

Internet Engineering Task Force
Internet Draft

[draft-ward-bgp4-ibb-00.txt](#)

David Ward
Internet Engineering
Group, LLC

John Scudder
Internet Engineering
Group, LLC
June, 1999

BGP Notification Cease: I'll Be Back

<[draft-ward-bgp4-ibb-00.txt](#)>

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/lid-abstracts.txt>

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

1. Abstract

Many recent router architectures decouple the routing engine from the forwarding engine, so that packet forwarding can continue even if routing software is not active. The current definition of the BGP protocol does not support this. We propose a new variety of CEASE NOTIFICATION message (IBB) which indicates to a peer that the router sending the notification expects to be able to continue forwarding traffic for a certain period of time without an established BGP peering session. We also propose a new OPEN message (ICB) that if received during the HOLDDTIME period, does not require conventional reestablishment of the BGP peering session. These capabilities are useful for orderly and non-intrusive routing

software updates, operating system updates, AS number migration, redundancy and catastrophic event handling.

Ward, Scudder Internet Draft June 1999

page 1

<[draft-ward-bgp4-ibb-00.txt](#)>

June, 1999

2. 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 [RFC-2119](#).

3. Introduction

Goals:

- a. Continued forwarding in the absence of an Established BGP peering session
- b. Traffic shall continue to flow over the preferred path which would be used if the BGP speaker had not closed the session
- c. Routes will not be flapped.

Applications:

- a. Support minimally intrusive upgrade of routing software, operating system, hardware, etc.
- b. Support minimally intrusive AS, IP, interface, etc. renumbering
- c. Support minimally intrusive catastrophic software events

4. Operation

IBB introduces a new OPEN option, a new CEASE NOTIFICATION option, and a new Capabilities Negotiation [[BGP-CAP](#)] option.

BGP operation is modified as follows:

4.1. Capability Negotiation

IBB must be negotiated at session startup time using Capability Negotiation. (See [Section 5](#) for discussion of why this is necessary.)

The capability encoding for IBB is as follows:

Capability Code: TBD (1 octet)

Capability Length: 6 (1 octet)

Capability Value:

Flags: reserved, must be transmitted as zero (2 octets)

Maximum IBB timeout in seconds: (2 octets unsigned)

Maximum route refresh timeout in seconds: (2 octets)

unsigned)

The IBB and route refresh timeouts specify the maximum timeout values the BGP speaker is willing to accept. The maximum timeout values are a matter of local configuration. 360 seconds is suggested as a reasonable default value for both maxima. The actual timeouts which will be used are based on the timeouts proposed in the IBB CEASE and ICB OPEN; see below.

Ward, Scudder Internet Draft June 1999

page 2

[<draft-ward-bgp4-ibb-00.txt>](#)

June, 1999

4.2. Closing a Session With IBB CEASE

After IBB has been successfully negotiated, if a BGP speaker wants to temporarily disconnect the session but is capable of continuing to forward packets, it MAY close the session using a special CEASE NOTIFICATION message called the `_I'll be back_`, or IBB CEASE. The IBB CEASE adds the following option to the standard CEASE NOTIFICATION message:

Error code = 6 (Cease) (one octet)
Error subcode = 1 (IBB) (one octet)
Flags = Reserved, must be sent as zero (two octets unsigned)
Data0 = IBB timeout in seconds (two octets unsigned)
Data1 = not used (two octets unsigned)

The semantics of the IBB CEASE are that the sender,

- a. Will attempt to reestablish the session prior to the expiration of the IBB timeout, and
- b. Will be able to continue forwarding packets in the interim.

A BGP speaker MUST NOT send an IBB CEASE unless these criteria are met. It MUST be possible for a router administrator to cause a BGP session to be closed with a conventional CEASE instead of an IBB CEASE.

When a BGP speaker has multiple IBGP peers to which it will send an IBB CEASE, it MUST NOT set the IBB timeout as a value greater than the minimum of all maximum IBB timeout values negotiated by the IBGP peers. A BGP speaker MUST NOT send an IBB CEASE to any IBGP peer unless all IBGP peers have successfully negotiated the IBB option. (See [Section 5](#) for discussion of why this is necessary, and for a discussion of special considerations for route reflectors.)

The IBB timeout selected SHOULD NOT greatly exceed the time needed for the BGP speaker to re-initiate its BGP connections; i.e. it has the sense of a `_reboot time_`. It MUST NOT exceed the maximum value established by the peer during capability negotiation. (There are

further restrictions for IBGP peers; see previous paragraph.)

Upon receiving the IBB CEASE, the connection to the peer which sent the CEASE should be closed, just as with a normal CEASE. However, in place of marking the routes from the peer as invalid, as specified in [section 6](#) of the BGP specification [[BGP-4](#)], the routes are scheduled for later cleanup as follows:

- a. Create a timer scheduled to expire at the lesser of the IBB timeout received in the CEASE and the locally-configured maximum. If the received IBB timeout exceeds the locally-configured maximum, an error SHOULD be logged.
- b. Mark the routes from the peer which sent the CEASE to be deleted when the timer expires.

Ward, Scudder Internet Draft June 1999

page 3

<[draft-ward-bgp4-ibb-00.txt](#)>

June, 1999

- c. If the IBB timeout expires, delete all marked routes immediately.
- d. If a new session is opened with the peer without the ICB option (see below) being used, or if a session is attempted but fails (i.e., an error is detected before the session enters ESTABLISHED state) delete all marked routes immediately, and cancel the timer.

