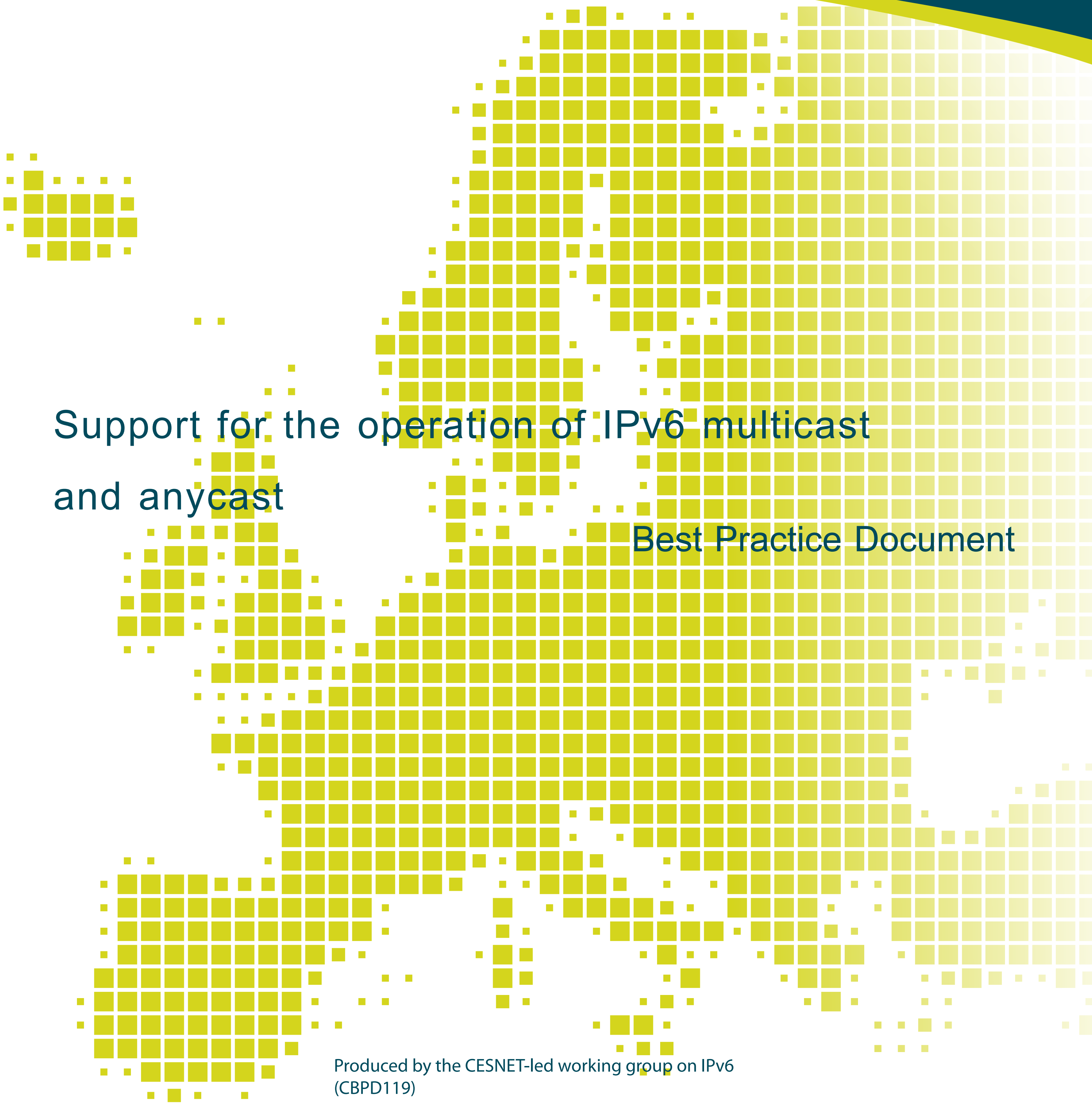

Support for the operation of IPv6 multicast and anycast

Best Practice Document

Produced by the CESNET-led working group on IPv6
(CBPD119)

Contributors:
Tomáš Podermański
Vladimír Veselý

November, 2012

Document No:	GN3-NA3-T4-CBPD116
Version / date:	1 November 2012
Original language:	Czech
Original title:	“ IPv6 Mýty a skutečnost, díl VII. - Podpora Multicast a anycast provozu”
Original version / date:	1 March 2011
Contact:	tpoder@cis.vutbr.cz

This translated version is based on the Czech version published in the electronic journal, Lupa.cz, on February 17, 2011.

