IDR Working Group
Internet-Draft
Expires: August 27, 2008

S. Hares
Green Hills Software
P. Bose
Lockheed
February 24, 2008

Dynamic AS Re-Association At Confederation Edge
draft-hares-asconfed-edge-03.txt

Status of this Memo

Copyright Notice

Abstract

This document provides a mechanism for Autonomous Systems within an AS Confederation to survive the disconnection to other AS within the AS confederation without dropping peers. When all links to the other AS in the Confederation break, this mechanism allows the AS to revert to local AS to continue communication with E-BGP peers. This mechanism has two parts: Capability signaling between the two parties at connection start to save two AS (internal and AS Confederation AS) and a mechanism to signal the switch between AS Confederation AS and internal AS.

Table of Contents

1. Introduction

This mechanism provides a mechanism for an Autonomous System within an AS confederation to survive disconnection from the rest of the Autonomous Systems within the AS Confederation. When an AS is connected to the rest of an AS confederation, it acts as a single AS. If all links between the AS to other members of the AS confederation are broken, the AS Confederation is broken in two (or more) parts, and the individual sub-Autonomous Systems (sub-AS-es) within the confederation may need to "back off" to their local AS number to restore connectivity through some external path.

If a router along the edge of an AS determines the sub-AS has lost its connection to the remainder of the confederation AS, it will need to change the AS number with which it is peering to eBGP peers. This restart of all EBGP connections can be onerous for the AS that has broken away from the AS Confederation. This draft provides a mechanism for the AS within the AS confederation to use a pre-agreed upon fail-over to the internal AS, so its eBGP connections will not be reset.

Upon return of the AS Confederation links, this mechanism can signal the Edge AS returning to the AS Confederation.

2. 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 RFC 2119 [1].

3. Mechanism overview for Dynamic AS Confederation Edge Re-association

The mechanism has parts: Open Capability and Dynamic Capability.

1) An Dynamic AS-Confederation Edge Open capability

The AS-Confederation capability signals the ability to fail-over upon "AS confederation disconnect" by changing the local AS number without resetting the eBGP peering session.

The format of the ASConfed-Edge capability is described in section 4 and contains the AS of the Confederation and a list of Internal AS that the BGP peer will back off to. This capability also indicates the mechanism by which the node will signal the switch via the dynamic capabilities.

Note: The detection of the "AS confederation disconnect" is a locally determined feature that includes (but is not limited to): determining that all AS Confederation BGP peers are disconnected from this peer.

2) Dynamic Capablity - Signaling AS Fail-over and Restoration

2.1)Signaling the AS Fail-over to Internal AS

Signaling an AS fail-over is done via a dynamic capability with the ASConfed_Edge capability with AS flag on. The ASConfed_Edge dynamic capability requires a dynamic capability with an "ACK" be sent to the originating BGP peer. Whether the "ACK" signals a success or failure, the full ASConfed_Edge capability must be sent.

The AS peer wishing to switch from an AS confederation to an internal AS, signals this to a peer by sending a ASConfed_Edge dynamic capability with the AS-in-Use flag set to the internal AS.

Upon receiving this dynamic capability the BGP speaker associated with the AS-Confederation Edge switches from the AS confederation to the AS number specified for the session to the internal session.

After switching to the new AS:

-All checking of the local AS in BGP packets utilizes the new AS.

-All new routes will be announced with the new AS number.

-All older routes will be re-announced based on the AS resend flag.

2.2) Signaling a AS restoring to AS confederations

When the AS Confederations links are re-established, the BGP speaker on the AS Confederation sends a Dynamic Capability with the ASConfed_Edge Capability with "AS in Use flag" set AS Confederation flag. All AS checking for the local BGP speaker reverts to the original Confederation AS.

After switching back to the AS Confederation:

-All Checking of the local AS in BGP packets utilizes the AS Confederation number.

-All new routes will be announced with the AS confederation number

All older routes will be re-announced with the AS confederation number based on the AS resend flag.

4. BGP Peer mechanisms

The mechanism has two parts:

1) Negotiate Open Capability with AS Confederation Edge

The ASConfed-Edge capability signals that a BGP peer ability to use the Dynamic AS Dynamic-Capability exchange on an AS Confederation Edge.