[4.3](#). Opening a Session With OPEN ICB

When a peer which sent an IBB CEASE wishes to establish a new session, it must do so by negotiating IBB as specified in [section 4.1](#), with the addition of the `_I Came Back_` (or ICB) OPEN parameter, which is encoded as follows:

 Parm. Type: TBD (one octet)
 Parm. Length: 3 (one octet)
 Parm. Value: Routerrefresh timeout in seconds (two octets unsigned)
 Flags: Reserved, must be sent as zero (one octet unsigned)

An OPEN carrying the ICB parameter is known as an ICB OPEN. The semantics of the ICB OPEN are that the sender,

- a. Previously sent an IBB CEASE, or terminated the previous session without sending a CEASE (e.g., due to a crash),
- b. Has preserved the forwarding table it had prior to sending the preceding IBB CEASE (the `_old forwarding table_`), and
- c. Will not remove any NLRI from the old forwarding table prior to the expiration of the route refresh timeout. (Note that it MAY update the NLRI, however.)

A BGP speaker MUST NOT send an ICB OPEN unless these criteria are met. A BGP speaker SHOULD NOT send an IBGP peer a route refresh timeout value which exceeds the minimum of the previously-negotiated route refresh timeouts for all IBGP peers. Note that this MAY require writing route refresh timeout values to stable storage as they are negotiated. (See [Section 5](#) for discussion of why this is advisable.)

The route refresh timeout value should be selected such that routing will typically have reconverged prior to its expiration. The exact means of selecting the value are implementation-specific, but MAY include manual configuration or heuristics based on the size of the Loc-RIB prior to session restart. 180 seconds MAY be used as a reasonable default value.

When an ICB OPEN is received:

- a. If there is a pending IBB timer, the timer is rescheduled to expire at the lesser of the route refresh timeout and the locally-configured maximum.

Ward, Scudder Internet Draft June 1999

page 4

[<draft-ward-bgp4-ibb-00.txt>](#)

June, 1999

- b. If there is not a pending IBB timer, but there is already a session in ESTABLISHED state with the peer from which the ICB OPEN was received, and if that session had negotiated IBB, then the ESTABLISHED session should be terminated immediately, as if an IBB CEASE had been received. (The effect will be to create a timer with a timeout value as given in (a), and to enqueue the peer's routes on that timer.) This rule provides for, e.g., non-intrusive transition from a primary to a backup route processor in the event of the failure of the primary in a router with redundant route processors.

If a BGP session is begun with a peer whose previous session terminated with an IBB CEASE, if the new session does not begin with an ICB OPEN, then the pending IBB timer should immediately be expired, i.e. the peer's old routes should immediately be flushed. Likewise, if a session is begun which terminates with an error (i.e., a condition which causes the connection to be terminated with a NOTIFICATION code other than CEASE) before reaching ESTABLISHED state, the peer's old routes should be flushed.

Under normal circumstances, the connection to the peer should be re-established in less than the IBB timeout period. When new routes are received from the peer, they may either depict wholly new NLRI (in which case they are added to the Adj-RIB-In as per the BGP specification) or they may depict NLRI which are already present in the Adj-RIB-In waiting on the deletion timer. In this

case, the marked route is replaced by the refreshing route. Such routes are said to have been refreshed, and are no longer candidates for deletion when the route refresh timer expires.

A `_previous session_` as discussed in this section is defined as a session with a BGP speaker whose IP address is the same as the IP address of the new session. Note that router ID SHOULD NOT be used to determine if a session is the `_previous session_`; this facilitates using IBB to non-intrusively change the router ID of a BGP speaker.

[4.4. Route Reflectors](#)

Note that it is only necessary that all direct IBGP peers of the BGP speaker support IBB, not all IBGP speakers in the routing domain if route reflection is in use. If route reflection is in use, then if an IBB cease is sent to a reflector which implements IBB, then the reflector simply won't propagate withdrawals until the timeout period expires.

The reflector itself is a special case. It MAY send an IBB notify to any subset of peers which all support IBB -- that is, if all the reflector's clients support IBB, an IBB cease MAY be sent to all the clients. If all the regular peers support IBB, an IBB cease MAY be sent to those peers.

Ward, Scudder Internet Draft June 1999

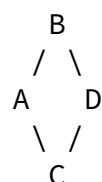
page 5

[<draft-ward-bgp4-ibb-00.txt>](#)

June, 1999

[5. Deployment](#)

The IBB cease may be used with external BGP peers with impunity. In the IBGP case, it's only safe to use IBB if all IBGP neighbors of the BGP speaker understand the IBB cease. To understand why this is the case, consider the following topology:



The topology is fully IBGP meshed; the diagram shows physical topology.

- A injects prefix X with Localpref 200
- B injects prefix X with Localpref 100
- A and D support IBB
- B and C do not

Internet Engineering Group, LLC
122 South Main Street, Suite 280
Ann Arbor, MI 48104
dward@ieng.com

John Scudder
Internet Engineering Group, LLC
122 South Main Street, Suite 280
Ann Arbor, MI 48104
jgs@ieng.com

Ward, Scudder Internet Draft June 1999

page 7

[<draft-ward-bgp4-ibb-00.txt>](#)

June, 1999

Full Copyright Statement

"Copyright (C) The Internet Society (date). All Rights Reserved.
This document and translations of it may be copied and furnished to
others, and derivative works that comment on or otherwise explain
it or assist in its implementation may be prepared, copied,

published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE."