

Network Working Group  
Internet Draft  
Expiration Date: January 2004  
Obsoletes [RFC2858](#)

Tony Bates (Cisco Systems)  
Ravi Chandra (Redback Networks)  
Dave Katz (Juniper Networks)  
Yakov Rekhter (Juniper Networks)

## Multiprotocol Extensions for BGP-4

[draft-ietf-idr-rfc2858bis-03.txt](#)

### **1. Status of this Memo**

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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.''

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

### **2. Abstract**

Currently BGP-4 is capable of carrying routing information only for IPv4. This document defines extensions to BGP-4 to enable it to carry routing information for multiple Network Layer protocols (e.g., IPv6, IPX, etc...). The extensions are backward compatible - a router that supports the extensions can interoperate with a router that doesn't support the extensions.

### **3. 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](#) [[RFC2119](#)].

### **4. Overview**

The only three pieces of information carried by BGP-4 [[BGP-4](#)] that are IPv4 specific are (a) the NEXT\_HOP attribute (expressed as an IPv4 address), (b) AGGREGATOR (contains an IPv4 address), and (c) NLRI (expressed as IPv4 address prefixes). This document assumes that any BGP speaker (including the one that supports multiprotocol capabilities defined in this document) has to have an IPv4 address (which will be used, among other things, in the AGGREGATOR attribute). Therefore, to enable BGP-4 to support routing for multiple Network Layer protocols the only two things that have to be added to BGP-4 are (a) the ability to associate a particular Network Layer protocol with the next hop information, and (b) the ability to associate a particular Network Layer protocol with NLRI. To identify individual Network Layer protocols associated with the next hop information and semantics of NLRI this document uses a combination of Address Family, as defined in [[RFC1700](#)], and Subsequent Address Family (as described in this document).

One could further observe that the next hop information (the information provided by the NEXT\_HOP attribute) is meaningful (and necessary) only in conjunction with the advertisements of reachable destinations - in conjunction with the advertisements of unreachable destinations (withdrawing routes from service) the next hop information is meaningless. This suggests that the advertisement of reachable destinations should be grouped with the advertisement of the next hop to be used for these destinations, and that the advertisement of reachable destinations should be segregated from the advertisement of unreachable destinations.

To provide backward compatibility, as well as to simplify introduction of the multiprotocol capabilities into BGP-4 this document uses two new attributes, Multiprotocol Reachable NLRI (MP\_REACH\_NLRI), and Multiprotocol Unreachable NLRI (MP\_UNREACH\_NLRI). The first one (MP\_REACH\_NLRI) is used to carry the set of reachable destinations together with the next hop information to be used for forwarding to these destinations. The second one (MP\_UNREACH\_NLRI) is used to carry the set of unreachable destinations. Both of these attributes are optional and non-transitive. This way a BGP speaker that doesn't support the multiprotocol capabilities will just ignore the information carried



in these attributes, and will not pass it to other BGP speakers.

## 5. Multiprotocol Reachable NLRI - MP\_REACH\_NLRI (Type Code 14):

This is an optional non-transitive attribute that can be used for the following purposes:

- (a) to advertise a feasible route to a peer
- (b) to permit a router to advertise the Network Layer address of the router that should be used as the next hop to the destinations listed in the Network Layer Reachability Information field of the MP\_NLRI attribute.
- (c) to allow a given router to report some or all of the Subnetwork Points of Attachment (SNPAs) that exist within the local system

The attribute is encoded as shown below:

```

+-----+
| Address Family Identifier (2 octets) |
+-----+
| Subsequent Address Family Identifier (1 octet) |
+-----+
| Length of Next Hop Network Address (1 octet) |
+-----+
| Network Address of Next Hop (variable) |
+-----+
| Number of SNPAs (1 octet) |
+-----+
| Length of first SNPA(1 octet) |
+-----+
| First SNPA (variable) |
+-----+
| Length of second SNPA (1 octet) |
+-----+
| Second SNPA (variable) |
+-----+
| ... |
+-----+
| Length of Last SNPA (1 octet) |
+-----+
| Last SNPA (variable) |
+-----+
| Network Layer Reachability Information (variable) |

```



+-----+

The use and meaning of these fields are as follows:

Address Family Identifier (AFI):

This field in combination with the Subsequent Address Family Identifier field identifies the Network Layer protocol associated with the Network Address of Next Hop and the semantics of the Network Layer Reachability Information that follows.

Presently defined values for the Address Family Identifier field are specified in [RFC1700](#) (see the Address Family Numbers section).

Subsequent Address Family Identifier (SAFI):

This field in combination with the Address Family Identifier field identifies the Network Layer protocol associated with the Network Address of the Next Hop and the semantics of the Network Layer Reachability Information that follows.

Length of Next Hop Network Address:

A 1 octet field whose value expresses the length of the "Network Address of Next Hop" field as measured in octets.

