

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2015

L. Yong
D. Cheng
W. Hao
D. Eastlake
Huawei Technologies Ltd.
A. Qu
MediaTek
J. Hudson
Brocade
U. Chunduri
Ericsson Inc.
March 4, 2015

IGP Multicast Architecture
draft-yong-pim-igp-multicast-arch-00

Abstract

This document specifies the architecture of IP multicast routing using an Interior Gateway Protocol (IGP).

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 5, 2015.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/license-info) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
1.1.	Overview	3
1.2.	Motivation	3
1.3.	Conventions used in this Document	4
1.4.	Terminology	4
2.	An Overview of IGP	5
3.	Scope	6
4.	Routing IP Multicast Packets	6
4.1.	Multicast Distribution Tree	7
4.1.1.	Bidirectional Distribution Tree	8
4.2.	Advertising Multicast Group Membership	9
4.3.	Requirements of Edge Routers	9
4.4.	Intra-Area Multicast Routing	10
4.5.	Inter-Area Multicast Routing	11
4.5.1.	Behavior of IS-IS L2 Router	11
4.5.2.	Behavior of OSPF ABR	11
4.6.	Heterogeneous Environment	11
4.7.	TE (Traffic Engineering) Support	12
4.8.	Applications with Overlay Model	12
4.9.	IPv4 and IPv6	12
5.	Acknowledgement	13
6.	References	13
6.1.	Normative References	13
6.2.	Informative References	13
	Authors' Addresses	14

[1.](#) Introduction

1.1. Overview

In an IP network, IGP is used to route and forward IP unicast packets. In doing so, the routers collect and maintain the network information and store them in their database. The network information includes the identity of the routers and their interconnections. In traffic engineering enabled network, the information also includes traffic related parameters such as link bandwidth. The network information that has already maintained on routers, along with some minor IGP protocol extension as proposed in this document, are sufficient to route IP multicast packets. This means a single IGP can be used for routing both unicast packets and multicast packets. This document describes the architecture of routing IP multicast packets using the network information that is disseminated by IGP.

1.2. Motivation

With the explosion of IP technology based applications, the support of IP multicast delivery over the same IP network that carries IP unicast traffic becomes mandatory. In many aspects, some basic requirements for routing IP multicast packets are the same as those for routing IP unicast packets; e.g., the "plug and play" nature of bringing up the routing engine and enabling the packets forwarding. It is desirable to use IGP that requires minimum configuration and currently only routes and forwards IP unicast packets, also to route and forward IP multicast packets.

Currently in an IP network, a separate protocol such as Protocol Independent Multicast (PIM - [[RFC4601](#)]) must be used to route and forward IP multicast packets, whereby some network information are actually retrieved from IGP. Using a single protocol, i.e., an IGP, to route both IP unicast and multicast packets is with much efficiency; e.g., there is no additional convergence time otherwise would be introduced by the second protocol. Using one protocol also reduces the operational complexity.

In an advanced data center network, it requires the decoupling of network IP space from service IP space, e.g., a VxLAN based network overlay [[RFC7348](#)]. To support all service applications, such IP network fabric must support both unicast and multicast. Decoupling network IP space from service IP address space also provides network agility and programmability. If network IP space is decoupled from service IP space, the network itself no longer needs manual configuration; automatically forming an IP network fabric can be done.

1.3. Conventions used in this Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

1.4. Terminology

This document makes use of the following terms:

- o Edge Router: A router that has direct interfaces with one or more IP hosts.
- o Distribution Tree: a rooted distribution tree with one root and one or more leaves and it is used facilitate routing multicast packets.
- o IGP: Interior Gateway Protocol.
- o Intra-Area: Refer to the communication between IGP routing nodes within a single IGP's area.
- o Inter-Area: Refer to the communication between IGP routing nodes across area boundary.
- o IP Multicast Group
- o Link State Database: The database is constructed and maintained by a router running link state based routing algorithm such as IS-IS and OSPF. It contains network based information including identity of routers and their interconnections, reachable IP addresses, etc.
- o Local Group Database: The database is constructed and maintained by an edge router that stores and maintains entries of multicast-address, host pair.
- o Pruned Tree: A subset of IGP's topology graph and with a tree root, from where, multicast packets are forwarded to one or more destination nodes with optimization of the usage of links and nodes.
- o Root Node: A router served as a root in a multicast distribution tree.
- o TE (Traffic Engineering) Database: The database is constructed and maintained by a router running link state based routing algorithm with TE extensions such as ISIS-TE and OSPF-TE. It contains TE

parameters (such as bandwidth) that are associated with links and nodes.

