IDMR Working Group
Internet Engineering Task Force
INTERNET-DRAFT
26 February 1999
Expires August 1999

Dave Thaler
Microsoft
Brad Cain
Nortel Networks

BGP Attributes for Multicast Tree Construction
<draft-ietf-idmr-bgp-mcast-attr-00.txt>

Status of this Memo

Copyright Notice

Abstract

The Multiprotocol Extensions for BGP-4 [MBGP] allow Network Layer Reachability Information to contain prefixes used for multicast forwarding. This document defines extensions to BGP-4 [BGP-4] which can be used to annotate such prefixes with information that can be used by multicast routing protocols when constructing trees.

Copyright Notice

Copyright (C) The Internet Society (1999). All Rights Reserved.

1. Introduction

The Multiprotocol Extensions for BGP-4 [MBGP] allow Network Layer Reachability Information to contain prefixes used for multicast forwarding. These prefixes may be "come-from" unicast prefixes or multicast "go-to" prefixes (i.e. Class-D addresses). Multicast routing protocols use these prefixes to construct multicast distribution trees.

We describe two BGP path attributes that may be used with prefixes used for multicast forwarding. The Forwarder Preference attribute is used by BGP speakers to elect a single forwarder on a multi-access link. The Data Flow Direction attribute describes the "directional" nature of the multicast tree for a given prefix. It may be used by multicast routing protocols which have the ability to construct multiple tree types (unidirectional and bidirectional trees), or to limit which routes can be used by multicast routing protocols which can only construct a single tree type.

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 RFC 2119 [RFC2119].

2. Forwarder Preference - FWDR_PREF (Type Code XXX):

In the case of a multi-access data link layer, multicast packets are sent natively on the link, and require special handling. Specifically, when packets are multicast natively on the link, each packet must only be sent onto the link by a single router, or duplicates will result.

In order to reduce the number of duplicates sent on multi-access links, current multicast routing protocols elect a designated forwarder per route or use a data-driven "assert" mechanism. Some protocols (e.g. DVMRP [DVMRP]) elect a designated forwarder per route. A multicast tree branch using that route then uses the designated forwarder on the multi-access link. Other protocols [PIM-DM,PIM-SM] use data-driven mechanisms since they may use route exchange protocols which do not

elect a designated forwarder, and run over links with multiple route exchange protocols. However, this mechanism has the disadvantage of periodically generating short bursts of duplicates.

The forwarder preference attribute is for multicast routing protocols which use multi-protocol BGP for route selection. This attribute allows a designated forwarder to be elected on a multi-access link. Designated forwarder election only works, however, when full-peering exists on the multi-access link, but this is indeed the case on multicast friendly interconnects [MIX] today. Operation over multi-access links in the absense of full-peering is outside the scope of this document.

2.1. FWDR_PREF Details

FWDR_PREF is an optional non-transitive attribute for the purpose of electing a single designated forwarder on a multi-access link. FWDR_PREF is used to inform other BGP speakers on the same link of the originating speaker's degree of preference for an advertised route for the purpose of electing a single designated forwarder on a multi-access link. FWDR_PREF is a four-octet non-negative integer with type code of XXX.

A BGP speaker MUST include the same FWDR_PREF value on a given route when advertising it to each peer on the same multi-access link.

To calculate the designated forwarder for a given route, the higher degree of preference is preferred.

3. Data Flow Direction - DIRECTION (Type Code XXX):

Multicast routing protocols use a multicast-RIB to construct multicast forwarding trees. Multicast trees may be either uni-directional or bi-directional, source rooted or core rooted. Some multicast routing protocols [PIM-SM,BGMP] can build multiple tree types. The data flow direction attribute may be used to tag a route with a direction for which data is allowed to flow. This may be used for special links (e.g. satellite), or for enforcing policy. The data flow direction attribute is ONLY a route policy attribute. Its use is limited by the multicast routing protocol in use.

Uni-directional trees provide aggressive loop prevention using a Reverse Path Forwarding (RPF) check mechanism whereby a specific incoming interface is chosen to accept packets from a destination. Uni-

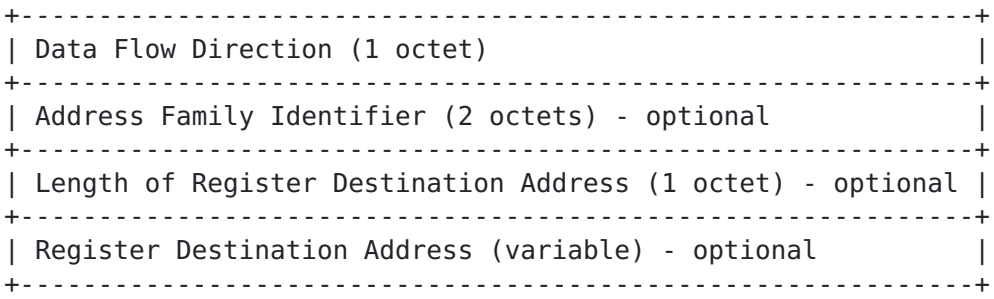
directional trees may provide a coarse grain policy whereby senders may be restricted to only coming from the correct RPF direction. For protocols which support multiple tree types, the data flow direction attribute may be used for tagging a route as uni-directional. This uni-directional route may be used for unidirectional shared trees (i.e. G-RIB route attribute) or for restrictions for non-member senders.

