


IST-2001-32603	Deliverable D3.4.2	
----------------	--------------------	--

Project Number:	IST-2001-32603
Project Title:	6NET
CEC Deliverable Number:	32603/Partner/DS/No./A1
Contractual Date of Delivery to the CEC:	
Actual Date of Delivery to the CEC:	
Title of Deliverable:	D3.4.2 – Inter-domain Multicast
Work package contributing to Deliverable:	WP3
Type of Deliverable*:	R
Deliverable Security Class**:	PU
Editors:	João Nuno Ferreira
Contributors:	Stig Venaas, Jerome Durand, Mickael Hoerd, Jean-Jacques Pansiot
Reviewers:	Mónica Domingues, João Nuno Ferreira

* Type: P - Prototype, R - Report, D - Demonstrator, O - Other

** Security Class: PU- Public, PP – Restricted to other programme participants (including the Commission), RE – Restricted to a group defined by the consortium (including the Commission), CO – Confidential, only for members of the consortium (including the Commission)

Abstract: A general purpose multicast inter-domain service is lacking for IPv6, and in practice also for IPv4. We describe two possible solutions to provide this service: inter-domain ASM multicast with the Embedded-RP mechanism, and inter-domain SSM multicast with a source discovery mechanism.

Keywords: IPv6, Multicast, Inter-domain, ASM, Embedded-RP, SSM

Table of Contents

1. INTRODUCTION..... 3

2. EXTENSIONS TO BGP (MBGP) 3

3. THE ASM CASE..... 4

3.1. INTER-DOMAIN MULTICAST WITH IPV4 4

3.2. INTER-DOMAIN MULTICAST WITH IPV6: THE EMBEDDED-RP SOLUTION..... 4

3.2.1. Introduction..... 4

3.2.2. The proposal 5

3.2.3. Required modifications to PIM..... 6

3.2.4. Impacts on the network..... 6

4. THE SSM CASE 7

4.1. WHAT IS THE INTER-DOMAIN WITH SSM..... 7

4.2. PIM-SSM ROUTING 7

4.3. THE SSMSPIFIER - ASM APPLICATION SERVICES OVER SSM 8

5. CONCLUSION AND FUTURE WORK..... 8

6. MULTICAST GLOSSARY 9

7. REFERENCES..... 11

1. Introduction

While deploying multicast in a single domain is now well understood, things are not so well defined in the inter-domain case. Moreover different problems and solutions are different for ASM and SSM and whether we consider IPv4 or IPv6. In this document we consider only PIM-SM as the multicast routing protocol, since this is the only protocol that is both available and usable in inter-domain. In section 2 we present the multicast extensions of BGP (included in MBGP) needed for the operation of PIM-SM (both ASM and SSM). Then in section 3 and 4 we deal first with the ASM case and then with the SSM case.

2. Extensions to BGP (MBGP)

PIM-SM relies on unicast routing tables to perform so-called RPF (reverse path forwarding) checks. RPF is used to decide whether to accept a data packet, because with multicast the same data packet might be received on multiple interfaces and RPF is needed to avoid duplication and loops; and is also used to decide which neighbour a join message should be sent to. Inside a single domain one usually tries to make sure that unicast and multicast topologies are congruent. One can then use the unicast routing table for RPF and both configuration and management is relatively easy. If the topologies are not the same, one will need to maintain multicast routing tables. That is, separate routing tables, sometimes called MRIB, that are used only for multicast RPF checks.

For inter-domain multicast the unicast and multicast topologies can be different because there might be several paths between two multicast domains where only some can do multicast. If a domain has unicast routes to another multicast domain through multiple peering points, but multicast only on some of them, then one also need to distinguish between unicast and multicast routes in the domain.

A routing protocol is needed to exchange the multicast routes between different domains, and sometimes also inside the same domain. MBGP [RFC2858] is a widely used protocol for this.