The translation of this report has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 238875, relating to the project 'Multi-Gigabit European Research and Education Network and Associated Services (GN3)'.

Contents

Abstract

Multicast support under the IPv6 protocol is, in many ways, similar to multicast under IPv4. However, the additional address length has enabled the integration of some improvements. This document examines IPv6 multicasting in detail.

1 Introduction

Multicast is a concept used to distribute data from a single source to multiple recipients under the following fundamental conditions:

- to deliver the same data to multiple recipients in the most efficient way, making optimal use of links, or making certain that each data flow passes through each link only once and only branches out when it is absolutely necessary;

- to guarantee that every node or link where there are no recipients is not "burdened" by the data flow.

The simple definition of multicasting implies the use of online multimedia data transfer. However, its applications far exceed this use. For instance, it is also used for software installation on multiple PCs. One of the crucial roles of multicasting is the distribution of signalling protocols, in particular on the local network level.

1.1 Where multicast comes from

If we want to adopt multicasting, we must understand exactly what it is and how it works. In fact, it does not differ greatly from a television receiver, where channels are used in the distribution process. In IP terminology, the **multicasting group** serves the same purpose. The multicasting source sends out the data to the relevant multicast groups and clients in response to reports informing the network infrastructure that they are interested in receiving data from (joining) the relevant multicast group.

Specially allocated blocks of IP addresses fulfil the roll of multicast groups in the IP protocol, for both IPv4 and IPv6. The multicast data source simply sends the data to the IP address defined as a multicast group by stating the relevant group's address as the target address in the packet. From the address's format, the network recognises it as a multicast operation, treating the packets differently than if it were a unicast operation.

Multicast requires the client to behave differently than it does with unicast. While unicast requires the client to "passively" wait for some data to be delivered, with multicast, the client must first request it. The client sends a signal through the network that it requests a given group of multicast data. Based on this request, the network infrastructure will deliver the data to the client for as long the client requires.

1.2 Multicast addresses

A multicast group is identified by a special block of addresses in the form of `ff00::/8`.

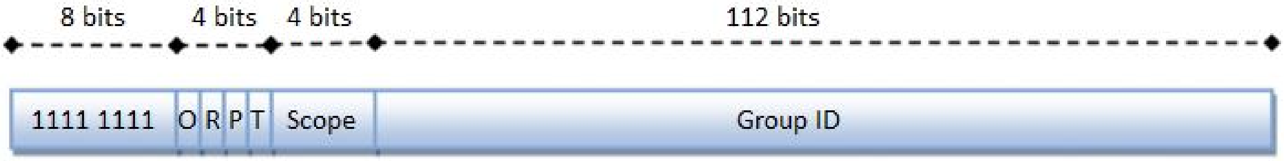

Figure 1: The general format of multicast addresses

The first 8 bits of a multicast address are set to a value of 1, which identifies it as a multicast address. The second 8 bits specify, in greater detail, the type of multicast address, and its range according to the following key:

- **0** – its use is presently reserved and must be set to 0 with each implementation;
- **R** – a 1bit flag or character identifying the Embedded-RP address (see Section 1.3);
- **P** – a 1bit flag identifying an address based on a unicast prefix (see Section 1.3);
- **T** – a 1bit flag defining it as a permanent/known network service multicast address assigned by IANA. A value of 1 is used if the address is dynamic or temporary.
- **Scope** – range of a multicast group. Because the *Hop Limit* field in the body of the Base Header[i] need not always be flexible; separate ranges are defined for multicast addresses according to the following table.