The format of the ASConfed-Edge capability is described in section 5 and contains a list of Autonomous systems that the BGP peer may re-associated to. This capability also indicates the mechanism by which the node will signal the switch is the dynamic capabilities message. The resend flags indicated in the type of resend flags the peer will support.

Within an AS, any BGP peer that will send an ASConfed-Edge Capability to an Exterior peer MUST send the ASConfed-Edge capability to all IBGP peers. Only if all IBGP Peers successfully negotiated the ASConfed-Edge capability, can any BGP peer dynamically switch over from the AS Confederation to an internal AS using this mechanism.

2) Signaling the Dynamic AS Confederation Switch -Originator

Signaling a Dynamic Switch is done via the Dynamic Capability message DYN-CAP [3] with the Dynamic AS capability (format in section 6).

The BGP peer decides to initiate the AS switch over by using local policy and implementation specific mechanisms. To signal the Dynamic AS switch over, the initiating BGP peer has two steps.

Step 1: Dynamic AS Confederation AS change to all IBGP peers

Upon receiving a "Dynamic AS Confederation change" indication to the BGP process, the BGP process will send to IBGP peers a dynamic capability message with Dynamic AS Confed_Edge capability with flags of:

Source flag to IBGP

AS use set to either InternalAS /AS Confederation,

Reserve/Start set to reserve,

Success flag set to Success (0) and,

Resend flags set to peers desired resend capability

If the Dynamic AS Confederation is switching from the AS Confederation to the internal peer, the flag is set to internal peer. If peer is requesting a switch from the internal AS to the AS Confederation, the flag is set to AS Confederation

Upon receiving the ACK from all IBGP peers for the Dynamic AS Capability,the BGP peer will send a capability message with the AS Confed_EDGE capability with flags set to:

source flag to IBGP,

the AS use set to InternalAS/AS Confederation,

Reserve/Start set to Start,

Success/Failure byte - set to Success, and

resend flags set to peers desired resend capability.

Upon receiving an ACK from all IBGP peers for the dynamic AS Confed_EDGE capability with start step 2.

In case of the receiving IBGP peer's local BGP implementation detecting a failure to switch to a new AS when it receives the dynamic capability with the "Reserve" flag set, Capability will be signaled with a "failure" flag. This failure will halt the originating Peer switch to the new AS. The failure is signaled by responding to the Dynamic ASConfed_Edge dynamic capability with the following bits set:

source flag set to IBGP

AS use flags set to internal/AS confederation,

Reserve/Start set to reserve,

resend flags set to original resend

If the IBGP peers are part of a Route-Reflection hierarchy, a Route Reflector MUST wait to send an ack to the Dynamic AS change after it has signaled all of its clients and all of it total mesh peers. In this way, when the initiating IBGP peer receives the Dynamic Capability ACK, the rest of the IBGP peer has been informed.

Note: that the Dynamic Capability ACK may pass back success or failure. A failure anywhere in an IBGP cloud will not allow the BGP to progress to step 2. Each AS Confed_Edge capability MUST be responded to with an dynamic capability with an ACK.

Step 2: Send all EBGP peers the Dynamic AS Confederation Change

Each EBGP peers signal the Dynamic AS change to its neighbor with ASConfed_EDGE Dynamic Capability. When sent to an EBGP peer, the source flag is set to "EBGP" flag.

Upon receiving this dynamic capability from an E-BGP peer, the BGP speaker processes the switch of the peer from the current AS number to the one specified in the capability. This change in processing includes:

-All checking of the local AS in BGP packets utilizes the new AS.

-All new routes will be announced with the new AS number.

-All older routes will be re-announced based on the AS resend flag.

Upon successful processing, the EBGP peer will acknowledge the Dynamic AS capability by sending a Dynamic Capability with an Ack in the Dynamic Capability flag and an "ack-succeess" indication in the Dynamic Capabiity flag word.

3) E-BGP Peer Receiving a Signaling the Dynamic AS Switch-over

Upon receiving a "Dynamic AS Change" Dynamic capability from an EBGP peer, the EBGP peer will follow 2 steps:

Step 1: Upon receiving a Dynamic AS Confed_Edge capability the BGP process will send to all IBGP peers the Dynamic AS Confed_Edge dynamic capability with an the as switch over flags.