Network Address of Next Hop:

A variable length field that contains the Network Address of the next router on the path to the destination system. The Network Layer protocol associated with the Network Address of the Next Hop is identified by a combination of <AFI, SAFI> carried in the attribute.

Number of SNPAs:

A 1 octet field which contains the number of distinct SNPAs to be listed in the following fields. The value 0 SHALL be used to indicate that no SNPAs are listed in this attribute.

Length of Nth SNPA:

A 1 octet field whose value expresses the length of the "Nth SNPA of Next Hop" field as measured in semi-octets



#### Nth SNPA of Next Hop:

A variable length field that contains an SNPA of the router whose Network Address is contained in the "Network Address of Next Hop" field. The field length is an integral number of octets in length, namely the rounded-up integer value of one half the SNPA length expressed in semi-octets; if the SNPA contains an odd number of semi-octets, a value in this field will be padded with a trailing all-zero semi-octet.

#### Network Layer Reachability Information (NLRI):

A variable length field that lists NLRI for the feasible routes that are being advertised in this attribute. The semantics of NLRI is identified by a combination of <AFI, SAFI> carried in the attribute.

When the Subsequent Address Family Identifier field is set to one of the values defined in this document, each NLRI is encoded as specified in the "NLRI encoding" section of this document.

The next hop information carried in the MP\_REACH\_NLRI path attribute defines the Network Layer address of the router that should be used as the next hop to the destinations listed in the MP\_NLRI attribute in the UPDATE message.

The rules for the next hop information are the same as the rules for the information carried in the NEXT\_HOP BGP attribute (see [Section 5.1.3](#) of [[BGP-4](#)]).

An UPDATE message that carries the MP\_REACH\_NLRI MUST also carry the ORIGIN and the AS\_PATH attributes (both in EBGp and in IBGP exchanges). Moreover, in IBGP exchanges such a message MUST also carry the LOCAL\_PREF attribute.

An UPDATE message that carries no NLRI, other than the one encoded in the MP\_REACH\_NLRI attribute, SHOULD NOT carry the NEXT\_HOP attribute. If such a message contains the NEXT\_HOP attribute, the BGP speaker that receives the message SHOULD ignore this attribute.

An UPDATE message SHOULD NOT include the same address prefix (of the same <AFI, SAFI>) in more than one of the following fields: WITHDRAWN ROUTES field, Network Reachability Information fields, MP\_REACH\_NLRI field, and MP\_UNREACH\_NLRI field. The processing of an UPDATE message in this form is un-defined.





**6. Multiprotocol Unreachable NLRI - MP\_UNREACH\_NLRI (Type Code 15):**

This is an optional non-transitive attribute that can be used for the purpose of withdrawing multiple unfeasible routes from service.

The attribute is encoded as shown below:

```

+-----+
| Address Family Identifier (2 octets)          |
+-----+
| Subsequent Address Family Identifier (1 octet) |
+-----+
| Withdrawn Routes (variable)                  |
+-----+
```

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

**Address Family Identifier (AFI):**

This field in combination with the Subsequent Address Family Identifier field identifies the semantics associated with the Network Layer Reachability Information (NLRI) that follows.

Presently defined values for the Address Family Identifier field are specified in [RFC1700](#) (see the Address Family Numbers section).

**Subsequent Address Family Identifier (SAFI):**

This field in combination with the Address Family Identifier field identifies the semantics associated with the Network Layer Reachability Information (NLRI) that follows.

**Withdrawn Routes Network Layer Reachability Information:**

A variable length field that lists NLRI for the routes that are being withdrawn from service. The semantics of NLRI is identified by a combination of <AFI, SAFI> carried in the attribute.

When the Subsequent Address Family Identifier field is set to one of the values defined in this document, each NLRI is encoded as specified in the "NLRI encoding" section of this document.

An UPDATE message that contains the MP\_UNREACH\_NLRI is not required



to carry any other path attributes.

## **7. NLRI encoding**

The Network Layer Reachability information is encoded as one or more 2-tuples of the form <length, prefix>, whose fields are described below:

```
+-----+
| Length (1 octet)      |
+-----+
| Prefix (variable)     |
+-----+
```

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. A length of zero indicates a prefix that matches all (as specified by the address family) addresses (with prefix, itself, of zero octets).

### **b) Prefix:**

The Prefix field contains an address prefix 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.

## **8. Subsequent Address Family Identifier**

This document defines the following values for the Subsequent Address Family Identifier field carried in the MP\_REACH\_NLRI and MP\_UNREACH\_NLRI attributes:

1 - Network Layer Reachability Information used for unicast forwarding

2 - Network Layer Reachability Information used for multicast forwarding

An implementation MAY support all, some, or none of the Subsequent



Address Family Identifier values defined in this document.