Value	**Title**	**Description**
0×1	Interface-Local	the range is within a single interface for loopback communication, which makes the communication between the individual interface addresses possible
0×2	Link-Local	the range is the same as with link addresses, usually the local network segment
0×4	Admin-Local	the smallest range that must be manually configured and does not depend on the physical topology of the network – used to define the subnet
0×5	Site-Local	the range covering the "space" (again, we come across the vague concept of space, which could be defined as a bank branch or building)
0×8	Organisation-Local	the range covering an entire organisation (such as a bank's entire network)
0×6, 0×7, 0×9, 0×A, 0×B, 0×C, 0×D	Unassigned	available for assignment and internal use for further segmentation of the range of multicast addresses, e.g., the CESNET2 network uses 0×A to define the entire country's range
0×E	Global	worldwide scope
0×0	Reserved	if a router finds a packet with such a range, it will simply delete it

0×3	Reserved	will not be assigned because it is reserved for another, future use
0×F	Reserved	if a router finds such a packet, it will deal with it as if it had a global range

- **Group ID** is a 112bit group identifier, usually divided into smaller sections with specific significance. RFC 3307[ii] implements the following hierarchy for the most accurate 32 bits for this item, where the most important bit copies the required T bit value.

Scope	**Description**
0000:0001 – 3FFF:FFFF	the permanent multicast addresses as they are assigned according to the IANA[iii] list
4000:0000 – 7FFF:FFFF	the permanent multicast group identifier, where important services (such as NTP or mens) are assigned the most accurate 32 bits from this group and where preceding bits do not play any role, because the given service has already been properly identified
8000:0000 – FFFF:FFFF	dynamic, multicast addresses serving non-allocated services and applications

Strict rules apply to the application of multicast addresses within IPv6, i.e., they must never appear as a source address within the *Base header* or in the *Routing header* (one of the *Distribution headers*). A router must never be able to direct multicast packets outside of its authorised range.

1.3 Multicast addresses based on an interface's prefix and identifier

One of the practical problems with multicasting is the choice of suitable multicast group addresses. These should always be chosen to avoid conflicts when one group is assigned to multiple sources. This is resolved by assigning a part of the multicast group address, either from the network prefix where the equipment is located (RFC 3306[iv]), or from the interface identifier (RFC 4489[v]). A multicast group address based on the network's prefix has the following format:

- **R** – a 1bit flag set to 0 when using this type of address;

- **P** and **T** – a 1bit flag assigned the value of 1 when using this type of address;

- The **Plen** or **Prefix Length** – determines the network prefix length (usually 48 or 64), with a maximum of 64 for this type of address;

- **Network Prefix** – separate network prefix value. For prefixes shorter than 64 bits the remaining digits/flags are set to 0;

- **Group ID** is 32bit use, according to the above-mentioned rules for identifying a group.

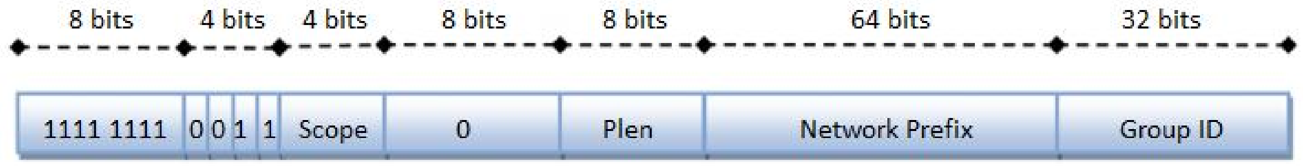

Figure 2: A multicast address based on the network prefix

Addresses based on an **interface identifier** have a very similar structure. In these cases, the *Plen* position (prefix length) is set to 0 and the *Network Prefix* is filled in by the interface identifier, typically created from a modified EUI 64. The use of addresses created by the interface identifier is most important when a global network prefix is not assigned (such as, in isolated networks using only link-local addresses).