Upon receiving the Dynamic Capability with the "Ack" bit set from all IBGP peers for the Dynamic ASConfed_EDGE Capability, the BGP peer can start step 2.

Again, if the IBGP peers of the receiving BGP are part of a Route-Reflection hierarchy, a Route Reflector MUST only send an ACK to the Dynamic AS Confed_Edge change after it has successfully sent the Dynamic Capability to its clients and all of it total mesh peers. In this way, when the initiating IBGP peer receives the Dynamic capability ACK, the rest of the IBGP

peer has been informed.

In case of errors in resetting the Dynamic AS capability, the receiving IBGP peer can set the "Failure" flag in the Dynamic capability that is being ACK. Any failures will be signaled to the originating AS, and the Dynamic AS switch terminated.

Termination of the AS_confederation switch can be done by deleting the Dynamic AS_Confederation capability

Step 2: Respond to E-BGP AS with Dynamic AS change

The E-BGP peer responds to the originated AS with a Dynamic AS change with an EBGP flag set and the Failure bit off.

Upon receiving this dynamic capability from an E-BGP peer, the BGP speaker associated with the AS Edge process the switch of the peer from the current AS number to the one specified in the capability. This switch includes:

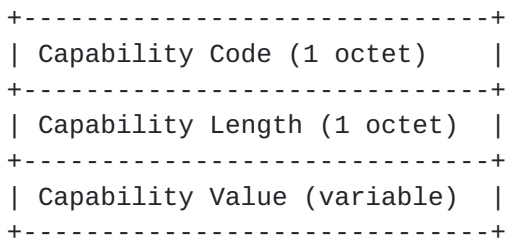
-All checking of the local AS in BGP packets utilizes the new AS.

-All new routes will be announced with the new AS number n

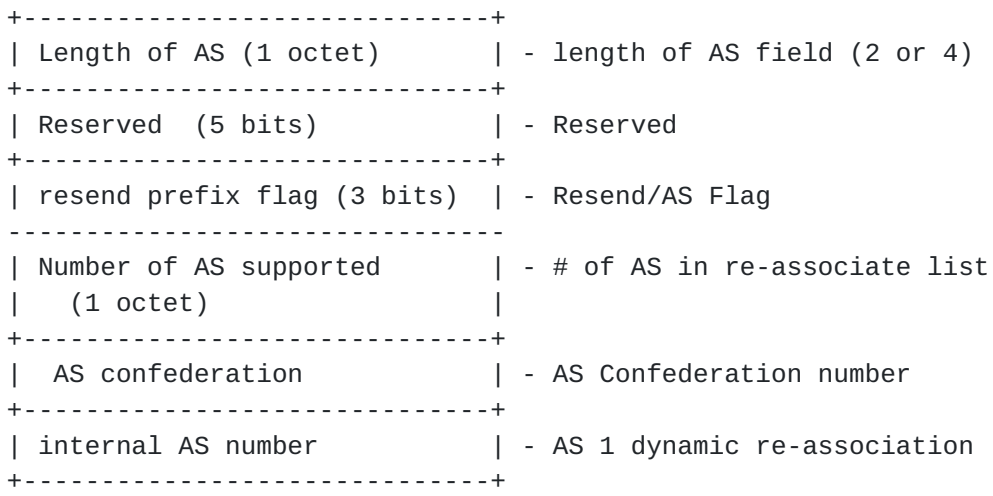
-All older routes will be re-announced based on the AS resend flag.

5. Open Dynamic AS Capability

RFC 3992 [2] describes the open capability mechanisms. This document describes a new Capability: Dynamic ASConfeder_EDGE.



Where the Capability value is:



Dynamic AS Confederation Edge Open Capability Bytes

The resend prefix flag indicates when the AS will resend the routes with the new AS. The flag values are set as a bit pattern to indicate that

0x00 - Resend routes based on local timer (in bataches)

0x01 - Resend routes immediately

0x02 - Don't resend routes (leave with old AS confederation)