BGP is usually used for inter-domain unicast routing and by using MBGP (multiprotocol BGP) one can exchange both unicast and multicast routes with BGP. MBGP adds additional attributes to BGP for specifying whether information is IPv4 or IPv6, and whether it's unicast, for multicast RPF checking or both. MBGP is sometimes, at least in multicast circles, called multicast BGP.

There is no difference between using MBGP for multicast in IPv4 and IPv6. Configuration should be quite similar, and it's used in the same way. Note that MBGP is needed not only when doing ASM, but also for SSM.

3. The ASM Case

3.1. Inter-domain multicast with IPv4

Here we will describe how inter-domain any-source multicast is done with IPv4. This differs quite a bit from how it is done with IPv6. For source-specific multicast there is no difference.

For both IPv4 and IPv6, PIM-SM is the most common multicast routing protocol. PIM-SM ensures that when a host in a domain joins a group, a shared tree is built from the RP in that domain, and also when a host in the domain is sending, the RP in that domain receives the data and is aware of the source. Doing this, PIM-SM can provide connectivity between sources and listeners when they are using the same RP. So by only using PIM-SM, it's necessary that all sources and receivers use the same RP.

For IPv6 this is basically how inter-domain ASM multicast is done with the Embedded-RP solution. For IPv4 however, there is a protocol called MSDP [RFC3618] that provides for communication between RPs in different domains. Using MSDP each multicast domain can have its own RP for groups of global scope, and a listener in one domain is able to receive from a source in another.

With MSDP, one sets up peerings between pairs of RPs. When an RP learns of a new source from PIM-SM, it will announce it to its MSDP peers. Also, a router receiving a source announcement from one peer, will forward it to its other peers. In this way the source announcements can be flooded throughout a network of peers. When an RP receives a source announcement for a group with local interest, someone in the RPs domain has joined the group, it will send a source specific join towards the source, building a SPT from the source in the other domain. Data received on the SPT can then be forwarded as usual.

By allowing each domain to have their own RP for all the global groups and using MSDP, one obtains a distributed model where no specific entity is needed to serve a group. This makes it easier to support multicast groups and sessions where no one in particular is responsible for it. One example might be the SAP [RFC2974] used for session announcements. This is a global service on a standardized group address. In contrast, for IPv6 there must for a given global group be a single RP, and the one operating the RP more or less owns the group or session. Note that MSDP is not scalable to a large number inter-domain of groups and sources since for each source there is a periodic flooding of a source announcement in the mesh of participating RPs. This is why MSDP has been considered by the IETF as an interim solution, waiting for another solution such as the now defunct BGMP architecture. MSDP has been deployed on a limited scale but there has been there as been problems with DDoS attacks. This explains why MSDP has not been specified for IPv6.

3.2. Inter-domain multicast with IPv6: the Embedded-RP solution

3.2.1. Introduction

Without MSDP, the construction of a shared tree with PIM-SM requires that all the PIM routers are configured with the same RP-set. A multicast group must then map to a unique RP in the entire Internet, as RPs cannot exchange information about IPv6 active sources. It is difficult to imagine a

protocol that would exchange information about existing RP's and corresponding multicast prefixes all over the Internet. BGMP had a similar goal but was too complicated and work on it has been terminated.

3.2.2. The proposal

A simple proposition came out then: to embed the RP address in the multicast address. This proposal is called Embedded-RP [RFC3956]. This seems impossible as both the RP address and the multicast address are 128 bits long. But making assumptions on the interface identifier of the RP, this solution is applicable.

An embedded-RP address has the following structure:

FF	0111	scope	res	RPad	plen	network prefix	group ID
8 bits	4 bits	4 bits	4 bits	4 bit	8 bits	64 bits	32 bits

Structure of an embedded-RP address

The following example describes how to build an IPv6 multicast address from the RP address. For an RP with the address 2001:660:3307:125::3, a multicast address can be derived the following way:

- *FF7*: Embedded-RP addresses start with "FF" since they are IPv6 multicast addresses. The 4 flag bits are "0111" or 7 in decimal.
- *scope*: 4 bits specifying the scope, similar to other multicast addresses.
- *res* (Reserved): the 4 bits of this field are set to 0;
- *RPad*: contains the last 4 bits of the interface identifier of the RP. In this particular example, RPad value is 3;
- *plen* (Prefix length): this field is the network prefix length for the RP address. In this example, the value is 0x40 (or 64 in decimal);
- *prefix*: contains the network prefix of the RP (2001:660:3007:125 for the example chosen);
- *group-ID*: this is the group identifier. It follows the specifications of RFC3513 and RFC3307.

