Network Working Group
Internet Draft
Expiration Date: July 2000

Yakov Rekhter
Eric C. Rosen
Cisco Systems, Inc.

January 2000

Carrying Label Information in BGP-4

draft-ietf-mpls-bgp4-mpls-04.txt

Status of this Memo

Abstract

When BGP is used to distribute a particular route, it can be also be used to distribute an MPLS label which is mapped to that route [MPLS-ARCH]. This document specifies the way in which this is done. The label mapping information for a particular route is piggybacked in the same BGP Update message that is used to distribute the route itself.

Table of Contents

1. Specification of Requirements

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.

2. Overview

When BGP is used to distribute a particular route, it can also be used to distribute an MPLS label that is mapped to that route [MPLS-ARCH]. This document specifies the way in which this is done. The label mapping information for a particular route is piggybacked in the same BGP Update message that is used to distribute the route itself.

This can be useful in the following situations:

- If two immediately adjacent Label Switched Routers (LSRs) are also BGP peers, then label distribution can be done without the need for any other label distribution protocol.

- Suppose one's network consists of two "classes" of LSR: exterior LSRs, which interface to other networks, and interior LSRs, which serve only to carry traffic between exterior LSRs. Suppose that the exterior LSRs are BGP speakers. If the BGP speakers distribute MPLS labels to each other along with each route they distribute, then as long as the interior routers support MPLS, they need not receive any of the BGP routes from the BGP

speakers.

If exterior router A needs to send a packet to destination D, and A's BGP next hop for D is exterior router B, and B has mapped label L to D, then A first pushes L onto the packet's label stack. A then consults its IGP to find the next hop to B, call it C. If C has distributed to A an MPLS label for the route to B, A can push this label on the packet's label stack, and then send the packet to C.

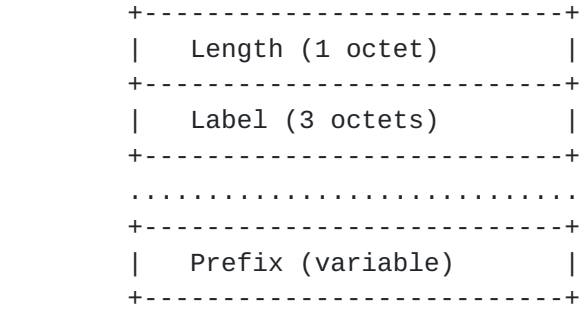
If a set of BGP speakers are exchanging routes via a Route Reflector [BGP-RR], then by piggybacking the label distribution on the route distribution, one is able to use the Route Reflector to distribute the labels as well. This improves scalability quite significantly. Note that if the Route Reflector is not in the forwarding path, it need not even be capable of forwarding MPLS packets.

Label distribution can be piggybacked in the BGP Update message by using the BGP-4 Multiprotocol Extensions attribute [RFC 2283]. The label is encoded into the NLRI field of the attribute, and the SAFI ("Subsequent Address Family Identifier") field is used to indicate that the NLRI contains a label. A BGP speaker may not use BGP to send labels to a particular BGP peer unless that peer indicates, through BGP Capability Negotiation, that it can process Update messages with the specified SAFI field.

3. Carrying Label Mapping Information

Label mapping information is carried as part of the Network Layer Reachability Information (NLRI) in the Multiprotocol Extensions attributes. The AFI indicates, as usual, the address family of the associated route. The fact that the NLRI contains a label is indicated by using SAFI value 4 [assignment to be confirmed by IANA].

The Network Layer Reachability information is encoded as one or more triples of the form <label, length, prefix>, whose fields are described below:



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

a) Length:

The Length field indicates the length in bits of the address prefix plus the label(s).

b) Label:

The Label field carries one or more labels (that corresponds to the stack of labels [MPLS-ENCAPS]). Each label is encoded as 3 octets, where the high-order bit contains "Bottom of Stack" (as defined in [MPLS-ENCAPS]). The following high-order three bits must be zero. The remaining 20 bits contain the label value.

c) Prefix:

The Prefix field contains address prefixes followed by enough trailing bits to make the end of the field fall on an octet boundary. Note that the value of trailing bits is irrelevant.

The label(s) specified for a particular route (and associated with it address prefix) must be assigned by the LSR which is identified by the value of the Next Hop attribute of the route.

When a BGP speaker redistributes a route, the label(s) assigned to that route must not be changed (except by omission), unless the speaker changes the value of the Next Hop attribute of the route.

A BGP speaker can withdraw a previously advertised route (as well as the binding between this route and a label) by either (a) advertising a new route (and a label) with the same NLRI as the previously advertised route, or (b) listing the NLRI of the previously advertised route in the Withdrawn Routes field of an Update message. The label information carried (as part of NLRI) in the Withdrawn Routes field should be set to 0x800000.

4. Advertising Multiple Routes to a Destination

A BGP speaker may maintain (and advertise to its peers) more than one route to a given destination, as long as each such route has its own label(s).

