

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: April 15, 2011

M. Boucadair  
C. Jacquenet  
France Telecom  
October 12, 2010

Constrained Multiple BGP Paths  
draft-boucadair-idr-constrained-multiple-path-00

## Abstract

It is commonly agreed the continuous increase of routing tables is a sensitive issue which may question the sustainability of the Internet. This document describes a solution which makes use of multiple paths without inducing drastic increase of routing tables. A constrained procedure to install additional paths in the RIB is described. Multiple paths are installed according to rules agreed between adjacent peers and also based on external events (e.g., proactive detection of link congestion).

## 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](#) [[RFC2119](#)].

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on April 15, 2011.

## Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

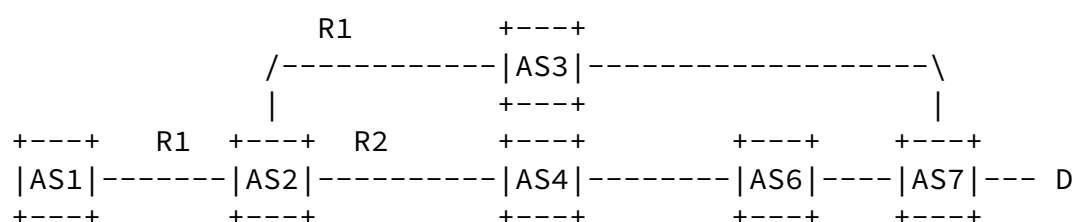
<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">1.1.</a>	Contribution of this I-D . . . . .	<a href="#">4</a>
<a href="#">2.</a>	Extended NLRI Encoding . . . . .	<a href="#">4</a>
<a href="#">3.</a>	Maximum Path Capability . . . . .	<a href="#">5</a>
<a href="#">4.</a>	NOTIFICATION Error Code . . . . .	<a href="#">6</a>
<a href="#">5.</a>	Procedure . . . . .	<a href="#">6</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">7</a>
<a href="#">7.</a>	Security Considerations . . . . .	<a href="#">8</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">8</a>
<a href="#">9.</a>	References . . . . .	<a href="#">8</a>
<a href="#">9.1.</a>	Normative References . . . . .	<a href="#">8</a>
<a href="#">9.2.</a>	Informative References . . . . .	<a href="#">8</a>
	Authors' Addresses . . . . .	<a href="#">8</a>

## 1. Introduction

[I-D.ietf-idr-add-paths] defines a procedure to support multiple paths. The support of this feature may exacerbate the increase of routing tables which is seen as a critical issue for the sustainability of the overall Internet [[I-D.irtf-rrg-recommendation](#)] [[I-D.narten-radir-problem-statement](#)].

In the meantime, allowing to store multiple paths in the RIB is motivated in several scenarios such as the following:

- o Load balancing;
- o Proactive reaction due to congestion events. As an illustration, Figure 1 shows an architecture where three paths to reach D are received by AS2. After applying the route selection process defined in [\[RFC4271\]](#), only R1 is selected. This route is then propagated to AS1. If AS3 experiences congestion on its link to AS7 for instance, part of the traffic is likely to be lost. If the procedure described in this document is applied, then once a congestion event is observed in AS3 (local to an ASBR, intra-domain links, involved intra-domain routers or on inter-domain links) an UPDATE message is sent by AS3 to AS2 to notify a congestion by means of a dedicated flag in the extended NLRI. Once this UPDATE message is received by AS2, the route selection process is applied and an additional route is installed in the RIB. An UPDATE message including the extended NLRI conveying two paths is then sent to AS1, R1 being tagged as a congested route (See Figure 2). This document focuses on this use case.



