Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: 8 January 2026

S. Krier
J. Horn
Cisco Systems
M. Ciurea
Swisscom AG
7 July 2025

SRv6 BGP Unreachable Prefix Announcement (UPA)
draft-krierhorn-idr-upa-00

Abstract

Summarization is often used in multi-domain networks to improve network efficiency and scalability. With summarization in place, there is a need to signal loss of reachability to an individual prefix covered by the summary. This enables fast convergence by steering traffic away from the node which owns the prefix and is no longer reachable.

This mechanism, referred to as Unreachable Prefix Announcement (UPA), has been specified for IGPs. This document specifies an and equivalent BGP mechanism for multi-AS networks where BGP is used to carry summary routes.

Commented [TG1]: You might want to describe and refer that aggregation is performed in BGP https://datatracker.ietf.org/doc/html/rfc4271#section-9.2.2.2 and relate throughout the document.

About This Document

This note is to be removed before publishing as an RFC.
Status information for this document may be found at https://datatracker.ietf.org/doc/draft-krierhorn-idr-upa/.
Discussion of this document takes place on the Inter-Domain Routing Working Group mailing list (mailto:idr@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/idr/. Subscribe at https://www.ietf.org/mailman/listinfo/idr/.

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 8 January 2026.

Copyright Notice

Table of Contents

1. Introduction

In modern networks, route summarization is a common practice to reduce routing table size and improve scalability. However, summarization can mask the loss of reachability of specific prefixes covered by the summary route, leading to slower convergence times. To address this, Interior Gateway Protocols (IGPs) have implemented an Unreachable Prefix Announcement (UPA) mechanism [I-D.ietf-lsr-igp-ureach-prefix-announce] to explicitly signal the loss of specific prefixes, enabling fast convergence mechanisms like BGP Prefix Independent Convergence (PIC) [I-D.ietf-rtgwg-bgp-pic] on ingress devices even when BGP route aggregation is being used.

This document proposes a similar UPA mechanism for BGP. In multi-AS networks, particularly those leveraging SRv6, where IGP is not running end-to-end, a BGP-based UPA is crucial. It ensures that the loss of reachability for an SRv6 locator or an egress PE loopback, which might be part of a summarized route, can be quickly communicated across AS boundaries, thereby maintaining fast convergence and network stability.

2. Conventions and Definitions

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.

3. Terminology

* UPA: Unreachable Prefix Announcement.
* SRv6: Segment Routing over IPv6.
* BGP PIC: BGP Prefix Independent Convergence.
* PE: Provider Edge router.
* AS: Autonomous System.
* RIB: Routing Information Base.
* MP_UNREACH: Multiprotocol Unreachable NLRI.
* ExtCom: BGP Extended Community.
* AFI: Address Family Identifier.

* SAFI: Subsequent Address Family Identifier.

4. Reference Deployment Scenario

The primary deployment scenario for BGP UPA ~~is a~~are multi-AS networks with ~~an~~ SRv6 deployments. In ~~this~~ these environments, BGP is used to carry SRv6 locators across AS boundaries, and ~~summarization~~ aggregation is performed at these boundaries to maintain scalability. When a specific SRv6 locator within a ~~summary~~ aggregate becomes unreachable, the UPA mechanism ~~is~~ neededs to signal this event across the ASes to the ingress PEs to trigger BGP-PIC mechanism.

This document considers two primary BGP transport options for SRv6:

* BGP IPv6 Unicast (AFI=2, SAFI=1)
* BGP CAR for SRv6 (AFI=2, SAFI=83)

While both options are viable, the rest of this document primarily considers the use of BGP IPv6 Unicast but the described UPA mechanism is applicable to just as well to BGP CAR or any other BGP transport routing deployment that uses route ~~summarization~~aggregation.

5. BGP UPA Message Format

A BGP UPA ~~message is used to~~ announces the loss of reachability of a specific ~~prefix~~path.

The specific prefix whose reachability ~~is~~ has been lost is encoded in the MP_UNREACH_NLRI attribute [RFC4760~~].~~ ~~The~~and a UPA BGP Extended Community attribute (as defined in Section 5.1) ~~is the only other attribute that applies to a UPA message~~ is being added. ~~An Update message carrying a UPA MUST only contain UPA prefixes (i.e., no other reachability advertisements or withdrawals) due to the presence of the UPA Extended Community.~~

5.1. UPA Extended Community

A new Transitive IPv4-Address-Specific Extended Community is defined for UPA.

The structure of this Extended Community is as follows:

* Type Field: TBD (assigned by IANA).
* Sub-Type Field: TBD (assigned by IANA).

Commented [TG2]: I suggest to add references to other documents where these terms are defined.

Commented [TG3]: Shouldn't this be appliccable to IPv4 and IPv6?

* Global Administrator Field (4 bytes): This field carries the BGP Router-ID of the node originating the UPA in BGP. This is helpful for network observability, specifically to troubleshoot~~ing and tracing~~ the originator in ~~a~~ multi-domain networks. It is assumed that BGP Router-IDs are unique within the ~~operator's managed~~BGP ASes.
* Local Administrator Field (2 bytes): This field is set to zero.

6. Trigger for UPA Origination in BGP

UPA origination in BGP can be triggered by two main scenarios:

6.1. Scenario A: IGP Redistribution of Summary into BGP

When an IGP summary route is redistributed into BGP, and a specific component prefix within that summary loses reachability in the IGP, the UPA indication is conveyed from IGP to BGP. The details of this mechanism is implementation specific and outside the scope of this document.

