METACenter uses also not-so-widely deployed middleware, therefore we tested servers and applications to ensure that IPv6 support is usable.

Even though some applications support IPv6 quite well from the “functional point of view,” we encounter minor issues mostly in configuration files and command line syntax. It is quite common that some applications do not accept IPv6 addresses in configuration files, do not support IPv6 addresses written in square brackets or the link local interface (with %) syntax. It does not necessarily mean that applications do not support IPv6, it is sufficient to use domain names instead.

4.1 Kerberos

Kerberos is the authentication mechanisms employed to ensure security of METACenter. Heimdal~[2] versions 0.6.3 and 0.7 and MIT Kerberos¹ version 1.5.3 were tested. All of them were used as clients and servers. In all combinations, we were able to use kinit over IPv6 to obtain Kerberos tickets using users’ passwords. We have encountered only one minor issue with the Heimdal libraries not interpreting correctly the IPv6 format of address. Addressing the problems would be quite easy when necessary, moreover, configuration files strictly use domain names anyway.

An IPv6 enabled key distribution centre (containing data identical to other METACenter KDCs) runs at `kdc1.ipv6.ics.muni.cz`.

4.2 SSH

OpenSSH² handles IPv6 since version 2.7 without problems (including full address syntax).

METACenter machines are available via ssh under domain names corresponding to IPv6 addresses (see Section 2.3). We also verified that Kerberos authentication can be established between IPv6 enabled client and server. These tests were done on our internal testbed where machines are provided only IPv6 addresses. However, we experienced problems with Kerberos authentication in ssh, when we moved to the METACenter production environment where machines are configured with both IPv4 and IPv6 that map to different host names. Such a configuration makes the ssh server unable to find the correct Kerberos key to verify client’s messages. When the machines are not provided with both types of IP addresses, the authentication goes on smoothly. At the moment, we are considering possible scenarios and evaluating if dual-addresses configurations are needed and how to possibly solve the issue.

4.3 Apache Web Server

Apache³ web server supports IPv6 since version 2.0. METACenter web portal can be accessed via IPv6 home page⁴.

¹ <http://web.mit.edu/kerberos/www/>

² <http://www.openssh.com/>

³ <http://www.apache.org/>

⁴ <http://www.meta6.cesnet.cz>

4.4 Network Filesystems

Major AFS server OpenAFS⁵ in current version 1.4.5 does not support IPv6. The roadmap document⁶ states that IPv6 support is planned in 20 to 25 months. During the last year, the expected time has been postponed several times. OpenAFS code uses IPv4 addresses as internal node identifiers, therefore the changes to the code are likely to be very difficult. Obviously, another network file system will be necessary to consider.

Parallel Virtual File System⁷ (PVFS) in version 2.7 doesn't support IPv6. Nevertheless the system is well coded and we expect changes to be small.

Although NFSv4⁸ is reported to work with IPv6 with patched 2.6.27 Linux kernel, the documentation is so incomplete we did not succeed configuring it in a reasonable amount of time.

4.5 DHCPv6

We supposed the physical machines to have fixed addresses assigned by DHCPv6. Moreover, pools of addresses shall be available for testing and installing machines before they are integrated to the production cluster infrastructure. We therefore require a DHCPv6 implementation to be able to assign addresses based on DHCPv6 Unique Identifier (DUID) generated out of MAC addresses. Although it is recommended to use MAC address with timestamp, this DUID type is not very suitable for cluster address assignment because the value of the DUID is not known until it is generated on the client. On the contrary, MAC addresses are very easy to obtain.

Three open-source DHCPv6 implementations exist: dhcpv6, wide-dhcpv6, and Dnsmasq. Abilities of all the packages are quite limited.

Dnsmasq⁹ does not support other DUID types than MAC address with timestamp. Although the documentation describes assigning addresses per DUID, it does not work.

Wide-dhcpv6¹⁰ is declared not to support temporary addresses (which violates the RFCs). Setting addresses based on DUIDs works, the only DUID type supported by the client is MAC address with timestamp. It is possible to use a client supporting other DUID types. It is not possible to configure several address pools per interface. The client does not really set the addresses on the interface.