2 Multicast routing

Multicast distribution and the setting up of paths from the source to the receiver will be a more complex process than it was with unicast. A router must be equipped with the means to allow communication and to determine whether there is interest within the end network for a certain multicast group. Furthermore, this entire process must be played out over a relatively short period of time (within a few seconds). Such a router uses a special signalling protocol referred to as the *multicast routing protocol*.

In general, we can divide multicast routing protocols into two basic groups. The first is a protocol entirely independent of routing information acquired from the unicast routing protocol (usually DVRMP[vi]), while the second group applies the unicast routing protocol. Presently this group mostly represents the *Protocol Independent Multicast* (PIM) protocol family. The definition, *Independent,* may be somewhat confusing here. In this case, the name only expresses its independence from a specific type of unicast routing protocol (OSPF, IS-IS, RIP) and not from the network layer protocol.

2.1 PIM Dense Mode (PIM-DM)

It is not difficult to think of several basic principles for distributing multicast routing information. The simplest regime in PIM protocol terminology is referred to as DM – Dense Mode. The router sends the relevant group's data to all interfaces, excluding the interface from which it obtained the data. In its initial stage, the network is literally flooded with the group's data. Based on this flood, the individual routers decide whether or not they want to receive it (join the group). If the router (which may be connected to the end network) decides that it does not want to join the multicast group, it sends a signal back to the source router with these instructions. In this way, the multicast distribution is gradually reduced, starting from the receiving end and moving towards the source end, when the receiving end does not express interest in receiving a multicast datagrams.

In a larger network, this trivial method of multicast distribution can cause relatively unpleasant complications during the initial flooding. This is why it is more suitable for smaller networks, where most clients are multicast receivers.

2.2 PIM Sparse Mode (PIM-SM)

The opposite of the Dense Mode is the Sparse Mode (PIM-SM). In this case, the end network router asks for direct connection to the multicast group. However, the end router has no idea where in the network the multicast data source may be found and the shortest path from it. For this reason, there must be a specific connection point within the network, which all routers are aware of and from where they know they can draw the multicast data/datagram. Such a connection point is referred to as the *Rendezvous Point* (RP). Once the initial contact to the agreed point (RP) has been made, optimisation can be made and the shortest path from the source can be established.

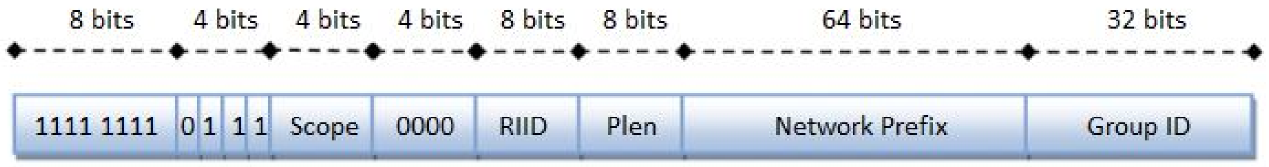

Figure 3: A multicast address with Embedded-RP support

The distribution of information concerning where an RP is located posed a significant problem under IPv4. It was necessary to apply a special protocol to distribute the specific RP information (Bootstrap Router Mechanism for PIM[vii]) and to pass this information between the individual RPs (Cisco-RP, MSDP[viii]). IPv6 offered a significant improvement, because the RP address is coded directly into the multicast address (RFC 3956[ix]). This type of address is referred to as an Embedded-RP multicast address.

- **R, P** and **T** – these 1bit flags are assigned the value of 1 when using this type of address;
- **RIID** – RP interface ID, where 4 bits in a single domain should be sufficient when creating addresses for several RP and may have a value ranging from 0x1 to 0xF;
- **Plen** – the prefix length, which must not be set to 0 and must not be greater than 64;
- **Network Prefix** – separate network prefix value. For prefixes shorter than 64 bits the remaining, unused digits are set to 0;
- **Group ID** – 32 bits used according to the above-mentioned rules to identify a group.

The Embedded-RP addresses are created in the following manner. If the multicast address falls within the scope `FF7x:y40:2001:DB8:BEEF:FEED::/96`, the resulting RP address will be `2001:DB8:BEEF:FEED::y`.