- o Transit Router: A router that is capable of receiving an IP multicast packet, then replicates it and sends to one or more other routers on the downstream direction in the same multicast distribution tree.

2. An Overview of IGP

There are currently two most deployed IGPs, and they are IS-IS [[RFC1195](#)]/[[RFC5308](#)] and OSPF [[RFC2328](#)]/[[RFC2740](#)]. IS-IS and OSPF are different in many aspects, but they both use link-state algorithm and the network information they disseminate for the same IP network are the same, including routers' IP addresses, routers' interconnections, reachable IP addresses, the network topology, etc.

An IGP operation is with hierarchy. An IGP runs within an area, where each participating router originates and advertises its own information (router's identity, interface IP addresses, identity of directly connected neighbors, etc.), and this information converges throughout the entire area but not beyond. As a result, within an IGP area, each participating router maintains the information of all routers and their interconnections. We call the collection of the network information as Link State Database, which is currently used as a base to calculate IP routing table for unicast packets within an IGP area. Sometimes we refer to the topology within an IGP area as a topology graph. Separate IGP areas may be interconnected and between areas, only reachability information is advertised across area boundary by Level-2 router in IS-IS or Area Border Router (ABR) in OSPF.

[RFC1195] specifies an IGP for routing IPv4 unicast packets using IS-IS protocol (ISO), whereas [[RFC5308](#)] specifies the extensions to support routing IPv6 unicast packets.

OSPFv2 [[RFC2328](#)] is an IGP for routing IPv4 unicast packets whereas OSPFv3 [[RFC2740](#)] is an IGP for routing IPv6 unicast packets.

The link state based routing algorithm in OSPF and IS-IS calculates the shortest path from the source to the destination. A routing table for routing unicast packets is generated on every participating IGP router.

For some applications, path restrictions (e.g., link bandwidth) need to be considered. As a result, extensions are added to both IS-IS and OSPF to support traffic engineering based unicast routing as follows:

- o [[RFC3630](#)] - Traffic Engineering (TE) Extensions to OSPF Version 2
- o [[RFC3784](#)] - Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)
- o [[RFC5329](#)] - Traffic Engineering Extensions to OSPF Version 3

A TE-capable IGP router, in addition to construct a Link State Database, also constructs and maintains a TE Database that stores the traffic parameters (e.g., bandwidth) associated with links and nodes, where the information is used for constraint based consideration during normal shortest path calculation.

3. Scope

To support IP multicast routing, either IS-IS or OSPF can be used and in the perspective of this document, there is no difference in choosing. And there requires no change in IS-IS or OSPF, except that extensions are needed in both protocols to advertise and store distribution tree root node address and multicast group receiver information, refer to [Section 4.2](#).

Using IGP to route IP multicast packets is within IGP's architecture and routing paradigm. IP multicast routing within an IGP area is called intra-area multicast routing, and IP multicast routing across IGP area is called inter-area multicast routing. The concept, rules and behavior regarding intra-area unicast routing and inter-area unicast routing are all similarly applicable to intra-area and inter-area multicast routing, respectively.

In an IPv4 network, IPv4 multicast packets can be routed using IS-IS (based on [[RFC1195](#)]) or OSPFv2 as introduced by this document. Similarly in an IPv6 network, IPv6 multicast packets can be routed using IS-IS (based on [[RFC5308](#)]) or OSPFv3 [[RFC2740](#)]. As the networking industry is currently under transition from IPv4 to IPv6, co-existence of the two is sometimes required. Using the architecture described in this document, IPv4 multicast packets can be transported over an IPv6 network and vice versa, IPv6 multicast packets can be transported over an IPv4 network.

4. Routing IP Multicast Packets

As illustrated in Figure 1, a single IGP can be deployed to support both IP unicast and multicast routing.

This section describes routing IP multicast packets using the existing network information that IGP collects, the related functions

and characteristics, along with the required extensions to existing IGPs.