Then an embedded-RP IPv6 multicast address for the RP having the address 2001:660:3307:125::3 will be FF7x:340:2001:660:3007:125:aabb:ccdd (aabb:ccdd being the *group-ID* chosen in this example).

The embedded-RP solution requires that the RPs have unicast addresses with interface identifiers having all bits, except the 4 low order bits, set to zero.

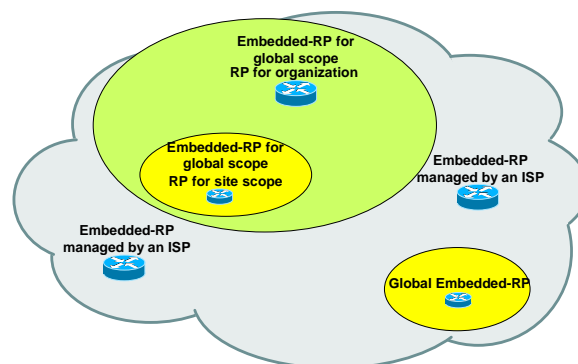
3.2.3. Required modifications to PIM

Embedded-RP requires that the group-to-RP mapping algorithm is changed. When a packet comes into the router with a destination address in FF70::/12 prefix, the RP address is retrieved as explained above.

Embedded-RP must be supported on all the routers of the shared tree, the RP and the DR of the sources and receivers. The support on the source tree is not required because the messages exchanged are PIM (S,G) prune/join. Nevertheless, one has to be aware that support on all routers eases the deployment and management of embedded-RP.

3.2.4. Impacts on the network

The IPv6 inter-domain multicast with embedded-RP is very different from what is done today in IPv4. The biggest change is that the PIM domains disappear with embedded-RP: the IPv6 multicast internet is a unique PIM "domain" where multiples rendezvous points are configured.



Embedded-RP model

- A side benefit of embedded-RP is that the address allocation problem is much easier than with IPv4. It is difficult to say now if this model will be accepted and deployed. Even if tests showed that the technology worked fine, some questions remain about the impacts caused by differences with the model known and deployed today. Also some general problems of the ASM case remain. For example: the third party dependency, multicast packets from domain A to domain B have to go through the domain of the RP;
- there is no source control: any source on the Internet can send into any inter-domain group;
- multicast data packets are encapsulated towards the RP (at least initially).

4. The SSM case

SSM (Source Specific Multicast) defines an IP multicast session with a single multicast source S , known by the set of multicast receivers. At the network level, only S is allowed to send in the channel identified by (S,G) . G is generally allocated by a local process running on the host having the IP address S . (S,G) is a channel identifier where groups subscribe specifically with an SSM capable host to router protocol like MLD version 2. We focus here only on the routing part of SSM, i.e. how the source specific multicast tree is built between SSM receivers and the source. We describe an end-host solution which enables multi-source sessions over SSM only networks.

4.1. What is the inter-domain with SSM

Strictly speaking, SSM is a subset/simplification of ASM and this is naturally the case for their implementation, respectively PIM-SM and PIM-SSM. The specifications of PIM-SSM are now included in PIM-SM specifications, so from now on we will call PIM-SSM the behavior of PIM-SM with SSM addresses (the $FF3x::/32$ prefix), and PIM-SM the behavior with ASM addresses.

The additional components needed for inter-domain ASM are used for the in-band automatic source discovery function and for group address allocation. These functions are not needed with SSM because S from (S,G) is supposed to be previously known by each receivers and because (S,G) is supposed to be globally unique in the Internet as S has been allocated by IANA.