Dnsmasq¹¹ supports all DUID types. It is not possible to configure addresses per DUID and a pool of addresses at the same time. Configuring Dnsmasq is quite complex, addresses are configured in separate classes which encapsulate special parameter settings and address pools. They allow also accept and reject rules for DUIDs, reject constructs seem to be ignored by the software.

⁵ <http://www.openafs.org/>

⁶ <http://www.openafs.org/roadmap.html>

⁷ <http://www.pvfs.org/>

⁸ <http://www.nfsv4.org/>

⁹ <http://dhcpv6.sourceforge.net/>

¹⁰ <http://sourceforge.net/projects/wide-dhcpv6/>

¹¹ <http://klub.com.pl/dhcpv6/>

To sum up, there is no generally production quality open source “out of the box” DHCPv6 implementation. In order to ensure stable address assignment, we configured the addresses of the physical hosts manually.

We have nevertheless patched wide-dhcpv6 server to support multiple pools of addresses and we run this implementation in the Laboratory of Advanced Network Technologies, Faculty of Informatics, MU Brno, to test stability of the server and clients. The results so far with all of the clients are completely unsatisfactory.

4.6 LDAP

The METACenter LDAP service described in [6] started its experimental IPv6 support. It is accessible with both versions of IP protocol. Following section describe the current state of IPv6 support and experiences with that experiment.

4.6.1 MetaLDAP server

The LDAP server for METACenter is physically located in WEBnet, University of West Bohemia (UWB) network, outside the L2 METACenter network cloud. The server, known as `demeter.zcu.cz`, with service DNS alias `meta-ldap.cesnet.cz` in IPv4 world, was transformed into dual stack server, adding IPv6 support without changing the IPv4 part. It was assigned an global static IPv6 address from UWB prefix (`2001:718:1801::/48`) and new DNS names `demeter.ip6.zcu.cz` and `ldap.meta6.cesnet.cz`.

The LDAP server is based on OpenLDAP¹² implementation where support of IPv6 is ready without upgrades or changes. As the server provides secure access methods (SASL/GSSAPI and X.509), the change in DNS names described above introduces also a need of server trusted identity (`krb5.keytab`, new service X.509 certificate) change/extension.

4.6.2 LDAP Clients

Majority of clients currently used is based on OpenLDAP command line utilities (METACenter admin tools, `mk-gridmap` utility). So the first tests of IPv6 accessibility were focused on `ldapsearch` utility.

The anonymous LDAP test queries using the `ldapsearch` utility are working without problems. Example testing query:

```
ldapsearch -x -h ldap.meta6.cesnet.cz -b o=VOCE,dc=eu-egee,dc=org
```

Let us note that using IPv6 address in `-h` (host name) parameter is not possible because of URL escaping done here. If the user wants to use an IP address it is possible with `-H` parameter (URI):

```
ldapsearch -x -H 'ldap://[2001:718:1801:1052::211]' \
  -b o=VOCE,dc=eu-egee,dc=org
```

¹² <http://www.openldap.org/>

4.6.3 Security

For production usage of the MetaLDAP service, Kerberos authentication mechanism is needed. MetaLDAP suffers the same issue as described in Section 4.2 concerning multiple host names per host. Taking into account that multiple addresses (and hence also host names) are common part of IPv6 design, rigorous research is necessary to address this topic.

4.7 Job Scheduling Systems

PBSPro¹³ used in METACenter to schedule jobs does not support IPv6. The official website and documentation contains no information about this topic, the lack of support has been confirmed in mail conversation with a producer representative.

Magrathea^[3] is a system for managing virtual machines on physical nodes. IPv6 is not supported, nevertheless the network communication module is well encapsulated, so porting Magrathea to IPv6 should be a straightforward task.

4.8 Virtual Machines

Xen¹⁴ virtual machine monitor is transparent for IPv6. IPv6 support depends purely on kernels in hypervisor and user domains. We have encountered an inconvenience related to setting addresses for user domains in Xen. The preferred method is to configure IPv4 addresses on kernel command line in the Xen configuration. This way is not supported for IPv6, so the addresses must be set in the virtual machine image itself.

Linux VServer¹⁵ in current versions (i.e., 2.2) has to be additionally patched, see the IPv6 page¹⁶ for details. With this patch, the implementation is stable and functional.