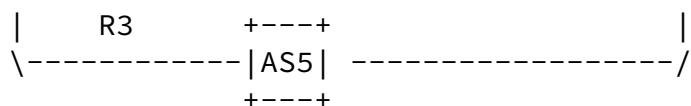


Figure 1: Example Architecture

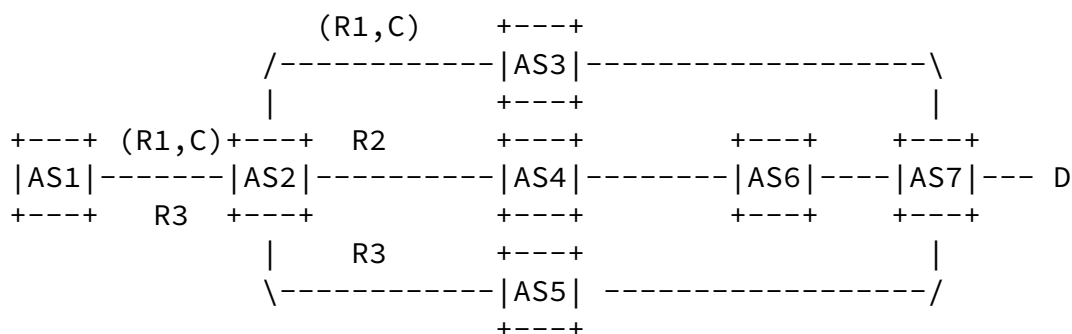


Figure 2: Congested link

### [1.1.](#) Contribution of this I-D

This document proposes a constrained multiple path procedure which allows BGP peers to manage multiple paths towards a given destination in a controlled fashion.

This document provides a scenario with congestion. Other use cases could also be considered, such as QoS-inferred route tagging capabilities.

This document re-uses some of the encodings proposed in [\[I-D.ietf-idr-add-paths\]](#).

## [2.](#) Extended NLRI Encoding

The current encoding defined in [[I-D.ietf-idr-add-paths](#)] does not allow to tag a specific enclosed path described in the Extended NLRI encoding. The NLRI encoding depicted in Figure 3 is used in this document instead of the one specified in [[I-D.ietf-idr-add-paths](#)].

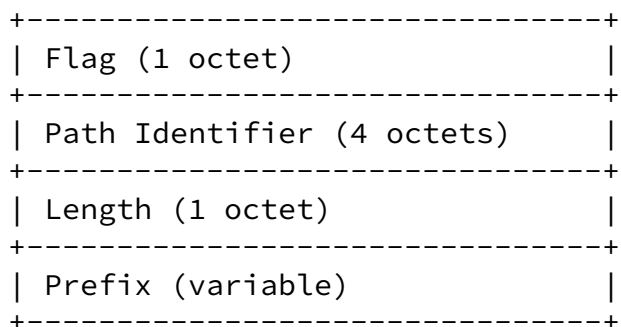


Figure 3: Extended NLRI

Except the Flag field, the remaining fields are similar to what is defined in Section 3 of [[I-D.ietf-idr-add-paths](#)].

The structure of the Flag field is shown in Figure 4.

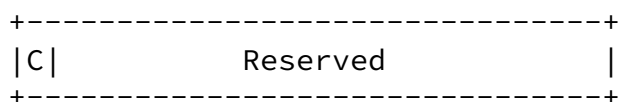


Figure 4: Flag

The first bit (called C bit) is used to indicate whether the path is congested (C=1) or not (C=0). The remaining bits are set to 0.

Further values can be defined in the future if required.

### 3. Maximum Path Capability

This section describes a new BGP Capability [[RFC5492](#)] called Maximum Path Capability. The format of this new Capability is shown in Figure 5.

```
+-----+
| Multiple Paths Max (1 octet)           |
+-----+
| Address Family Identifier (2 octets)    |
+-----+
| Subsequent Address Family Identifier (1 octet) |
+-----+
| Send/Receive (1 octet)                 |
+-----+
```

Figure 5: Multiple Paths Capability

Multiple Paths Max field indicates the maximum number of multiple paths to a given destination prefix that can be supported by a given BGP peer. The number of multiple paths that can be exchanged between

two BGP peers MUST NOT exceed the Multiple Paths Max threshold.