We have seen that with PIM-SM, Embedded-RP is a candidate solution for large scale multicast deployment as it resolves partially the multicast group address allocation and the rendezvous point discovery issues. With this solution, the IPv6 ASM multicast Internet is an unique "PIM" domain where multiple PIM-RP are configured. SSM does not need any rendezvous point and address allocation scheme, and rely on a subset of PIM-SM to function. The IPv6 SSM multicast Internet is still an unique "PIM" domain, possibly not congruent with the unicast one with the help of MBGP but a lot easier to debug compared to PIM-ASM.

4.2. PIM-SSM Routing

A PIM-SSM router only needs to implement the upstream and downstream (S,G) state machine from the specification. It needs to implement the (S,G) assert and state machine for multiple PIM routers sharing a common LAN. Also Hello messages, neighbor discovery DR discovery and packet forwarding rules must be implemented.

With this subset, the implementation is greatly simplified compared to the $(*,G)$, (S,G,RPT) , $(*,*,RP)$ downstream an upstream state machine, the register state machine the $(*,G)$, assert state machine, the bootstrap RP election, the Keep alive Timer and the SptBit from a PIM-SM router. Also, the forwarding rules are more complex in PIM-SM.

By putting source discovery and address allocation outside of the network, SSM multicast routing does exactly what it's name indicates: multi-destination routing. Because the multicast network function is greatly simplified, debugging, securing, deploying an SSM multicast network is greatly simplified too.

Network operators may choose to deploy PIM-SSM only or PIM-SM, but if PIM-SM is deployed, PIM-SSM is implicitly deployed too. Note that PIM-SM and PIM-SSM can be used simultaneously since they use different address ranges.

4.3. The *ssmsdpifier* - ASM application services over SSM

The main missing feature with SSM is automatic source discovery in multi source applications. If we consider that some domains might deploy SSM only (no ASM), it is necessary to have this feature implemented in end hosts. We have specified and implemented a new application protocol called SSMSDP (Source Specific Multicast Source Discovery Protocol). In this solution, a multisource session is identified by a control channel (C, G). Session receivers listen to this control channel. Sources send announcements to the controller (C). These announcements are then forwarded into the control channel. This allows receivers to learn about new sources and to join the corresponding SSM channel. The controller can be seen as some kind of RP but at the session level, and only for signaling.

The implementation of SSMSDP consists in a low level library aimed at offering at least the same functionalities as ASM in the network does. We have developed a tool called "*ssmsdpifier*" based on this library. This tool allows the users to launch ASM applications (that is applications developed to run with ASM multicast), to be launched over SSM-only networks without patching/recompiling. Several well-known applications have shown to function without problems, including *vic*, *rat* and the multicast beacon.

Note that the lack of MLDv2 support in some operating systems (notably MS Windows) is currently limiting IPv6 SSM usage.

5. Conclusion and future work

A general purpose inter-domain IPv6 multicast service can be deployed using two different architectures : either ASM with embedded-RP, or SSM with a source discovery mechanism such as SSMSDP. It is hard to tell if a solution or the other will be widely accepted and deployed.

Embedded-RP is quite new, and it's not clear yet how the service best can be provided. One issue is how to provide failover or possibly load-balancing mechanisms. In order to do this, anycast-RP mechanisms may be needed. For IPv6 MSDP is not available, but a possible solution is "Anycast-RP using PIM", see draft-ietf-pim-anycast-rp-02.txt [ANYCASTRPPIM].

Another issue is address assignment. Users should not need to know what the RP address is and compute group addresses. Ideally applications or hosts should not care about RP addresses at all. But there should be some way the user or application can be told what group address to use, or at least a range to pick them from. Finally, another open issue is how to control usage of Embedded RPs. A provider or organization configuring a router as an Embedded RP, may wish to use it only for sessions for which at least some participants are customers or related to the organization.