The encoding described above allows a single BGP Update message to carry multiple routes, each with its own label(s).

In the case where a BGP speaker advertises multiple routes to a

destination, if a route is withdrawn, and a label(s) is specified at the time of withdrawal, only the corresponding route with the corresponding label is withdrawn. If a route is withdrawn, and no label is specified at the time of withdrawal, then only the corresponding unlabeled route is withdrawn; the labeled routes are left in place.

5. Capability Negotiation

A BGP speaker that uses Multiprotocol Extensions to carry label mapping information should use the Capabilities Optional Parameter, as defined in [BGP-CAP], to inform its peers about this capability. The MP_EXT Capability Code, as defined in [BGP-MP], is used to negotiate the (AFI, SAFI) pairs available on a particular connection.

A BGP speaker should not advertise this capability to another BGP speaker unless there is a Label Switched Path (LSP) between the two speakers.

A BGP speaker that is capable of handling multiple routes to a destination (as described above) should use the Capabilities Optional Parameter, as defined in [BGP-CAP], to inform its peers about this capability. The value of this capability is TBD.

6. When the BGP Peers are not Directly Adjacent

Consider the following LSR topology: A--B--C--D. Suppose that D distributes a label L to A. In this topology, A cannot simply push L onto a packet's label stack, and then send the resulting packet to B. D must be the only LSR that sees L at the top of the stack. Before A sends the packet to B, it must push on another label, which was distributed by B. B must replace this label with yet another label, which was distributed by C. In other words, there must be an LSP between A and D. If there is no such LSP, A cannot make use of label L. This is true any time labels are distributed between non-adjacent LSRs, whether that distribution is done by BGP or by some other method.

This document does NOT specify any procedure for ensuring in real time that label distribution between non-adjacent LSRs is done only when the appropriate MPLS infrastructure exists in the network or networks connecting the two LSRs. Ensuring that the proper infrastructure exists is an issue for network management and operation.

7. Security Considerations

When an LSR A is directly connected to an LSR B via a point-to-point interface, then when A receives packets over that interface, it knows that they come from B. This makes it easy for A to discard any packets from B whose top labels are not among the labels that A distributed to B. That is, A can easily ensure that B only uses those labels which it is entitled to use. This technique can be used to prevent "label spoofing", i.e., the situation in which an LSR imposes a label which has not been properly distributed to it.

The procedures discussed in this document would commonly be used when the label distribution peers are separated not merely by a point-to-point link, but by an MPLS network. This means that when an LSR A processes a labelled packet, it really has no way to determine which other LSR B pushed on the top label. Hence it cannot tell whether the label is one which B is entitled to use. In fact, when Route Reflectors are in use, A may not even know the set of LSRs which receive its label mappings. So the previous paragraph's technique for preventing label spoofing does not apply.

It is possible though to use other techniques to avoid label spoofing problems. If, for example, one never accepts labeled packets from the network's "external" interfaces, and all the BGP-distributed labels are advertised via IBGP, then there is no way for an untrusted router to put a labeled packet into the network. One can generally assume that one's IBGP peers (or the IBGP peers of one's Route Reflector) will not attempt label spoofing, since they are all under the control of a single administration.

This condition can actually be weakened significantly. One doesn't need to refuse to accept all labeled packets from external interfaces. One just needs to make sure that any labeled packet received on an external interface has a top label which was actually distributed out that interface.

Then a label spoofing problem would only exist if there are both trusted and untrusted systems out the same interface. One way to avoid this problem is simply to avoid this situation.

8. Acknowledgments

Thanks to Ravi Chandra, Enke Chen, Srihari Ramachandra, Eric Gray and Liam Casey for their comments.

9. References

[BGP-4] RFC 1771, "A Border Gateway Protocol 4 (BGP-4)", Y. Rekhter, T. Li, 3/95

[BGP-CAP] "Capabilities Negotiation with BGP-4", R. Chandra, J. , draft-ietf-idr-bgp4-cap-neg-04.txt, 9/99

[BGP-MP] RFC 2283, "Multiprotocol Extensions for BGP-4", T. Bates, R. Chandra, D.Katz, Y. Rekhter, 2/98

[BGP-RR] RFC 1966, "BGP Route Reflection: An alternative to full mesh IBGP", T. Bates, R. Chandra, 6/96.

[MPLS-ARCH] "Multiprotocol Label Switching Architecture"A Proposed Architecture for MPLS", E. Rosen, A. Vishwanathan, R. Callon, draft-ietf-mpls-arch-06.txt, 8/99.

[MPLS-ENCAPS] "MPLS Label Stack Encoding", E. Rosen, et al, draft-ietf-mpls-label-encaps-07.txt, 9/99

10. Author Information

Yakov Rekhter
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
email: yakov@cisco.com

Eric Rosen
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA 01824
email: erosen@cisco.com