## 9. Error Handling

If a BGP speaker receives from a neighbor an Update message that contains the MP\_REACH\_NLRI or MP\_UNREACH\_NLRI attribute, and the speaker determines that the attribute is incorrect, the speaker MUST delete all the BGP routes received from that neighbor whose AFI/SAFI is the same as the one carried in the incorrect MP\_REACH\_NLRI or MP\_UNREACH\_NLRI attribute. For the duration of the BGP session over which the Update message was received, the speaker then SHOULD ignore all the subsequent routes with that AFI/SAFI received over that session.

In addition, the speaker MAY terminate the BGP session over which the Update message was received. The session SHOULD be terminated with the Notification message code/subcode indicating "Update Message Error"/"Optional Attribute Error".

## 10. Use of BGP Capability Advertisement

A BGP speaker that uses Multiprotocol Extensions SHOULD use the Capability Advertisement procedures [[BGP-CAP](#)] to determine whether the speaker could use Multiprotocol Extensions with a particular peer.

The fields in the Capabilities Optional Parameter are set as follows. The Capability Code field is set to 1 (which indicates Multiprotocol Extensions capabilities). The Capability Length field is set to 4. The Capability Value field is defined as:

0	7	15	23	31
+-----+-----+-----+-----+				
	AFI		Res.	SAFI
+-----+-----+-----+-----+				

The use and meaning of this field is as follow:

AFI - Address Family Identifier (16 bit), encoded the same way as in the Multiprotocol Extensions

Res. - Reserved (8 bit) field. Should be set to 0 by the sender and ignored by the receiver.

SAFI - Subsequent Address Family Identifier (8 bit), encoded



the same way as in the Multiprotocol Extensions.

A speaker that supports multiple <AFI, SAFI> tuples includes them as multiple Capabilities in the Capabilities Optional Parameter.

To have a bi-directional exchange of routing information for a particular <AFI, SAFI> between a pair of BGP speakers, each such speaker MUST advertise to the other (via the Capability Advertisement mechanism) the capability to support that particular <AFI, SAFI> routes.

## **11. IANA Considerations**

As specified in this document, the MP\_REACH\_NLRI and MP\_UNREACH\_NLRI attributes contain the Subsequence Address Family Identifier (SAFI) field. The SAFI name space is defined in this document. The IANA will maintain and register values for the SAFI namespace as follows. SAFI value 0 is reserved. SAFI values 1 and 2 are assigned in this document. SAFI values 4 through 63 are to be assigned by IANA using the "IETF Consensus" policy defined in [RFC2434](#). SAFI values 64 through 127 are to be assigned by IANA, using the "First Come First Served" policy defined in [RFC2434](#). SAFI values 128 through 255 are for "private use", and values in this range are not to be assigned by IANA.

## **12. Comparison with [RFC2858](#)**

This document makes the use of the next hop information consistent with the information carried in the NEXT\_HOP BGP path attribute.

This document removes the definition of SAFI = 3.

## **13. Comparison with [RFC2283](#)**

This document restricts the MP\_REACH\_NLRI attribute to carry only a single instance of <AFI, SAFI, Next Hop Information, ...>.

This document restricts the MP\_UNREACH\_NLRI attribute to carry only a single instance of <AFI, SAFI, ...>.

This document clarifies handling of an UPDATE message that carries no NLRI, other than the one encoded in the MP\_REACH\_NLRI attribute.

This document clarifies error handling in the presence of MP\_REACH\_NLRI or MP\_UNREACH\_NLRI attributes.





This document specifies the use of BGP Capabilities Advertisements in conjunction with Multi-protocol extensions.

Finally, this document includes the "IANA Consideration" Section.

#### **14. Security Considerations**

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

#### **15. Acknowledgements**

The authors would like to thank members of the IDR Working Group for their review and comments.

#### **16. References**

[BGP-CAP] "Capabilities Advertisement with BGP-4", R. Chandra, J. Scudder, [RFC2842](http://www.rfc-editor.org/rfc/rfc2842.txt), May 2000

[BGP-4] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", [RFC 1771](http://www.rfc-editor.org/rfc/rfc1771.txt), March 1995.

[RFC1700] "Assigned Numbers", J. Reynolds, J. Postel, [RFC1700](http://www.rfc-editor.org/rfc/rfc1700.txt), October 1994 (see also <http://www.iana.org/iana/assignments.html>)

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](http://www.rfc-editor.org/rfc/rfc2119.txt), [RFC 2119](http://www.rfc-editor.org/rfc/rfc2119.txt), March 1997.

#### **17. Author Information**

Tony Bates  
Cisco Systems, Inc.  
email: [tbates@cisco.com](mailto:tbates@cisco.com)

Ravi Chandra  
Redback Networks  
e-mail: [rchandra@redback.com](mailto:rchandra@redback.com)

Dave Katz  
Juniper Networks, Inc.  
email: [dkatz@juniper.com](mailto:dkatz@juniper.com)



Yakov Rekhter  
Juniper Networks, Inc.  
email: yakov@juniper.com