There are also issues regarding SSM. One possible problem if SSM becomes widely used, is that routers in the Internet may end up with a lot of multicast state, since state must be store for every

single (S,G). This is a problem with current IPv4 ASM solutions as well. How to reduce this state might be a topic for new research into multicast protocols. There are some protocols solving this in the intra-domain case that cannot be used effectively inter-domain.

An issue with SSMSDP as specified and implemented today is that there is a single point of failure. There is now work going on to have redundant controllers to achieve both reliability and load sharing. It is expected that some interdomain testing will be done on the 6NET network.

A common issue for both Embedded-RP and SSMSDP compared to the current IPv4 inter-domain multicast with MSDP, is that some entity must be responsible for providing the Embedded RP or the SSMSDP controller. There might be long lasting sessions that have no natural owner. Maybe someone from one organization starts a session and picks a group address using the organizations Embedded RP, or runs a SSMSDP controller. It might be that that session lasts for a long time with multiple participants, and that the creator wishes to leave. It may then be desired to use another RP or SSMSDP controller for the session. This would however require either a new group G or a new channel (S,G) to be chosen, and for the applications to migrate to that, ideally with no interruptions to the service.

Finally, to debug and manage inter-domain multicast, it's very useful to have support for mtrace. mtrace is sort of multicast traceroute. There are a few implementations for IPv4, while we're only aware of one for IPv6. There are attempts at specifying this in the IETF, but it's not clear whether it will be standardized and how quickly.

6. Multicast Glossary

Anycast-RP - A way of using the same RP-address on several RP-routers to do load-balancing, and also fast fail-over, see [RFC 3446](#).

ASM - Any Source Multicast - ASM is the classical multicast service model as described in [RFC 1112](#) where any host can join a given multicast group G, and any host can send a packet with destination address G, and have it delivered to all members of the group G. The sender does not need to be a member of G. Compare with SSM.

Auto-RP - A dynamic protocol for configuring the group-to-RP mappings in a multicast domain.

BGMP - Border Gateway Multicast Protocol - An inter-domain multicast protocol. For each active multicast group, it builds a shared tree between domains that have senders or receivers for the group. See, [RFC 3913](#).

Bi-directional PIM - Uses shared trees, not only from RP towards receivers like in PIM-SM, but also in the other direction from sources towards the RP. There are no PIM registers. See [draft-ietf-pim-bidir-07.txt](#).

BSR - Bootstrap Router Protocol - A dynamic protocol for configuring the group-to-RP mappings in a multicast domain. See also PIM-SM specification, [RFC 2362](#).

Channel - This is used as a term for the source-group pair (S,G) in SSM; see SSM.

DR - Designated Router - The PIM-SM router on a link that is acting on behalf of the hosts on the link. When a host starts sending multicast, the DR sends register messages to the RP. When there are multiple PIM-SM routers on a link, one of them is elected as the DR. Initially the DR will also send join messages on behalf of the hosts and maintains tree state, but in some cases another PIM router on the link can take over; see last-hop router. See also PIM-SM specification, [RFC 2362](#).

Embedded-RP - A way of encoding the IPv6 RP-address in an IPv6 group-address. By using groups derived from the RP-address, routers can compute the RP-address from the group address so that they don't need any prior RP configuration. See [RFC 3956](#).

IGMP - Internet Group Management Protocol - Protocol used between hosts and multicast routers. It's used by IPv4 hosts to report multicast group membership to routers. The latest version is IGMPv3, see [RFC 3376](#). For source-specific reports, like in SSM, v3 is required. For IPv6, see MLD.

Last-hop router - The PIM-SM router on a link that is responsible for sending join messages on behalf of the hosts and maintaining tree state. This is the last router to forward the packets before they reach the host. This is initially the DR, but with multiple PIM routers on the same link, another router may become the last-hop router. See also PIM-SM specification, [RFC 2362](#).

MBGP - Multi Protocol BGP - This is often used for multicast. One may not always want the same topology for multicast and unicast. Using MBGP one can have BGP peerings exchanging both unicast and multicast prefixes independent of each other. See [RFC 2858](#).

MLD - Multicast Listener Discovery - Protocol used between hosts and multicast routers. It's used by IPv6 hosts to report multicast group membership to routers. For MLD, see [RFC 2710](#). For source-specific reports, like in SSM, v2 is required, see [RFC 3810](#). For IPv4, see IGMP.

MSDP - Multicast Source Discovery Protocol - An inter-domain protocol. It connects RP's in different domains, so that information about new sources can be distributed between the RP's. See [RFC 3618](#).

PIM - Protocol Independent Multicast - An inter-domain multicast routing protocol. Called protocol independent because it makes use of the unicast routing table for RPF, but is independent of which unicast routing protocols are used to populate the table. There are several variants, see PIM-DM, PIM-SM and Bi-directional PIM.

PIM-DM - Protocol Independent Multicast - Dense Mode - The dense variant of PIM. Dense means that it does flood-and-prune instead of only forwarding where requested. Compare with PIM-SM.

PIM-SM - Protocol Independent Multicast - Sparse Mode - The sparse variant of PIM. Called sparse because it only forwards where requested. Compare with PIM-DM. See also [RFC 2362](#).

RP - Rendezvous Point - For PIM-SM, this is used for source discovery. New sources registers with the RP, and initially traffic is forwarded on RPT which is rooted at the RP. See also PIM-SM specification, [RFC 2362](#).

RPF - Reverse Path Forwarding - RPF is used to determine where to send join messages, or from whom a packet should arrive. The RPF neighbour for a given address, is computed from the routing tables, and is often the next-hop a unicast packet with that destination address would be forwarded to. PIM uses RPF to find out where to send join messages and also performs a so-called RPF check, discarding data packets arriving on the wrong interface. RPF is used by many multicast protocols and flooding mechanisms, also BSR. See also PIM-SM specification, [RFC 2362](#).

RPT - RP Tree (aka shared tree) - The tree that is rooted at the RP and created by (*,G) joins from last-hop routers. For ASM this tree is used at least initially, last-hop routers may create SPT's and switch to them. See also PIM-SM specification, [RFC 2362](#).

RTCP - RTP Control Protocol - RTCP is used by RTP receivers to report on the quality of the data distribution. It also includes a string identifying the receiver. See RTP specification, [RFC 3550](#).


RTP - Real-time Transport Protocol - RTP provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. Those services include payload type identification, sequence numbering, timestamping and delivery monitoring. It is often used with UDP multicast transport but can also work with other underlying transports. [RFC 3550](#).

SPT - Shortest Path Tree - This tree is rooted at the source, or rather at the source's DR, and the leaves are last-hop routers and/or the RP. It's built by (S,G)-joins. See also PIM-SM specification, [RFC 2362](#).

SSM - Source Specific Multicast - Instead of joining a group G, hosts join a so-called channel (S,G), and the host will only receive from the source S. A host can join several sources with the same group. The source does not need to be member of the group. See [draft-ietf-ssm-arch-06.txt](#).

7. References

- [RFC2858] RFC 2858: "Multiprotocol Extensions for BGP-4", Bates, T, Rekhter, Y., Chandra, R., Katz, D., June 2000 (Status: Internet Official Protocol Standards);
- [RFC2974] RFC 2974: "Session Announcement Protocol", Handley, M., Perkins, C. Whelan, E., October 2000 (Status: Experimental Protocol for the Internet community);
- [RFC3618] RFC 3618: "Multicast Source Discovery Protocol (MSDP)", Fenner, B., Meyer, D., October 2003 (Status: Experimental Protocol);

IST-2001-32603	Deliverable D3.4.2	
----------------	--------------------	--

[RFC3956] RFC 3956: “Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address”. P. Savola, B. Haberman. November 2004. (Format: TXT=40136 bytes) (Updates RFC3306) (Status: PROPOSED STANDARD)

[ANYCASTRPPIM] PIM Internet Draft: “Anycast-RP using PIM”, Dino Farinacci, June 2004.