

Network Working Group  
INTERNET DRAFT

Paul Traina  
Juniper Networks, Inc.  
Danny McPherson  
TCB  
John G. Scudder  
Cisco Systems, Inc.

SEPTEMBER 2002

Autonomous System Confederations for BGP  
<[draft-mcpherson-rfc3065bis-00.txt](#)>

## 1. Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC 2026](#).

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>.

## 2. Abstract

The Border Gateway Protocol (BGP) is an inter-autonomous system routing protocol designed for Transmission Control Protocol/Internet Protocol (TCP/IP) networks. BGP requires that all BGP speakers within a single autonomous system (AS) must be fully meshed. This represents a serious scaling problem that has been well documented in a number of proposals.

This document describes an extension to BGP which may be used to create a confederation of autonomous systems that is represented as a single autonomous system to BGP peers external to the confederation, thereby removing the "full mesh" requirement. The intention of this extension is to aid in policy administration and reduce the

INTERNET DRAFT

SEPTEMBER 2002

management complexity of maintaining a large autonomous system.

### [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 [\[8\]](#).

### [4.](#) Introduction

As currently defined, BGP requires that all BGP speakers within a single AS must be fully meshed. The result is that for  $n$  BGP speakers within an AS  $n*(n-1)/2$  unique IBGP sessions are required. This "full mesh" requirement clearly does not scale when there are a large number of IBGP speakers within the autonomous system, as is common in many networks today.

This scaling problem has been well documented and a number of proposals have been made to alleviate this [\[3,6\]](#). This document presents another alternative alleviating the need for a "full mesh" and is known as "Autonomous System Confederations for BGP", or simply, "BGP Confederations". It has also been observed that BGP Confederations may provide improvements in routing policy control.

This document is a revision of [RFC 3065](#) [\[5\]](#), which is itself a revision to [RFC 1965](#) [\[4\]](#). It includes editorial changes, clarifications and corrections based on deployment experience with BGP Confederations. These revisions are summarized in Appendices A and B.

### [5.](#) Terms and Definitions

#### AS Confederation

A collection of autonomous systems advertised as a single AS number to BGP speakers that are not members of the confederation.

#### AS Confederation Identifier

An externally visible autonomous system number that identifies the confederation as a whole.

Member-AS

An autonomous system that is contained in a given AS confederation.

Member-AS Number

Traina, et al.

[Page 2]

---

INTERNET DRAFT

SEPTEMBER 2002

An autonomous system number visible only within a BGP confederation.

## [6](#). Discussion

It may be useful to subdivide autonomous systems with a very large number of BGP speakers into smaller domains for purposes of controlling routing policy via information contained in the BGP AS\_PATH attribute. For example, one may choose to consider all BGP speakers in a geographic region as a single entity.

In addition to potential improvements in routing policy control, if techniques such as those presented here or in [\[6\]](#) are not employed, [\[1\]](#) requires BGP speakers in the same autonomous system to establish a full mesh of TCP connections among all speakers for the purpose of exchanging exterior routing information. In autonomous systems the number of intra-domain connections that need to be maintained by each border router can become significant.

Subdividing a large autonomous system allows a significant reduction in the total number of intra-domain BGP connections, as the connectivity requirements simplify to the model used for inter-domain connections.

Unfortunately, subdividing an autonomous system may increase the complexity of routing policy based on AS\_PATH information for all members of the Internet. Additionally, this division increases the maintenance overhead of coordinating external peering when the internal topology of this collection of autonomous systems is modified.

Therefore, division of an autonomous system into separate systems may adversely affect optimal routing of packets through the Internet.

However, there is usually no need to expose the internal topology of this divided autonomous system, which means it is possible to regard a collection of autonomous systems under a common administration as a

single entity or autonomous system, when viewed from outside the confines of the confederation of autonomous systems itself.

## [7.](#) AS\_CONFED Segment Type Extension

Currently, BGP specifies that the AS\_PATH attribute is a well-known mandatory attribute that is composed of a sequence of AS path segments. Each AS path segment is represented by a triple <path segment type, path segment length, path segment value>.

In [\[1\]](#), the path segment type is a 1-octet long field with the two following values defined:

Value	Segment Type
1	AS_SET: unordered set of ASs a route in the UPDATE message has traversed
2	AS_SEQUENCE: ordered set of ASs a route in the UPDATE message has traversed

This document specifies two additional segment types:

- |   |   |
|---|---|
| 3 | AS_CONFED_SEQUENCE: ordered set of Member-AS Numbers in the local confederation that the UPDATE message has traversed |
| 4 | AS_CONFED_SET: unordered set of Member-AS Numbers in the local confederation that the UPDATE message has traversed    |

## [8.](#) Operation

A member of a BGP confederation will use its AS Confederation Identifier in all transactions with peers that are not members of its confederation. This confederation identifier is the "externally visible" AS number and this number is used in OPEN messages and advertised in the AS\_PATH attribute.