6.2. Scenario B: BGP Aggregation~~/Summarization~~

When BGP itself is performing aggregation~~ or summarization~~, and a constituent specific route goes away, the UPA is triggered internally within BGP.

Implementations SHOULD provide a configurable option to specify which types of specific prefixes trigger UPA (e.g., only /48 prefixes for SRv6 locators).

7. UPA Origination in BGP

UPA origination trigger (in either of the two scenarios) is processed by BGP only when in the absense of a valid reachable route in BGP for that specific prefix. The origination of UPA indication involves the update generation of the BGP UPA message as specified in Section 5.

The UPA state for the prefix SHOULD be retained for a time period to ensure it has been propagated to its neighbors and avoid generation of multiple UPA messages for the same prefix.

8. UPA Propagation in BGP

The propagation of UPA messages in BGP follows the same principles as UPA origination. BGP speakers receiving a UPA will process it (refer Section 7) and propagate it to their peers as appropriate.

Formatted: Italian (Italy)

Commented [TG4]: I believe that the UPA mechanism is being started when in the LSDB and RIB pathes of a given summary or aggregation has been withdrawn correct?

Commented [TG5]: Is this time specified or configured? If configured how?

Commented [TG6]: Since it is a transitive attribute, a reference to https://datatracker.ietf.org/doc/html/rfc4271#section-5 would be helpful.

9. UPA Processing in BGP

A BGP speaker processes UPA messages only for those ~~prefixes~~ pathes for which it does not have an equal length valid reachable route. The processing of UPA message involves notification of unreachability within the router to trigger BGP PIC. The details of this mechanism are outside the scope of this document.

10. UPA Timer

The UPA state needs to be retained in the BGP table for a configurable duration. This is crucial to prevent unwanted flooding and to allow sufficient time for the UPA to be propagated to all relevant peers.

Commented [TG7]: Suggest to describe how it is being configured, ideally by augmenting draft-ietf-idr-bgp-model YANG modules, and which criteria's should be used to define the duration in the Operations and Manageability Considerations section and how it should be monitored in BGP.

11. Backwards Compatibility

The UPA mechanism is designed to be backwards compatible. Since a UPA is propagated as an MP_UNREACH_NLRI, a BGP speaker that does not understand the UPA Extended Community will simply discard or ignore the update as a withdrawal for a non-existent prefix. Implementations SHOULD provide a configuration knob to enable UPA propagation to specific neighbors. The default MUST be to not propagate UPA messages. This ensures that UPA propagation can be limited to the desired domain or network boundary.

Commented [TG8]: I assume that that the configuration applies to neighbor and neighbor-group configurations correct?

12. Security Considerations

The primary security consideration relates to the use of BGP IPv6 Unicast for carrying SRv6 locators. There is a potential for leakage of internal infrastructure details into the public Internet if filtering route policies are misconfigured. The explicit signaling of unreachable prefixes via UPA could reveal more granular internal network topology information if not properly contained. Operators SHOULD ensure robust filtering policies are in place at AS boundaries. The configurable knob to disable UPA propagation to specific neighbors (Section 11) can serve as a mitigation strategy to limit the scope of UPA messages to trusted domains.

Operations and Manageability Considerations

To gain visibility when and to which BGP paths UPA has been applied to and how the UPA message is being propagated throughout the BGP ASes to which BGP speakers, BGP monitoring protocol as defined in RFC 7854 can be leveraged.

With BMP Local RIB as defined in RFC 9069 and by defining a new UPA status code in draft-ietf-grow-bmp-path-marking-tlv#section-3.1, reflecting the UPA state at the RIB.

Commented [TG9]: Operations and Manageability Considerations is defined in https://datatracker.ietf.org/doc/html/draft-opsarea-rfc5706bis.

13. IANA Considerations

This document requests that IANA assign a new Transitive IPv4-Address- Specific Extended Community type and sub-type from the FCFS range for UPA.

14. References

14.1. 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/rfc/rfc2119>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/rfc/rfc4760>.

[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/rfc/rfc8174>.

14.2. Informative References

[I-D.ietf-lsr-igp-ureach-prefix-announce] Psenak, P., Filsfils, C., Voyer, D., Hegde, S., and G. S. Mishra, "IGP Unreachable Prefix Announcement", Work in Progress, Internet-Draft, draft-ietf-lsr-igp-ureach-prefix-announce-09, 2 July 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-lsr-igp-ureach-prefix-announce-09>.

[I-D.ietf-rtgwg-bgp-pic] Bashandy, A., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", Work in Progress, Internet-Draft, draft-ietf-rtgwg-bgp-pic-22, 20 April 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-bgp-pic-22>.

Commented [TG10]: Both document references should be normative. Add RFC 7854, 9069 and draft-ietf-grow-bmp-path-marking-tlv as informative references.

Acknowledgments

The authors would like to acknowledge the contribution of Ketan Talaulikar and Clarence Filsfils for their valuable input and review of this document.

Authors' Addresses

Serge Krier
Cisco Systems
De Kleetlaan 6a
1831 Diegem
Belgium
Email: sekrier@cisco.com

Jakub Horn
Cisco Systems
Milpitas, CA 95035
United States of America
Email: jakuhorn@cisco.com
Mihai Ciurea
Swisscom AG
Alte Tiefenaustrasse 6
CH-3048 Worblaufen
Switzerland
Email: mihai.ciurea@swisscom.com

Formatted: French (Switzerland)

Formatted: German (Switzerland)