High bandwidth, unidirectional transmission to low cost, receiver-only hardware is becoming an emerging network fabric, e.g. broadcast satellite links or some cable links. Such links can result in bidirectional islands connected via unidirectional links. One common solution to preserving dynamic routing is to add a layer between the network interface and the routing software to emulate bi-directional links through tunnels. However, this can result in non-optimal routing, especially since some routing protocols use forward-paths, while others use reverse-paths.

3.1. DIRECTION Details

DIRECTION is an optional transitive attribute that can be used for the purpose of annotating a route with the direction(s) in which data is allowed to flow. Optionally, for Come-From routes, a register destination may be present. This destination may be used for encapsulating packets when uni-directional trees are constructed.

The DIRECTION attribute contains a 1-octet "Data Flow Direction", optionally followed by three fields identifying a register destination if the data flow direction is Come-From-only. The DIRECTION attribute is type code XXX and is encoded as shown below:



The use and the meaning of these fields are as follows:

Data Flow Direction (DFD): 1 octet

The following values are defined:

1 - Go-To. The route is valid only for traffic travelling towards the origin of the route.

2 - Come-From. The route is valid only for traffic travelling away from the origin of the route. A register destination is only useful for Come-From routes, as the purpose of the register destination is to provide a Go-To channel where none exists otherwise.

3 - Both Go-To and Come-From. The route is valid both for traffic travelling towards and away from the origin of the route.

Address Family Identifier:

This field carries the identity of the Network Layer protocol associated with the Network Address that follows. Presently defined values for this field are specified in RFC1700 (see the Address Family Numbers section). This field MUST NOT be present unless DFD=2.

Length of Next Hop Network Address:

A 1 octet field whose value expresses the length of the "Network Address of Register Destination" field as measured in octets. This field MUST NOT be present unless DFD=2.

Network Address of Next Hop:

A variable length field that contains the Network Address of the router to which encapsulated data packets may be sent. This field MUST NOT be present unless DFD=2.

By default, if the DIRECTION attribute is not present, a route can be assumed to be "Both Go-To and Come-From". That is, the path may be assumed to provide bi-directional connectivity. However, it is the tree construction protocol which will ultimately chose the direction of data flow.

3.2. Direction Usage

When a route is advertised whose next-hop is on the other side of a

unidirectional link from the router to which the route is advertised, it should be annotated with the DIRECTION attribute. This can be used, for example, to provide a directional route over a satellite link with one set of attributes, and a separate (less preferred) route over a bidirectional tunnel. This lets traffic travelling in the direction of the physical link prefer to use that link. Traffic travelling in the opposite direction may prefer a separate path, while still being allowed to traverse the tunnel if desired (e.g. if a bidirectional path is required, and the tunnel is the only bidirectional link available).

Protocols which construct bidirectional trees [CBT,BGMP] allow data to flow in both directions along tree branches maintained with "Join" and "Prune" messages. Such bidirectional branches MUST follow a route which is "Both Go-To and Come-From" (DFD=3).

Several multicast routing protocols [DVMRP,PIMDM,PIMSM] can construct unidirectional source-specific trees by sending "Join" and "Prune" messages along the reverse path towards a source address, with the data flowing in the opposite direction. For this purpose, a Come-From route (DFD=2 or DFD=3) must be used.

Similarly, protocols which construct unidirectional shared-trees [PIMSM] send "Join" and "Prune" messages along the reverse path towards a Rendezvous Point (RP) address, with the data flowing in the opposite direction. Again, a Come-From route (DFD=2 or DFD=3) must be used. Some protocols which normally construct bidirectional trees [BGMP] can construct a unidirectional tree if no DFD=3 route exists.

Finally, shared-tree protocols [PIM-SM,CBT,BGMP] typically support non-member senders by forwarding data towards an address associated with the root (the RP/core's unicast address in PIM-SM and CBT, or the group address in BGMP). Packets from non-member senders may follow a Go-To route (DFD=1 or DFD=3) until they hit a router on the multicast distribution tree.

4. Security Considerations

This extension to BGP does not change the underlying security issues.

5. References

[BGP-4]
Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC

1771, March 1995.

[MBGP]
Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 2283, February 1998.

[RFC2119]
Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

[DVMRP]
Pusateri, T., "Distance vector multicast routing protocol", Internet Draft, March 1998.

[MOSPF]
Moy, J., "Multicast extensions to OSPF", RFC 1584, July 1993.

[PIMSM]
Estrin, D., Farinacci, D., Helmy, A., Thaler, D., Deering, S., Handley, M., Jacobson, V., Liu, C., Sharma, P., and L. Wei. "Protocol independent multicast-sparse mode (PIM-SM): Protocol specification", RFC 2362, June 1998.

[CBT]
Ballardie, A., "Core based trees (CBT version 2) multicast routing: Protocol specification", RFC 2189, September 1997.

[PIMDM]
Estrin, D., Farinacci, D., Helmy, A., Jacobson, V., and L. Wei, "Protocol independent multicast (PIM), dense mode protocol specification", Internet Draft, May 1997.

[MIX]
LaMaster, H., Shultz, S., Meylor, J., and D. Meyer, "Multicast-Friendly Internet Exchange (MIX)", Internet Draft, draft-ietf-mboned-mix-00.txt, November 1998.

6. Author Information

Dave Thaler
Microsoft
One Microsoft Way
Redmond, WA 98052
EMail: dthaler@microsoft.com

Brad Cain
Nortel Networks
3 **Federal Street**
Billerica, MA 01821
EMail: bcain@baynetworks.com

7. Full Copyright Statement