For the remaining fields, the same description as what is specified in Section 4 of [[I-D.ietf-idr-add-paths](#)] is assumed.

#### [4.](#) NOTIFICATION Error Code

This document defines a new NOTIFICATION error code:

Error Code	Name
-----	-----
TBA	Maximum Path

The following error subcode is defined:

Error SubCode	Name
-----	-----
TBA	Maximum Path Reached

#### [5.](#) Procedure

The procedure for exchanging constrained multiple paths is as follows:

- o During BGP session initialisation, a peer supporting the procedure described in this document includes the Maximum Path Capability in its Capability Set;
- o A BGP speaker can advertise multiple paths to a BGP peer only if a Maximum Path Capability was included in its Capability Set received from an adjacent peer;

- o Furthermore, if a BGP peer announces a number of routes towards a destination prefix that exceeds what has been negotiated during the OPEN phase, a NOTIFICATION message MUST be sent and SHOULD include an adequate Error Code/SubCode that corresponds to the exceeded multiple path threshold (See [Section 4](#)).
- o Once the capability negotiation phase is completed, BGP peers adopt the normal BGP behaviour as specified in [[RFC5492](#)];

- o Each AS would implement means/tools to monitor its intra and inter-domain links. These tools are out of the scope of this document. Once a threshold is reached (e.g., 75% of link utilisation), an event is sent to the ASBRs. These nodes send UPDATE messages to their peers to notify them about the status of advertised routes. C bit is set to 1 when a given route is seen as congested;
- o Once this UPDATE is received by a BGP peer, it re-computes the routes to be installed in the RIB. If the selected route is congested, a new route is added to the local RIB. Both routes will be advertised using the extended NLRI to adjacent BGP speakers;
- o This process can be re-iterated until reaching the maximum supported multiple paths. Note that only one route with the flag C set to 0 is installed in the local RIB. A local BGP speaker accept to install a new route if and only if the best route is congested; otherwise only one route is installed in the local RIB.
- o The removal of alternative path can be undertaken by a BGP speaker according to local or external events.

[[Note: The document does not elaborate on load balancing and how the traffic is distributed among available path.]]

[[Note: For load banaling purposes, a metric can be conveyed to inform a BGP peer about the BW of the downstream interconnection link (i.e., interconnection links with one hop adjacent ASes.]]

## 6. IANA Considerations

This document requests

- o a new code for the Maximum Path Capability
- o Maximum Path Error Notification code

- o Maximum Path Reached error sub-code



## [7.](#) Security Considerations

This document does not introduce any security issue other than those elaborated in [[RFC4271](#)].

## [8.](#) Acknowledgements

Many thanks to David Binet for his review.

## [9.](#) References

### [9.1.](#) Normative References

- [I-D.ietf-idr-add-paths]  
Walton, D., Retana, A., Chen, E., and J. Scudder,  
"Advertisement of Multiple Paths in BGP",  
[draft-ietf-idr-add-paths-04](#) (work in progress),  
August 2010.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate  
Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway  
Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement  
with BGP-4", [RFC 5492](#), February 2009.

### [9.2.](#) Informative References

- [I-D.irtf-rrg-recommendation]  
Li, T., "Recommendation for a Routing Architecture",  
[draft-irtf-rrg-recommendation-14](#) (work in progress),  
September 2010.
- [I-D.narten-radir-problem-statement]  
Narten, T., "On the Scalability of Internet Routing",  
[draft-narten-radir-problem-statement-05](#) (work in  
progress), February 2010.

Authors' Addresses

Mohamed Boucadair  
France Telecom

Email: [mohamed.boucadair@orange-ftgroup.com](mailto:mohamed.boucadair@orange-ftgroup.com)

Christian Jacquenet  
France Telecom

Email: [christian.jacquenet@orange-ftgroup.com](mailto:christian.jacquenet@orange-ftgroup.com)