2.3 Source-specific multicast (SSM)

So far, we have only been considering multicast distribution through a multicast group. However, a problem arises when more sources send out data to the same multicast group. The nodes must receive all the data to a particular group and then define the requested data as either for transmission or for the application level.

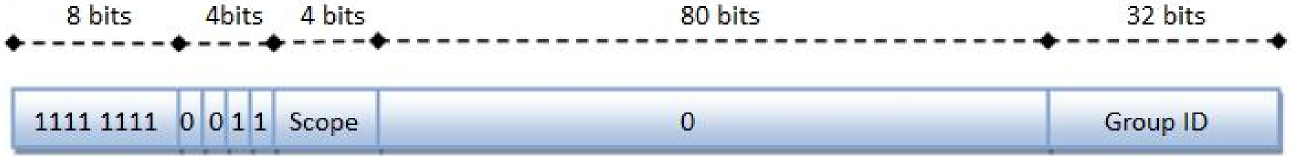

Figure 4: Multicast addresses for source-specific multicast

Source-specific Multicast[x] generally offers a solution to this. In this case, we use a combination of the source address and the multicast group to identify the multicast data. This pair is referred to as (S, G) and is called a

Channel. A distribution tree is created separately for each channel. In the same manner, the end nodes connect to the individual channels, or the (S, G) pair.

For SSM, IPv6 uses a special address format. This is derived from an address structure based on a unicast prefix and has the following format:

- **R** – a 1bit flag set to 0 when using this type of address;
- **P** and **T** – these 1bit flags are assigned the value of 1 when using this type of address;
- **Group ID** – 32 bits used according to the above-mentioned rules of identifying a group.

SSM addresses use the prefix `FF3x::/96` (where `x` is a valid range) and are differentiated only by the *Group ID*. At first glance, it appears that this could easily create a conflict when assigning the same group. However, SSM addresses are also tied to the source address, such that the tuple (source address and multicast group) defines a unique identifier of the multicast data.

3 Multicast in a local network

Multicast is applied somewhat differently in a local network. A local network has a limited scope and, in the simplest case, multicast may be distributed to all end nodes. In this way, multicast is practically pushed to an omnidirectional *(broadcast)* level of operation. The end equipment, now flooded with multicasting datagrams, simply chooses the one it is interested in or wants to join. All network devices have to treat multicast traffic in a different way. For this reason, multicast data in a network layer is mapped to special, link-layer, multicast addresses (MAC addresses). This mapping is accomplished by a very simple algorithm. The last 32 bits of an IPv6 multicast address group are used, mapped to the last 32 bits of a MAC address. The first 16 bits of a MAC address are assigned the value 33–33. An `FF02::1:FF68:12CB` IPv6 multicast group is mapped to a `33:33:FF:68:12:CB` Ethernet multicast. In this way, all equipment working on a link layer (switch) recognises it as a different type of operation from unicast and works with it in the appropriate manner.

3.1 MLD protocol

If we want to receive multicast from a different network, we must somehow inform the router on that network of this intention. Based on this request, the router takes the step to access the multicast from our local network. The MLD[xi] (*Multicast Listener Discovery*) protocol is used for this purpose. The end node expresses its interest in a multicast group by sending a report to the network to connect (*Join*) to the group, and based on this, the router can deliver the data from the relevant multicast group. In IPv4, the Internet Group Management Protocol[xii]) behaved in the same manner. A major upgrade under IPv6 is that MLD is an integral part of ICMPv6 reports, which was a separate protocol under IPv4. From a functional standpoint, both protocols are quite similar and, in principle, offer the same options.

IPv6	**IPv4**	**Description/options**
MLDv1	IGMPv2	entering or leaving a group, maintaining interest in a group
MLDv2	IGMPv3	expanded option to work with SSM (Source Specific Multicast), Filtering

3.2 MLD snooping

As was already stated, in the simplest case at the local network level, the multicast operation is mapped into the Ethernet multicast operation in the same manner as with omnidirectional broadcast. However, problems arise when multicast data is distributed to all equipment connected to the network. If, for example, a multicast operation is at 90Mb/s (such as, several streams from a larger number of cameras), each end node must somehow manage to process this operation. *MLD snooping* technology resolves this issue. Switches connecting the end nodes first listen for MLD communication initiated by the end nodes, and based on this, they can determine which multicast groups the end node want to join. Even though there are only L2 switches, equipment processing the MLD packets must understand the contents of the MLD protocol. For this reason, it must support the processing of these reports on the switch's interfaces. At present, MLD snooping tends to be supported by only higher priced equipment.