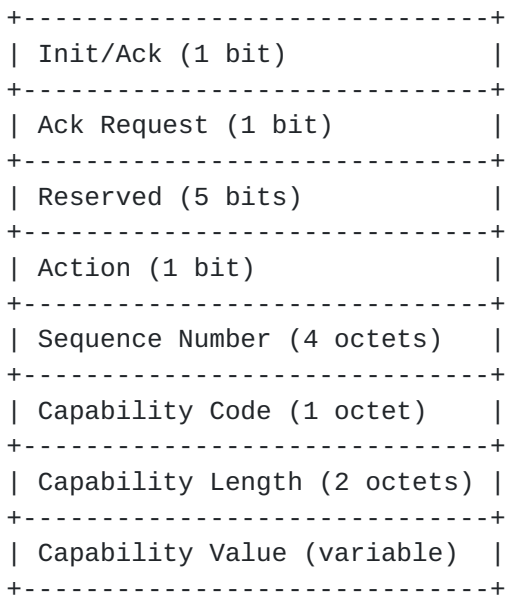
The number of AS supported field gives the number of the Autonomous Systems fin the dynamic re-association list.The Autonomous Systems in the AS list are the list of ASes that this peer may switch to in when

dynamically re-association from the original AS to a new AS.

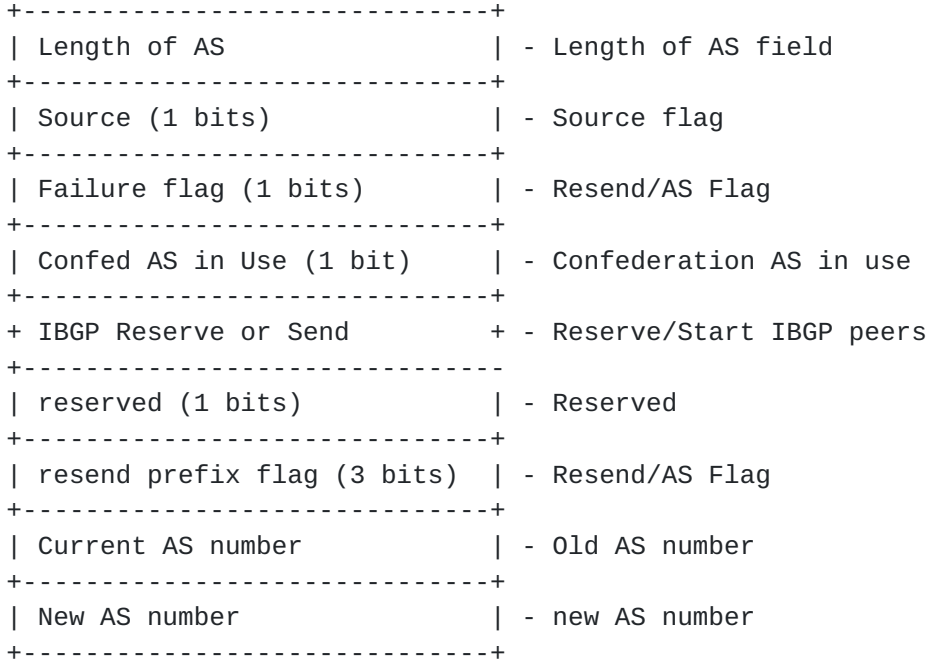
Each side of the peer will send a list of Autonomous Systems that it will dynamic re-associate with. Upon start-up the re-associations list can be check by policy to determine that each side can support the required re-associations.

6. Capability Message

This BGP dynamic capability uses the new BGP Dynamic Capability DYN-CAP [3] format of:



The capability value is:



Dynamic AS Confed Edge Dynamic Capability Bytes

bit definitions:

Source Flag flag (1 bits)

0 - node originated

1 - EBGP peer originated

Failiure Flag (1 bit)

1 - failure

0 - success

AS in USE:

0x0 - Internal AS number

0x1 - AS Confederation number

IBGP Reserve or Start

0x0 - Reserve IBGP peers only, Do not start EBGP peers negotiation.

0x1 - Start IBGP peers and Start EBGP negotiation

Resend flag values

0x00 - Resend routes based on local timer

0x01 - Resend routes immediately

0x02 - Don't resend routes (leave with old AS confederation)

7. Prevention of Loops

Because all IBGP nodes are synchronized before an EBGP peer is transition to a new AS, the local BGP logic can detect the full transition.

