Interdomain Working Group
Internet-Draft
Intended status: Standards Track
Expires: 11 April 2026

S. Litkowski
Cisco
J. Haas
HPE
K. Patel
Arrcus

8 October 2025

Inter Domain considerations for Constrained Route distribution
draft-litkowski-idr-rtc-interas-03

Abstract

RFC4684 defines Multi-Protocol BGP (MP-BGP) procedures that allow BGP speakers to exchange Route Target reachability information in order to limit the propagation of Virtual Private Networks (VPN) Network Layer Reachability Information (NLRI).

RFC4684 addresses both intra domain and inter domain distributions. Operational deployment experience shows that the current distribution model defined in RFC4684 for inter domain may cause some issue in specific scenarios.

This document proposes alternate route distribution rules for inter domain in order to address these specific scenarios.

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 https://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 11 April 2026.

Copyright Notice

Table of Contents

1. Requirements Language 2
2. Standard Inter AS RT membership NLRI propagation 2
3. Limitations of the current approach 4
3.1. Disjoint ASes 4
3.2. Slow convergence 6
3.3. Suboptimal routing 6
4. Considering multiple paths of the RT NLRI 6
5. Operational considerations 7
6. Security considerations 7
7. Acknowledgements 7
8. IANA Considerations 7
9. Normative References 7
Authors' Addresses 8

1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Standard Inter AS RT membership NLRI propagation

[RFC4684] Section 3.1 and 3.2 describes respectively inter-AS and intra-AS VPN route distribution and distinguish two types of Route Target Membership NLRIs (RT NLRIs):

* Locally originated NLRI where origin-as field of the NLRI is equal to the local AS number.

* External NLRI where origin-as field of the NLRI is different from the local AS number.

When the mechanism detailed in [RFC4684] is in place, inter AS VPN route propagation happens only on the shortest path towards the peer ASes by pruning of some branches of the distribution tree.

Existing implementations of [RFC4684] exhibit two flavors of pruning for interAS that are both compatible with [RFC4684].

* Pruning based on peering type: pruning is applied when RT membership path are learned from eBGP peers only. No pruning is applied when path is iBGP. The pruning applies regardless of the RT NLRI (locally originated or external).

* Pruning based on NLRI type: pruning is applied to external RT NLRIs (source AS different from local AS). This pruning rule applies both to eBGP and iBGP.

These different approaches in pruning can lead to different VPN route distribution behaviors as highlighted in the following example.

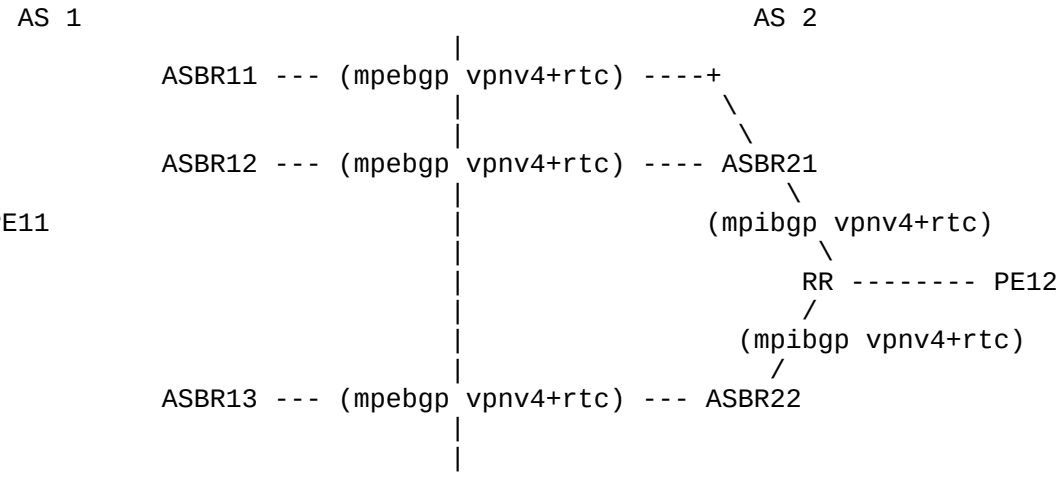

Figure 1: Inter-AS VPN scenario

An example is provided in Figure 1. ASBR11, ASBR12 and ASBR13 are part of the AS1. We consider that all PE11 and PE12 are interested in an any-to-any VPN using RT 65535:10. All ASBRs will generate and advertise the same RT NLRI 1:65335:10/96 towards ASBR21/22.