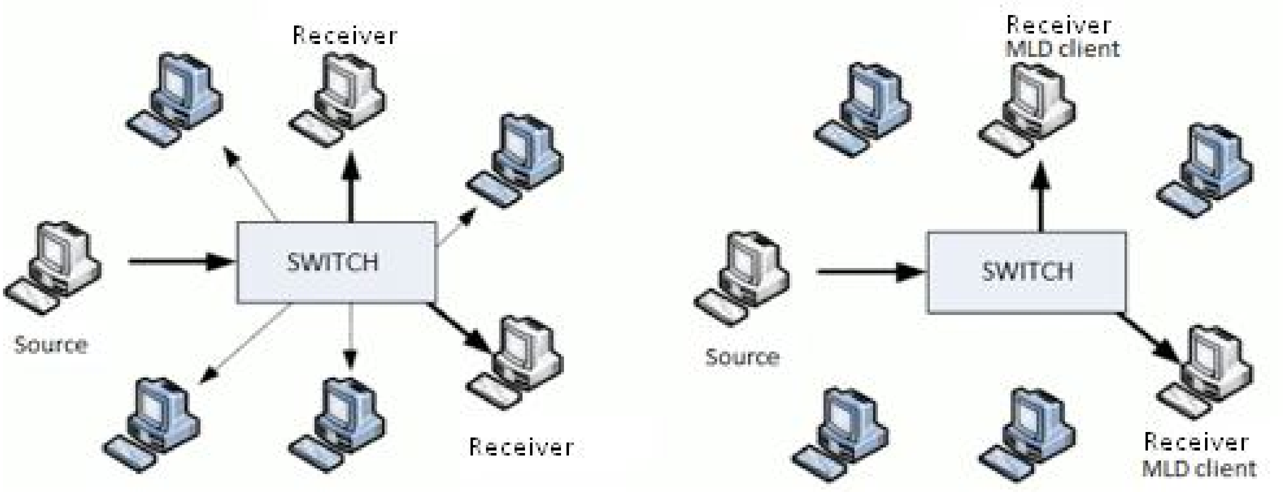

Figure 5: Switches with or without MLD snooping support

3.3 Broadcast and IPv6

IPv6 does not support omnidirectional broadcasting (*Broadcast*). The reason is rather simple. A broadcast is a special multicast case where all nodes in a network request to join a particular group. For these cases, it is not necessary to implement a special mechanism, requiring only the use of multicast.

However, if we look more closely at the signalling within the neighbour discovery protocol (*Neighbour Discovery for IP version 6*, RFC 4861[xiii]), we find that all nodes in IPv6 are required to follow the multicast group `FF02::1`. However, this poses a problem for equipment administering multicast groups (router, switch with MLD snooping support) where it is necessary to maintain the status information for all nodes connected to the network. For this reason, RFC 4861[xiv] instructs all packets containing the target multicast group `FF02::1` to be automatically transferred (broadcast) to all of the switch ports. In this way, some switches manage all multicast groups falling within the range `ff0x::/12`. The result is that the operation of selected multicast groups is distributed to all IPv6 nodes, so that this type of multicast behaves the same as a broadcast.

4 Anycast

Anycast is often compared to multicast. In reality, though, the two technologies do not have a great deal in common. While in multicasting, data from a single source is distributed to multiple end nodes, anycast operates in exactly the opposite manner. Data from multiple sources are distributed to the closest point in the network where such data may be processed. At first glance, it is clear that entirely different methods will be used for the distribution of anycast traffic.