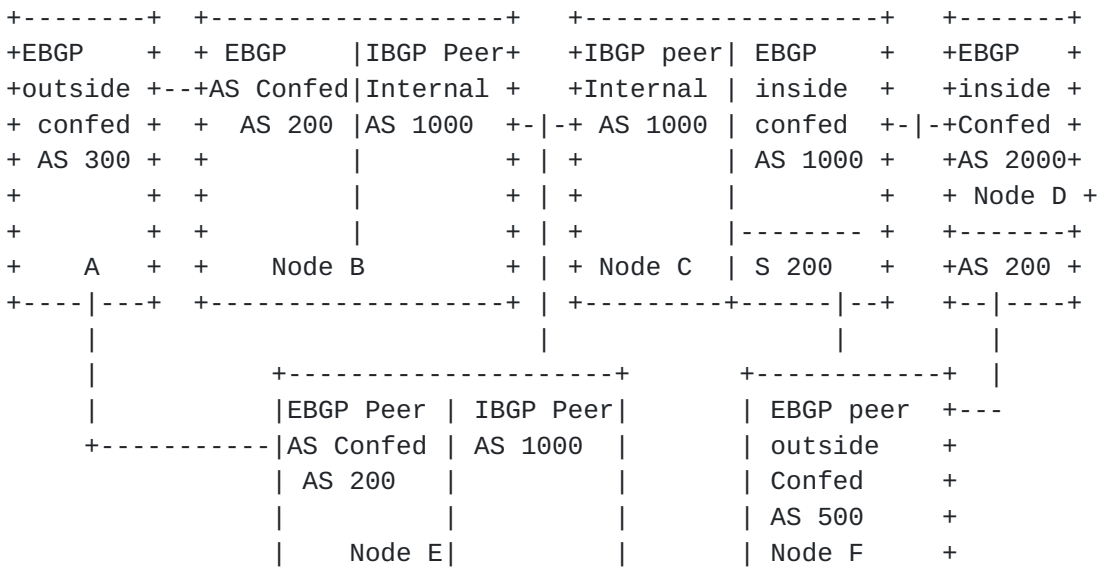
If any active IBGP peer is unable to transition, the whole transition to the new AS stops.

The two stages of IBGP negotation (IBGP only and EBGP/IBGP negotiation) allow the peer to negotiated the ability to AS change before information exterior peers. The two stage IBGP negotiation can reduce or eliminate chatter to EBGP peers while the IBGP peers settle on the decision to change the AS.

The two stage negotiation requires that IBGP peers deal with the multiple AS identity that AS confederation nodes have. Each AS confederation node has an internal AS and a Confederation Edge AS. This section describes two scenarios of using the two stage process. In the first scenario, the IBGP stage 1 synchronization fails. In the second example, the IBGP Stage 1 synchronization succeeds.

In the example below, Node B, C, D an E are inside the AS confederation. Nodes A and F are outside the AS confederation (denoted as AS 200).

When node D is disconnected from Node C, the dynamic AS Confederation Edge switch takes place. Node B, C, and E switch to AS 1000 for IBGP and EBGP. Their split view of the world (AS 200/1000) becomes just AS 1000. Node C detects the loss of the Connection to D, and sends an indication to switch to the Internal AS.



Dual AS identities for an IBGP with Dynamic AS Confederation

Loop Prevention

1) Two stage commit in IBGP peers with failure

1. Node C's sends a Capability message to Node B with sequence number 10 and an Add capability for Dynamic AS Confed_Edge Capability. Inside the Dynamic AS Confed_Edge capability the flags are set for: IBGP, Reserve and AS-in-Use Internal, Success (always on request capability). The AS Confederation is 200, and the Internal AS is 1000.

2. Node C's sends a Capability message to Node E with sequence number 30 and an Add capability for Dynamic AS Confed_Edge Capability. Inside the Dynamic AS Confed_Edge capability the flags are set for: IBGP, Reserve and AS-in-Use Internal, Success (always on request capability); the AS Confederation is 200, and the Internal AS is 1000.

3. Node B trys to engage the AS 1000 switch but fails due to some internal problem.

4. Node B sends a dynamic capability (seqence number 10) with an ACK flag set and a Dynamic AS Confed_Edge capability inside with flags set to: IBGP, Reserve, AS-in-Use Internal, Failure.