4.9 User Management

System Perun^[4] for managing user accounts is a client/server application based on Oracle database. The critical part for IPv6 support is the copying configurations from Perun master to managed machines. This part is not IPv6 ready, but it is a simple client server applications over TCP using Heimdal Kerberos, therefore the changes should be simple.

4.10 Host and Service Monitoring

METACenter deploys Nagios¹⁷ system for checking health of machines and services. Nagios essentially uses ssh to run external programs installed on watched remote machines. The programs check status of the machines and the textual infor-

¹³ <http://www.pbsgridworks.com/>

¹⁴ <http://www.citrixserver.com/>

¹⁵ <http://linux-vserver.org/>

¹⁶ <http://linux-vserver.org/IPv6>

¹⁷ <http://www.nagios.org/>

mation is returned back to the server that processes the information and stores it in a database.

Nagios is declared to support IPv6 since 1.4 series. We have tested version 3.1 (with a set of home-grown patches that should not affect network functionality). It is able to run the testing ssh over IPv6 without troubles.

5 Conclusion

We have described status of deployment of the IPv6 in the METACenter infrastructure. The report covers addressing and naming plan as well as description of the link layer support necessary to create flat addressing suitable for experiments with virtual machines.

The infrastructure would have no meaning without applications. We have verified status of IPv6 support in key METACenter middleware. While widely used tools like web servers usually support IPv6 well, other software packages have issues of various seriousness: from minor inconveniences in configuration files to designs that practically make porting to IPv6 impossible. Configuration file parsing is usually easy to correct, on the other hand, when IPv4 structures are hard-coded in the program code, it is often easier to rewrite the code from scratch.

Many programs belong to a “middle class” from the IPv6 support point of view. Those programs suppose a one-to-one mapping between IP addresses and machines. This assumption is wrong even in the IPv4 world, but it usually cause no problems as it is the most usual situation. The same is true about expecting domain names to canonically identify machines as we have seen in the description of Kerberos and ssh (Section 4.2). In pure IPv6 or mixed environment, having more addresses per interface is a natural part of the design. Even a not-so-badly coded software that either doesn't encapsulate network operations into a module and/or having unsuitable interface to the networking module can be quite difficult to port to IPv6.

References

- [1] ANTOŠ, D. *METACentrum IPv6 Addressing Plan*. METACentrum CVS, 2007.
- [2] DANIELSSON, J.; WESTERLUND, A. Heimdal - an independent implementation of Kerberos 5. In *USENIX Annual Technical Conference*, 1998.
- [3] DENEMARK, J.; RUDA, M.; MATYSKA, L. Magrathea – Grid Management Using Virtual Machine. In *Cracow Grid Workshop '06*, pages , Academic Computer Centre CYFRONET AGH, Krakow, Poland, 2007, p. 138–145. ISBN 83-915141-7-X.
- [4] KŘENEK, A.; SEBASTIANOVÁ, Z.; SITERA, J. *Perun – systém pro řízení přístupu uživatelů k prostředkům METACentra* (Perun – user access control system for METACentrum resources). Technical Report 1/2004¹⁸, Praha: CESNET, 2004. In Czech.

¹⁸ <http://www.cesnet.cz/doc/techzpravy/2004/perun-2.2/perun-2.2.pdf>

- [5] NOVÁK, V.; ŠMRHA, P.; VERICH, J. *Deployment of CESNET2+ E2E Services*. Technical Report 18/2007¹⁹, Praha: CESNET, 2007.
- [6] SITERA, J. *Aktuální stav LDAPu METACentra* (Current status of LDAP in METACentrum). Technical Report 29/2003²⁰, Praha: CESNET, 2003. In Czech.
- [7] SITERA, J. *LDAP service for VOCE*. Technical Report 15/2005²¹, Praha: CESNET, 2005.

¹⁹ <http://www.cesnet.cz/doc/techzpravy/2007/cesnet-e2e-services/>

²⁰ <http://www.cesnet.cz/doc/techzpravy/2003/metaldap/>

²¹ <http://www.cesnet.cz/doc/techzpravy/2005/voceldap/>