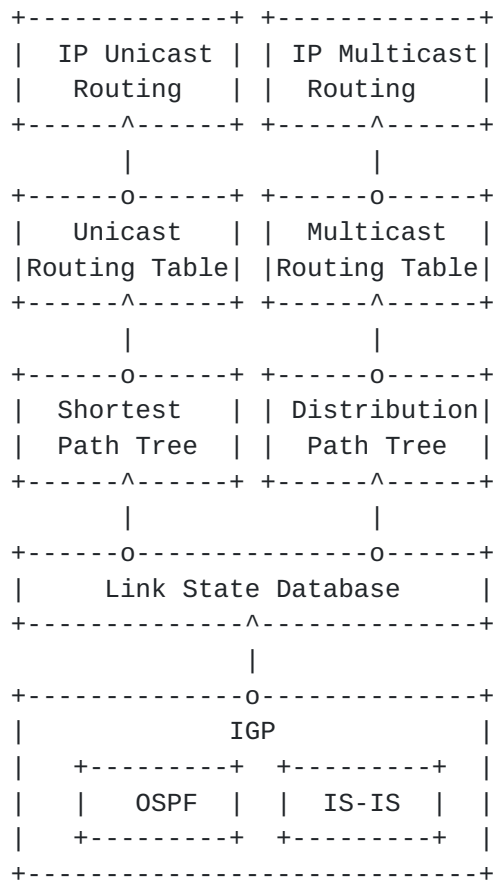


Figure 1: Using an IGP to Route both IP Unicast and Multicast Packets

4.1. Multicast Distribution Tree

To route IP multicast packets, it requires a distribution tree. A distribution tree consists of a tree root, one or more tree leaves, and some branch nodes. The tree root is identified by the IP address (or Router ID) of an arbitrary router. The tree root can be configured for a specific IP multicast address group, or automatically elected via an algorithm. A tree leaf is an edge router and is a multicast destination. A tree leaf is identified by an edge router's IP address and it is directly attached to one or more hosts that advertise the IP multicast group addresses (see [Section 4.2](#) for details). A router that is not a tree root but transmits a received IP multicast packet to another router is called a Transit Router, which is a branch node in the distribution tree.

In general, there is a single multicast distribution tree for each IP multicast address group. Once a distribution tree is formed, an IP packet with the multicast destination address is forwarded according to the multicast distribution tree, i.e., from the tree root to all tree leaves.

Via configuration, additional distribution tree can be constructed for the same IP multicast address group, however with different tree root and tree branches (paths). This option provides a redundancy for routing path protection, and it can also be used to support load balance.

When a leaf node of a multicast distribution tree is in the same IGP area as the tree node, the packet flow from the root to the leaf is within a single IGP area. This behavior is called IGP intra-area multicast routing.

When a leaf node of a multicast distribution tree is in a different IGP area as the tree node, the packet flow from the root to the leaf must cross IGP area boundary. This behavior is called IGP inter-area multicast routing.

Unicast routing in an IGP domain requires minimum configuration. This characteristic is inherited for multicast routing, i.e., there requires minimum configuration and a multicast distribution tree can generally be constructed quickly in the same manner as a unicast routing table.

4.1.1. Bidirectional Distribution Tree

The IP multicast distribution tree as described above is uni-directional, i.e., all leaf nodes can only receive multicast packets destined to a given multicast address. In this scenario, the tree root may be the traffic source and if not, the source must unicasts packets to the tree root, which then distributes the packets according to the distribution tree. The uni-directionality of distribution tree is useful for applications such as video broadcasting.

A multicast distribution tree can also be constructed as bi-directional. In a bi-directional distribution tree, IP multicast packets destined to a given multicast address can traverse on any tree branch in both directions; that means any leaf node on the tree can be a multicast receiver but also a sender. When a tree leaf node is a sender, it transmits its multicast packets to all other leaf nodes according to the bi-directional distribution tree. The bi-directionality of distribution tree is useful for applications such as network virtualization overlays ([[RFC7365](#)]) and video conference.

The algorithm to build a uni-directional distribution tree is in general different from that to build a bi-directional tree. In both cases, care must be taken in order to build an optimized multicast distribution tree, such as the consideration of the average path length from the root to leaf nodes, the total links (branches) used for the distribution, etc.

Configuration (along with a default) may be used to specify the directionality of an IP multicast distribution tree for a given IP multicast address group.

4.2. Advertising Multicast Group Membership

In order to support multicast routing, an IGP must be extended to store and advertise IP multicast addresses in the similar manner currently for IP unicast addresses.

Pairs of [multicast-group, host] can be configured on an edge router, or learned from the interaction with IGMP/MLD(see [Section 4.3](#)). In either case, the router would be required to advertise the IP multicast group membership throughout the IGP area. The advertising, refresh, aging, and removal of IP multicast addresses are handled in the same manner as the existing database element, i.e., LSP in IS-IS and LSA in OSPF.

IP multicast addresses can also be advertised across IGP area boundary using similar mechanism as for IP unicast addresses. IP multicast addresses may be summarized similar to that of IP unicast addresses for scaling purpose.

The details of storing and advertising IP multicast address using IS-IS is specified in a separate document.

The details of storing and advertising IP multicast address using OSPF is specified in a separate document.

4.3. Requirements of Edge Routers

To support routing IP multicast packets, edge routers, i.e., routers that have interfaces directly connected to IP hosts, are required to run IGMP (IGMPv2/[[RFC2236](#)] or IGMPv3/[[RFC3376](#)]) for IPv4 based hosts and MLD (MLD/[[RFC2710](#)] or MLDv2/[[RFC3810](#)]) for IPv6 based hosts.

As the result of interaction with hosts, an edge router would build a Local Group Database where each entry is a [multicast-group, host] pair, which indicates that the attached host belonging to the IP multicast group. This process is on-going in order to keep track of

the IP group membership addresses of attached hosts and strictly according to protocol specification of IGMP/MLD.

The Local Group Database is used in two folds. First, when an edge router receives an inbound IP multicast packet, it checks in the database to see if any entry that has any matching IP multicast-group address against the destination address in the received packet, and if so, the packet is forwarded to the local host(s); otherwise the packet is dropped. Note this behavior already exists on edge routers that support IP multicast forwarding.

Second, an edge router is required to advertise/flush the IP multicast addresses learnt/withdrew from IGMP/MLD procedure to/from other routers in the same IGP area, in the similar manner as advertising/flushing its own interface IP addresses. With this procedure, an IP multicast distribution tree can be built for each IP multicast address group. The details for advertising multicast addresses by IS-IS and OSPF will be documented separately.

In some deployment, a host as a multicast destination or source may connect to more than one edge routers for the purpose of reliability or/and load balance, as normally termed as multi-homing. In this scenario, care must be taken in order to prevent from forwarding loop as well as packets duplication.

4.4. Intra-Area Multicast Routing

An IP multicast distribution tree within an IGP area is in effect a sub-graph of the IGP's area topology graph (see [Section 2](#)). All routers that receive advertisement of IP multicast addresses in the IGP area must build the multicast distribution tree for each IP multicast address group. The construction of the distribution is based on the IGP's Link State Database, which is currently used for routing IP unicast packets. All routers in an IGP area must calculate and construct the intra-area distribution tree using IGP's Link State Database with the same algorithm, so that a pruned tree can be constructed for the distribution tree. Care must be taken to avoid forwarding loops and routing optimization is highly desired.

Note the algorithm for constructing an IP multicast distribution tree, and other related functions are outside of any specific IGP, i.e., there requires no change in IGP.

The algorithm and related details for intra-area multicast routing is specified in a separate document.

4.5. Inter-Area Multicast Routing

In inter-area unicast routing, an IP packet from one IGP area forwarded to another area is sent to an area border node (ABR for OSPF) or L2 router (for IS-IS) first, which then forwards the packet to the neighboring area. This is also the scenario for inter-area multicast routing, and as such, an ABR/L2-Router functions as a Transit Router, or a branch node in the multicast distribution tree.

Note that IGP's Link State Database is per area, so the multicast distribution tree constructed on routers in the transmitting area in generally terminated at the ABR/L2-Router due to lack of routing information. The ABR/L2-Router in question would require extending the distribution in the receiving area based on the separate Link State Database.

The procedure and related details for inter-area multicast routing is specified in a separate document.

4.5.1. Behavior of IS-IS L2 Router

For IS-IS, the area boundary is on the link, and so the L2 router in the receiving area extends the distribution tree for that area.

To support inter-area multicast routing, an IS-IS L2 Router is required to propagate IP multicast addresses received in one area to all L2 Routers in other areas it is connected. This behavior is similar to the advertisement of IS-IS Reachability Information PDU.

4.5.2. Behavior of OSPF ABR

For OSPF, the area boundary is on the ABR. When an ABR attached to both transmitting area and receiving area, it extends the distribution tree in the receiving area.