5. Node E successful can switch to the Internal AS. Node E sends the capability message with an ACK flag, sequence number 30, and an AS Confed_Edge capability inside. The flags inside the AS Confed_Edge capability are set to: IBGP, Reserve, AS-In Use Internal, and Success.

6. Node C sends sends a capability message to Node E with sequence number 31, Delete flag, Dynamic AS Confed_Edge capability with flag bits set to: IBGP, Reserve, AS-in-Use Internal, and Failure.

7. Node E clears all Dynamic AS Confederation state and respond with a capability messages with an "Ack" flag set, sequence number 31, Delete Flag, Dynanic AS Confed_Edge capability with flag bits set to:IBGP, Reserve, AS-in-Use Internal and Failure.

8. Node C, B, and E form one portion of AS confederation 200, and Node D forms another portion. (These loops are possible with

normal AS Confederations and the rejection by Node B stops the Dynamic AS Confederation Edge from preventing the loops.

2) Two stage commit with IBGP peers with success

1. Node C's sends a Capability message to Node B with sequence number 10 and an Add capability for Dynamic AS Confed_Edge Capability. Inside the Dynamic AS Confed_Edge capability the flags are set for: IBGP, Reserve and AS-in-Use Internal, Success (always on request capability). Inside the Dynamic AS Confed_Edge capability the AS Confederation is 200, and the Internal AS is 1000.

2. Node C's sends a Capability message to Node E with sequence number 30 and an Add capability for Dynamic AS Confed_Edge Capability. Inside the Dynamic AS Confed_Edge capability the flags are set for: IBGP, Reserve and AS-in-Use Internal,and sucess. Success (always on request capability). Inside the Dynamic AS Confed_Edge capability, the AS Confederation is 200, and the Internal AS is 1000.

3. Node B trys to engage the AS 1000 switch and succeeds. Node B responds to Node C with a capability message with sequence number 10, Ack Flag set, and a Dynamic AS Confed_Edge caability inside. The Dynmic AS Confed_Edge capability has flags set for: IBGP, Reserve, AS in-Use, and Success.

4. Node E successful can switch to the Internal AS. Node E sends the capability message with an ACK flag, sequence number 30, and an AS Confed_Edge capability inside. The flags inside the AS Confed_Edge capability are set to: IBGP, Reserve, AS-In Use Internal, and Success.

5. Node C's sends a capability message to Node B (seq 11) and Node E (sequence 31) with an Add capability for Dynamic AS Confed_Edge capability. Inside the Dynamic AS Confed_Edge Capability,the flags are set to: IBGP, Reserve and AS-in-Use Internal, Start, and Success (always on request capability). AS 200 is in the AS confederation. AS 1000 is in the internal AS.

6. Node B (sequence 11) sends back a capability messages with ACK bit set on the Add of the Dynamic AS Confed_Edge capability. Within the AS Confed_Edge capability, the flags are set to: IBGP, Start, AS-in-Use Internal, and Success. Node B starts to send capability mesages with Dynamic AS Confed_Edge capability to the E-BGP peers.

1. Node B sends to Node A a capability message with Add of Dynamic AS Confed_Edge capability with sequence number 5. Inside the Dynamic AS Confed_Edge capability, the flags are set to: EBGP, Start, AS-in-Use Internal, and Success.

2. Node A validates it can switch the E-BGP session to receive AS 1000 from Node B. Node A sends back a capability message with sequence number 5 with an ACK flag set with an Internal AS Confed_Edge Capability inside. Inside the AS Confed_Edge Capability the flags are set to: EBGP, Start, AS-in-Use Internal and Success.

7. Node E sends node C a capability messages with sequence number 31, Action Add and An Ack Flag set. The Dynamic AS Confed_edge capabilty is contained inside with flags set to: EBGP, Start, AS-in-Use Internal, and Success. Node E starts to send capability messages with Dynamic AS Confed_Edge capability to the E-BGP peers.

    1. Node E sends to Node A a capability messages with sequence number of 15. Action Add and a Dynamic AS Confed_Edge capability inside. Inside the Dynamic AS Confed_Edge capability the flags are set to: EBGP, Start, AS-in-Use Internal,and Success.

    2. Node A validates it can change the EBGP session to AS 300 to AS 1000, and then sends a capability to confirm the Ack. The capability messages contains sequence number 15, Add Action, Ack flag, and a Dynamic AS Confed_Edge capability inside with flags of: EBGP, Start, AS-in-Use Internal, and Success. Of course, the AS information is AS Confederation 200, As Internal 1000.