When implementation uses peering type based pruning: as ASBR21 has two ebgp paths for 1:65535:10/96, only the best path (ASBR11) will be picked up as a branch of the VPN route distribution tree because the RT NLRI is received over eBGP. However, RR in AS2 has two paths for 1:65535:10/96, one from ASBR21, one from ASBR22. Because the paths are iBGP, RR considers all paths as branches of the VPN distribution tree. As a consequence, VPN routes from PE12 will be distributed to ASBR21 and ASBR22. In AS1, ASBR11 and ASBR13 will receive the routes.

When implementation uses NLRI type based pruning: as ASBR21 has two ebgp paths for 1:65535:10/96, only the best path (ASBR11) will be picked up as a branch of the VPN route distribution tree because the RT NLRI is an external RT NLRI. Similarly, RR in AS2 has two paths for 1:65535:10/96, one from ASBR21, one from ASBR22. Because the NLRI is an external RT NLRI, RR selects also only the best path (e.g.: ASBR21) as a branch of the VPN distribution tree. As a consequence, VPN routes from PE12 will be distributed to ASBR21 only. In AS1, only ASBR11 will receive the routes.

3. Limitations of the current approach

The current model of distribution of VPN routes across ASes when RT membership is advertising is optimizing the number of VPN route states to be maintained on nodes. While this is a good goal, this may lead to some limitations/issues in some scenarios.

3.1. Disjoint ASes

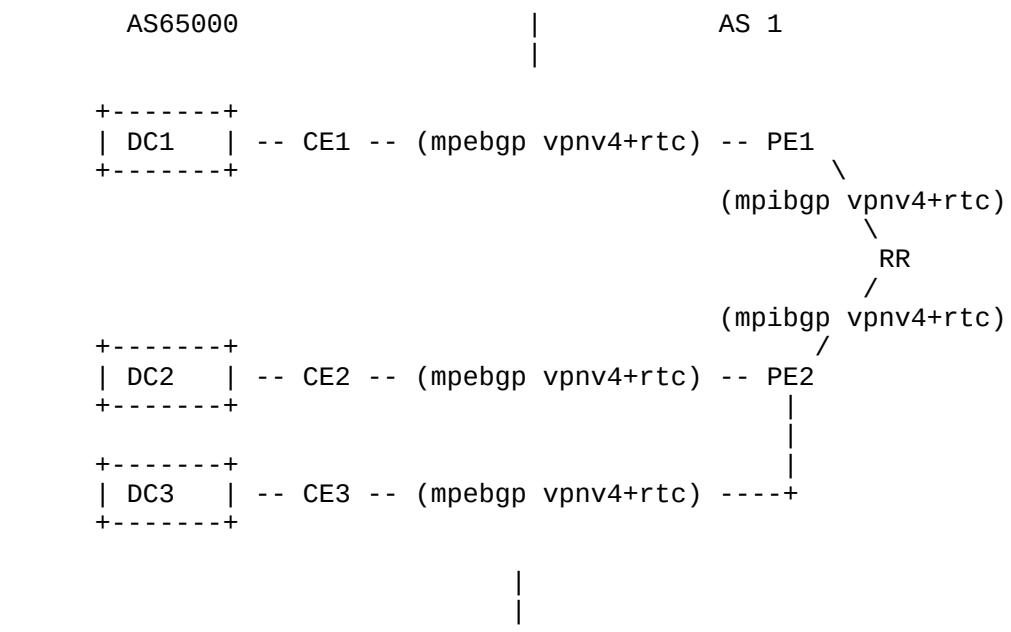

Figure 2: Inter-AS with disjoint ASes

Figure 2 presents a scenario where datacenters are connected through MPLS VPN inter-AS option B [RFC4364] to a service provider network. RT membership is distributed to optimize distribution of VPN routes. In this scenario, all datacenters are using the same AS number, generally a private ASN (65000). As we expect DCs to communicate between each other, some features like "as-override" are deployed on PEs to overcome ASPATH loop issue.