To support inter-area multicast routing, an OSPF ABR is required to propagate IP multicast addresses received in one area to all other areas it attached. This behavior is similar to the advertisement of OSPF Summary LSA.

4.6. Heterogeneous Environment

To deploy the IP multicast routing using IGP as described in this document, it requires all routers in the IGP area implement the following:

- o Implement the extension in IS-IS (documented separately) and in OSPF (documented separately) for advertising multicast addresses.

- o Support the new functions as described in [Section 4](#).

In a heterogeneous network environment, i.e., not all routers in an IGP area implement the above extensions. A multicast distribution tree within an area does not allow to be segregated, but tunneling mechanism can be used to support multicast routing here. When there are routers that would be required to be on a multicast distribution tree but not supporting the required extensions, a tunnel is constructed connecting two adjacent routers capable of routing multicast and across one or more not-capable routers, such that the tunnel becomes a single branch on the distribution tree. An IP multicast packet sent from a tunnel end to the other is encapsulated in an IP packet with the sending router's IP address as the source address and the receiving router's IP address as the destination address.

[4.7.](#) TE (Traffic Engineering) Support

The existing IP multicast routing practice (e.g., PIM) does not consider route constraints (e.g., link bandwidth). Both OSPF and IS-IS support traffic engineering based unicast routing by constructing and maintaining a TE Database. Like Link State Database, the TE Database can also be used to support IP multicast routing when one or more path constraints is under consideration.

Note to perform TE based multicast routing using IGP, routers must support TE extensions, and otherwise, there requires no other change in the IGP.

[4.8.](#) Applications with Overlay Model

Using a single IGP as a uniformed routing engine for both IP unicast and multicast routing enables a simple but highly efficient IP networking fabric that can serve varies applications above it as a overlay model. These applications are viewed as at the service level, completely decoupled with the underneath IP networking fabric however enjoy both IP unicast and multicast transportation infrastructure. In the multicast perspective, the applications can be IP based, but can also be level-2 based such as Ethernet.

[4.9.](#) IPv4 and IPv6

The architecture as outlined in this document supports IPv4 multicast routing in IPv4 networks, and also IPv6 multicast routing in IPv6 networks.

With mechanisms such as tunneling or address translation, the same architecture can also support IPv4 multicast routing in IPv6

networks, and IPv6 multicast routing in IPv4 networks. The details are specified in other document.

5. Acknowledgement

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

6.2. Informative References

- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", [RFC 1195](#), December 1990.
- [RFC2236] Fenner, W., "Internet Group Management Protocol, Version 2", [RFC 2236](#), November 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), April 1998.
- [RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", [RFC 2710](#), October 1999.
- [RFC2740] Coltun, R., Ferguson, D., and J. Moy, "OSPF for IPv6", [RFC 2740](#), December 1999.
- [RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", [RFC 3376](#), October 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", [RFC 3630](#), September 2003.
- [RFC3784] Smit, H. and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", [RFC 3784](#), June 2004.
- [RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", [RFC 3810](#), June 2004.
- [RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", [RFC 4601](#), August 2006.

- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", [RFC 5308](#), October 2008.
- [RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, "Traffic Engineering Extensions to OSPF Version 3", [RFC 5329](#), September 2008.
- [RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", [RFC 7348](#), August 2014.
- [RFC7365] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for Data Center (DC) Network Virtualization", [RFC 7365](#), October 2014.

Authors' Addresses

Lucy Yong
Huawei Technologies Ltd.
Austin, TX
USA

Email: lucy.yong@huawei.com

Dean Cheng
Huawei Technologies Ltd.
2330 Central Expressway
Santa Clara, CA 95135
USA

Email: dean.cheng@huawei.com

Weiguo Hao
Huawei Technologies Ltd.
101 Software Avenue
Nanjing 210012
China

Email: haoweiguo@huawei.com

Donald Eastlake
Huawei Technologies Ltd.
155 Beaver Street
Milford, MA 01757
USA

Email: d3e3e3@gmail.com

Andrew Qu
MediaTek
San Jose, CA 95134
USA

Email: laodulaodu@gmail.com

Jon Hudson
Brocade
130 Holger Way
San Jose, California 95134
USA

Email: jon.hudson@gmail.com

Uma Chunduri
Ericsson Inc.
300 Holger Way
San Jose, California 95134
USA

Email: uma.chunduri@ericsson.com