8. After Node C sends the capability messages to it's internal peers (Node E and Node B), it sends dynamic capabilities to it's external peers: Node D (in the AS confederation) and Node F (in AS 500 external to the AS Confederation.)

9. Node D processes C capability message

    1. Node D receives a capability message from Node C with sequence number 40, Add Action, and Dynamic AS Confed_Edge capability inside. Inside the Dynamic AS Confed_Edge capability the flags are set to: EBGP, Start, AS-in-Use Internal, and Succcess. The AS information is AS Confederation 200 and AS Internal 1000.

2. Node D determines it can switch the EBGP Session Node C to AS 1000, the internal AS. It peers with AS 1000 as the AS Confederation, and moves the AS value to 200.

3. Node D sends back an capability message with sequence number 40, Add action, and Dynamic AS Confed_Edge capability inside. Inside the Dynamic AS Confed_Edge capability, the flags are set to: EBGP, Start, AS-in-Use Internal and Success. The AS informaiton is AS confederation 200 and AS internal 1000.

10. Node F process C's capability message

1. Node F receives a capability message from Node C with sequence number 200, Add Action, and Dynamic AS Confed_Edge capability inside. Inside the Dynamic AS Confed_Edge capability the flags are set to: EBGP, Start, AS-in-Use Internal, and Success. The AS information is: AS Confederation 200 and AS Internal 1000.

2. Node F determines it can switch the EBGP session with Node C to AS 1000, the internal AS.

3. Node F sends back a capability message with sequence number 200, Add Action, and Dynamic AS Confed_Edge capabiliity inside. Inside the Dynamic AS Confeded_Edge capability, the flags are set to: EBGP, Start, AS-in-Use Internal and Success. AGain, the AS information is: AS Confederation 200 and AS internal 100.

11. At this point:

1. AS 300 has: Node A

2. AS 1000 has: Node C, B and E

3. AS 200 has Node D as an AS confederation with AS 2000 inside.

4. AS 2000 has only Node D.

5. AS 500 has Node F

6. AS 300 peers with AS 1000

7. AS 1000 peers with AS 200 (Node D), AS 300 (Node A), and AS 500 (Node F).

The two stage commit allows the internal AS one stage to confirm resources with IBGP peers prior to disturbing and E-BGP peers. If an E-BGP peer fails after the IBGP cloud has switched, the single E-BGP peer can be dropped and re-initialized.

If there is any local consideration that the AS has split and existing routes may cause a black hole, implementation MAY set the "re-announce all routes now" flag to prevent loops.

8. Security Considerations

The security of the BGP exchange is optionally secured by the TCP MD5 key.

Upon discussion with security reviewers, the addition of this feature will neither improve nor detract from the TCP MD5 level of security. The authors considered adding a "cookie" feature to further secure this exchange. Again, review with security experts indicated this "cookie" feature would not improve the security level.

The TCP session security will continue across the dynamic BGP peer re-association. The TCP sessions dynamic MD5 re-association or key switch would also allow TCP sessions to continue for a long period.

9. IANA Considerations

IANA will need to assign the following values:

a) Dynamic AS Confederation Edge Open Capability

b) Capability Code (1 octet) for the dynamic AS capability

Initial testing may be done by taking these codes from the experimental stage. IANA is requested to provide experimental values for both values.

10. References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", March 1997, <ftp://ftp.isi.edu/in-notes/rfc2119>.

[2] Redback Networks and Cisco, ""Capability Adverstisement with BGP-4"", November 2002, <http://www.ietf.org/rfc/rfc3392.txt>.

[3] Cisco and Cisco, ""Dynamic Capability for BGP-4"", November 2006, <http://www.ietf.org/internet-drafts/draft-ietf-idr-dynamic-cap-09.txt>.

Authors' Addresses

Susan Hares
Green Hills Software
825 Victors Way
Ann Arbor, MI 48108

Phone: +1-734-222-1610
Email: shares@ghs.com

Pratik Bose
Lockheed
700 N Frederick Ave
Gaithersburg, MD 20879

Phone: +1-301-428-4215
Email: pratik.bose@lmco.com

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).