There are many applications where anycast can be used. A good example is the use of anycast when communicating with recursive DNS servers. First, an agreement must be made that the relevant service will be accessible to specific IP addresses. The clients know of the agreement and, in response to a query to the recursive DNS server, they send a request to the agreed address. The network infrastructure sends the packet to the nearest recursive DNS server, which then processes the request. If a client moves to another network, requests sent to the recursive DNS server will be once again processed by the nearest server along the path. These factors reveal the positive aspects of anycast. In this way, the end system always receives the best possible response while this mechanism can be used to spread burden. Ten servers that are serving client requests can easily be covered by a single anycast address.

4.1 Format of anycast addresses

As was already explained, the transmission of anycast data does not greatly differ from regular unicast packet-transmission. Any unicast address or group of unicast addresses can become an anycast address. On a technical level, the anycast is handled by a special routing entry, which delivers the anycast packet to the required target. This would not be an issue in smaller networks, where we can afford either the enhancement of the routing entry on the routing protocol network or the “branching” of such an operation from the default direction at the level of the organisation that links the router. However, on a global level, anycast entries are unique within the routing table. This is okay as long as the list of entries is not too large. This means that a greater use of anycast by a greater number of services would result in a disproportional growth of entries in the global routing tables. For this reason, anycast is used in well-balanced and justified cases, or when a better method is not available.

One instance of the use of anycast is for root DNS server addresses. Considering that a basic knowledge of root DNS server addresses is required for the proper functioning of the entire DNS system, the root DNS server addresses must be as fixed and stable as possible. Some IPv6 transmission mechanisms and other services also use anycast technology.

5 Conclusion

The basic options offered by multicast and anycast under IPv6 are technically similar to those offered under IPv4. However, the ability to create a better assignment of multicast group addresses, in particular for SSM, and the ability to integrate IPv6 RP addresses directly into group addresses (Embedded RP) represents a substantial improvement.

IPv6 will continue to remain a technology that is reserved, for the most part, for special uses, in particular, for services in closed networks. Those who had thought that IPv6 would result in the global applicability of multicast will probably be disappointed. At the moment, this is not an option, and considering the present trends, probably never will be one (or at least, not for the foreseeable future), even though IPv6 greatly facilitates it and improves the mechanisms for its implementation.

The diagnostics of potential problems still remain the greatest issues for practical multicast operation, which is very difficult, especially considering the difficulty of its localisation. When a multicast is non-operational, it is practically impossible for an end user or service provider to determine whether the inoperability is caused by a configuration error in the local network (such as with one of the switches), or by an error with the router, RP, or the multicast routing protocol.

Anycast operation, as with IPv4, will always have room only for a limited set of specific applications, where the use of such anycast technology is well suited. On a global scale, this would only be limited to a few applications. Examples of this may be IPv6 protocol transmission mechanisms or access to root DNS servers.

Most of the mechanisms and principles of using the IPv4 protocol were implemented into IPv6 with minimal change, or implemented with minimal and uncontroversial improvements. If we consider some of the small differences in signalling (MLD), in practice, we would be able to work with multicast under IPv6 in the same way that we did under the IPv4 protocol.

6 List of illustrations

7 References

[i] http://www.lupa.cz/clanky/ipv6-myty-a-skutecnost-dil-v-zjednodusene-hlavicky/
[iiii] http://tools.ietf.org/html/rfc3307
[iii] http://www.iana.org/assignments/ipv6-multicast-addresses/ipv6-multicast-addresses.xml
[iv] http://tools.ietf.org/html/rfc3306
[v] http://tools.ietf.org/html/rfc4489
[vi] http://tools.ietf.org/html/rfc1075
[vii] http://tools.ietf.org/html/rfc5059
[viii] http://tools.ietf.org/html/rfc3618
[ix] http://tools.ietf.org/html/rfc3956
[x] http://tools.ietf.org/html/rfc4607
[xi] http://tools.ietf.org/html/rfc3810
[xii] http://tools.ietf.org/html/rfc4605
[xiii] http://tools.ietf.org/html/rfc4861
[xiv] http://tools.ietf.org/html/rfc4861