CE1, CE2 and CE3 are advertising RT 1:1 respectively to PE1 and PE2, the generated NLRI would be 65000:1:1/96. According to procedures defined in [RFC4684] Section 3.2, both PEs are using the standard BGP route selection and advertisement rules. PE2 has two paths for RT NLRI 65000:1:1/96, picks one as best (e.g.: CE2) and advertises it to the route-reflector. PE1 advertises its path learned from CE1 to the RR. RR in AS1 has two paths for 65000:1:1/96 and will pick one as best (e.g.: path from PE1). The VPN route distribution tree will be established to PE1 only, PE2 will never get VPN routes for RT 1:1. However, even if RR1 picked up PE2 has best path, because PE2 picked up CE2 as best path, CE3 will have never received VPN routes for RT 1:1.

3.2. Slow convergence

In Figure 1, a VPN route Pv with RT 65335:10 from PE12 is distributed only to ASBR11 by RR in AS2. In AS1, PE11 has a single path to reach Pv. If ASBR11 fails, BGP convergence should occur to provide a new path to PE12:

1. ASBR21 detects the failure of ASBR11 and picks up a new best path for RT 1:65335:10/96 through ASBR12.

2. ASBR21 sends Pv to ASBR12.

3. ASBR12 sends Pv to PE11.

This convergence process could be very slow in high scale scenarios, thus not fitting the service level agreements that the service provider maintains.

3.3. Suboptimal routing

In Figure 1, a VPN route Pv with RT 65335:10 from PE12 is distributed only to ASBR11 by RR in AS2. In AS1, PE11 has a single path to reach Pv from ASBR11. However, from a geographical point of view, ASBR11 may not be the best option for PE11 to reach PE12 (ASBR13 may provide a better end-to-end path). PE11 doesn’t have the ability to pick the best path to Pv from its point of view.

4. Considering multiple paths of the RT NLRI

This document proposes an alternative to the default behavior proposed by [RFC4684] for inter-AS VPN route distribution. Any alternate behavior SHOULD consider multiple paths for an external RT NLRI among the ones available to solve the limitations highlighted in this document. Implementations MAY propose one or more alternate behaviors to balance between adding more VPN routes states within the network and solving the limitations highlighted in this document.

As examples:

* An implementation MAY support the ability to consider all paths for external RT NLRIs as eligible to be part of the VPN route distribution tree regardless of the origin-as. Thus, all VPN routes will be propagated other all possible distribution paths.

* An implementation MAY support the ability to consider all paths for external RT NLRIs as eligible to be part of the VPN route distribution tree for only a subset of origin-as (e.g.: configured list of ASes, or all private ASes...). Thus, VPN routes will be propagated other all possible distribution paths only for a subset of destination ASes.

* An implementation MAY support the ability to consider the best and second best paths for external RT NLRIs as eligible to be part of the VPN route distribution tree regardless of the origin-as. This would allow to provide at least one alternate path for VPN routes in the destination ASes.

5. Operational considerations

Enabling alternate behaviors in consideration of external RT NLRIs that deviate from the default behavior specified in [RFC4684] may have operational implications as VPN routes may be distributed across additional paths leading to increase of BGP prefixes and paths on some devices. Users must carefully evaluate the impact of these changes on their network/router scale.

Choosing the intended behavior for considering paths of external RT NLRIs is a local decision. Different nodes can use different behaviors without breaking the overall functionality of the VPN: alternate behaviors are adding paths for the VPN routes. However, in order to overcome the limitations of [RFC4684] presented in Section 3, it is important to ensure that all nodes that are participating to the VPN route distribution (e.g.: RRs, ASBRs...) are configured with a behavior that fulfills the propagation requirements.

6. Security considerations

This document does not introduce any new security issue compared to [RFC4684].

7. Acknowledgements

Authors would like to thank Aravind Kumar Paramasivam for the useful comments and review.

8. IANA Considerations

There is no IANA consideration.

9. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-editor.org/info/rfc4684>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

Authors' Addresses

Stephane Litkowski
Cisco
Email: slitkows@cisco.com

Jeff Haas
HPE
Email: jhaas@juniper.net

Keyur Patel
Arrcus
Email: keyur@arrcus.com