A member of a BGP confederation will use its Member-AS Number in all transactions with peers that are members of the same confederation as the given BGP speaker.

A BGP speaker receiving an AS\_PATH attribute containing an autonomous system matching its own AS Confederation Identifier shall treat the path in the same fashion as if it had received a path containing its own AS number.

A BGP speaker receiving an AS\_PATH attribute containing an AS\_CONFED\_SEQUENCE or AS\_CONFED\_SET which contains its own Member-AS Number shall treat the path in the same fashion as if it had received

a path containing its own AS number.

#### 8.1. AS\_PATH Modification Rules

When implementing BGP Confederations Section 5.1.2 of [1] is replaced with the following text:

When a BGP speaker propagates a route which it has learned from another BGP speaker's UPDATE message, it shall modify the route's AS\_PATH attribute based on the location of the BGP speaker to which the route will be sent:

- a) When a given BGP speaker advertises the route to another BGP speaker located in its own autonomous system, the advertising speaker shall not modify the AS\_PATH attribute associated with the route.
- b) When a given BGP speaker advertises the route to a BGP speaker located in a neighboring autonomous system that is a member of the configured autonomous system confederation, the advertising speaker shall update the AS\_PATH attribute as follows:

- 1) if the first path segment of the AS\_PATH is of type AS\_CONFED\_SEQUENCE, the local system shall prepend its own Member-AS Number as the last element of the sequence (put it in the leftmost position).
  - 2) if the first path segment of the AS\_PATH is not of type AS\_CONFED\_SEQUENCE the local system shall prepend a new path segment of type AS\_CONFED\_SEQUENCE to the AS\_PATH, including its own Member-AS Number in that segment.
- c) When a given BGP speaker advertises the route to a BGP speaker located in a neighboring autonomous system that is not a member of the configured autonomous system confederation, the advertising speaker shall update the AS\_PATH attribute as follows:
- 1) if any path segments of the AS\_PATH are of the type AS\_CONFED\_SEQUENCE or AS\_CONFED\_SET, those segments shall be removed from the AS\_PATH attribute, leaving the sanitized AS\_PATH attribute to be operated on by steps 2 or 3.
  - 2) if the first path segment of the remaining AS\_PATH is of type AS\_SEQUENCE, the local system shall prepend its own AS Confederation Identifier as the last element of the sequence (put it in the leftmost position).
  - 3) if there are no path segments following the removal of the

first AS\_CONFED\_SET/AS\_CONFED\_SEQUENCE segments, or if the path segment of the remaining AS\_PATH is not of type AS\_SEQUENCE the local system shall prepend a new path segment of type AS\_SEQUENCE to the AS\_PATH, including its own AS Confederation Identifier in that segment.

When a BGP speaker originates a route:

- a) the originating speaker shall include an empty AS\_PATH attribute in all UPDATE messages sent to BGP speakers residing within the same autonomous system. (An empty AS\_PATH attribute is one whose length field contains the value zero).
- b) the originating speaker shall include its own Member-AS Number in an AS\_CONFED\_SEQUENCE segment of the AS\_PATH attribute of all

UPDATE messages sent to BGP speakers located in neighboring Member-ASs that are members of the local confederation (i.e., the originating speaker's Member-AS Number will be the only entry in the AS\_PATH attribute).

- c) the originating speaker shall include its own AS Confederation Identifier in an AS\_SEQUENCE segment of the AS\_PATH attribute of all UPDATE messages sent to BGP speakers located in neighboring autonomous systems that are not members of the local confederation. (In this case, the originating speaker's AS Confederation Identifier will be the only entry in the AS\_PATH attribute).

## [9](#). Error Handling

It is an error for a BGP speaker to receive an update message with an AS\_PATH attribute which contains AS\_CONFED\_SEQUENCE or AS\_CONFED\_SET segments from a neighbor which is not located in the same confederation. If a BGP speaker receives such an update message, it SHALL treat the message as having a malformed AS\_PATH according to the procedures of [\[1\] Section 6.3](#) ("UPDATE message error handling").

## [10](#). Common Administration Issues

It is reasonable for Member-ASs of a confederation to share a common administration and IGP information for the entire confederation.

It shall be legal for a BGP speaker to advertise an unchanged NEXT\_HOP and MULTI\_EXIT\_DISCRIMINATOR (MED) attribute to peers in a neighboring AS within the same confederation.

In addition, the restriction against sending the LOCAL\_PREFERENCE attribute to peers in a neighboring AS within the same confederation is removed.

Path selection criteria for information received from members inside a confederation MUST follow the same rules used for information received from members inside the same autonomous system, as specified in [1].

## 11. Compatability Considerations

All BGP speakers participating as member of a confederation MUST recognize the AS\_CONFED\_SET and AS\_CONFED\_SEQUENCE segment type extensions to the AS\_PATH attribute.

Any BGP speaker not supporting these extensions will generate a NOTIFICATION message specifying an "UPDATE Message Error" and a sub-code of "Malformed AS\_PATH".

This compatibility issue implies that all BGP speakers participating in a confederation MUST support BGP confederations. However, BGP speakers outside the confederation need not support these extensions.

## 12. Deployment Considerations

BGP confederations have been widely deployed throughout the Internet for a number of years and are supported by multiple vendors.

Improper configuration of BGP confederations can cause routing information within an AS to be duplicated unnecessarily. This duplication of information will waste system resources, cause unnecessary route flaps, and delay convergence.

Care should be taken to manually filter duplicate advertisements caused by reachability information being relayed through multiple Member-ASs based upon the topology and redundancy requirements of the confederation.

Additionally, confederations (as well as route reflectors), by



different locations in a confederation, have been shown to cause permanent oscillation between candidate routes when using the tie breaking rules required by BGP [1]. Care must be taken when selecting MED values and tie breaking policy to avoid these situations.

One potential way to avoid this is by configuring inter-Member-AS IGP metrics higher than intra-Member-AS IGP metrics and/or using other tie breaking policies to avoid BGP route selection based on incomparable MEDs.

### [13. Security Considerations](#)

This extension to BGP does not change the underlying security issues inherent in the existing BGP, such as those defined in [7].

### [14. Acknowledgments](#)

The general concept of BGP confederations was taken from IDRP's Routing Domain Confederations [2]. Some of the introductory text in this document was taken from [6].

The authors would like to acknowledge Bruce Cole for his implementation feedback and extensive analysis of the limitations of the protocol extensions described in this document and [5]. We would also like to acknowledge Srihari Ramachandra, Alex Zinin, Naresh Kumar Paliwal, Jeffrey Haas and Bruno Rijsman for their feedback and suggestions.

Finally, we'd like to acknowledge Ravi Chandra and Yakov Rekhter for providing constructive and valuable feedback on earlier versions of this specification.

### [15. References](#)

- [1] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", [RFC 1771](#), March 1995.
- [2] Kunzinger, C., Editor, "Inter-Domain Routing Protocol", ISO/IEC 10747, October 1993.
- [3] Haskin, D., "A BGP/IDRP Route Server alternative to a full mesh routing", [RFC 1863](#), October 1995.
- [4] Traina, P. "Autonomous System Confederations for BGP", [RFC 1965](#), June 1996.

- [5] Traina, P., McPherson, D. and Scudder, J., "Autonomous System Confederations for BGP", [RFC 3065](#), February 2001.
- [6] Bates, T., Chandra, R. and E. Chen, "BGP Route Reflection An Alternative to Full Mesh IBGP", [RFC 2796](#), April 2000.
- [7] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", [RFC 2385](#), August 1998.
- [8] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.

## [16.](#) Authors' Addresses

Paul Traina  
Juniper Networks, Inc.  
1194 N. Mathilda Ave.  
Sunnyvale, CA 94089 USA

Phone: +1 408 745-2000  
EMail: [pst+confed@juniper.net](mailto:pst+confed@juniper.net)

Danny McPherson  
TCB

EMail: [danny@tcb.net](mailto:danny@tcb.net)

John G. Scudder  
Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, CA 95134

Phone: +1 734.302.4128  
EMail: [jgs@cisco.com](mailto:jgs@cisco.com)

INTERNET DRAFT

SEPTEMBER 2002

17. [Appendix A](#): Comparison with [RFC 1965](#)

The most notable change from [4] is that of reversing the values AS\_CONFED\_SEQUENCE(4) and AS\_CONFED\_SET(3) to those defined in section "AS\_CONFED Segment Type Extension". The reasoning for this is that in the initial implementation, which was already widely deployed, they were implemented backwards from [4], and as such, subsequent implementations implemented them backwards as well. In order to foster interoperability and compliance with deployed implementations, they've therefore been changed here as well.

The "Compatibility Discussion" was removed and incorporated into other discussions in the document. The use of the term "Routing Domain Identifier" was replaced with Member-AS Number.

The mention of hierarchical confederations was removed due to the fact that it is not actually possible to deploy confederations hierarchically. That is, a Member-AS of confederation A can only be a simple autonomous system, it cannot itself be a confederation B (whose internal topology is hidden from confederation A).

Finally, the "Deployment Considerations" section was expanded a few subtle grammar changes were made and a bit more introductory text was added.

18. [Appendix B](#): Comparison with [RFC 3065](#)

Added discussion regarding inability to nest confederations.

Added text regarding not propagating confederation attributes beyond confederation boundaries.

Finally, made use of "Member-AS" and "AS Confederation Identifier" terminology more consistent.

