Interdomain Routing Working Group
Internet-Draft
Intended status: Standards Track
Expires: 17 September 2026

H. Wang
J. Dong
Huawei Technologies

16 March 2026

BGP-based Unreachable Prefix Advertisement for Inter-Domain Fast Reroute
draft-wang-idr-bgp-upa-00

Abstract

This document specifies a mechanism for advertising unreachable prefixes across Autonomous Systems using BGP. The mechanism enables fast convergence in VPN services when backbone source nodes become unreachable, by allowing Unreachable Prefix Advertisement (UPA) routes propagated through BGP across AS boundaries. This solution extends the IGP-based UPA mechanism defined in RFC9929 to inter-domain scenarios, ensuring remote PEs can promptly detect and react to failures in source domains. The route is not used for packet forwarding but solely for BGP next-hop reachability resolution, enabling fast failover of BGP VPN routes.

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 17 September 2026.

Copyright Notice

Table of Contents

1. Introduction

1.1. Background and Motivation

Segment Routing over IPv6 (SRv6) is increasingly adopted in metro and backbone networks. SRv6 forwarding paths rely on Locator routes distributed through the IGP domain. In large-scale deployments, to control the flooding domain size and routing table scale, multi-area or multi-level solutions are commonly employed. For instance, a non-backbone area typically aggregates routes before advertising them to the backbone area.

However, this approach presents a challenge: when a source node in a non-backbone becomes unreachable, remote backbone or non-backbone areas only receive aggregated routes and remain unaware of the specific node failure. Consequently, VPN services cannot converge promptly in response to the failure. This scenario has been described and addressed within a single domain in [RFC9929], which defines the IGP-based Unreachable Prefix Advertisement (UPA) mechanism.

This document extends this mechanism to the BGP control plane: the UPA route is introduced into BGP not to forward traffic, but to signal unreachability during BGP route recursion.

1.2. Problem Statement

While [RFC9929] provides a robust solution for intra-domain failure detection, it does not address inter-domain scenarios. Consider a case where Autonomous System A (ASa) is peering with Autonomous System B (ASb). ASa deploys the [RFC9929] UPA mechanism internally. When a source node in a non-backbone area of ASa fails:

1. The ASa border ASBR (Autonomous System Border Router) receives the IGP UPA for the unreachable prefix.
2. The ASa ASBR is aware of the failure via IGP UPA.
3. However, this information cannot be conveyed to ASb.
4. ASb PE routers continue to rely on aggregated routes advertised by ASa, unaware of the specific node failure.
5. VPN traffic from ASb to the affected prefix continues to be forwarded toward the failed node because ASb lacks the reachability information required for BGP convergence. Without IGP UPA information being propagated across AS boundaries, BGP cannot properly detect the failure and cannot complete convergence, resulting in unnecessary service disruption.

The key problem is the inability to propagate IGP UPA information across AS boundaries, preventing remote PEs in neighboring ASes from performing timely traffic rerouting.

This document defines a mechanism that enables IGP UPA information to be carried by BGP and propagated across AS boundaries, thereby extending fast reroute capabilities to inter-domain scenarios.

2. 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.

3. Overview of Operation

The proposed mechanism extends the IGP UPA concept defined in [RFC9929] into the BGP layer. The overall operation is as follows:

1. *Generation at ASBR*: When an ASBR receives an IGP UPA route from IGP, it generates a BGP UPDATE carrying a UPA route based on the IGP prefix and associated information.

2. *Advertisement to Neighbors*: The ASBR advertises this BGP UPA route to its external BGP peers in neighboring ASes.

3. *Validation at Receiver*: Upon receiving the BGP UPA route, the receiving router validates it against specific rules (see Section 5.1).

4. *IP RIB Installation*: If validated, the route is installed in the IP RIB with a special flag indicating it represents an unreachable prefix.

5. *BGP VPN Route Recursion*: When BGP VPN routes recurse to this IP RIB entry, the recursion process detects the unreachable state and triggers re-evaluation of the VPN route selection, resulting in fast failover to alternative paths.

6. *Propagation*: Depending on deployment, the BGP UPA route may be redistributed into IGP for intra-domain propagation (following [RFC9929]) or further propagated via BGP to interior PEs or other ASes.

The mechanism is designed to be backward compatible: routers that do not recognize the UPA-specific attributes will treat the BGP UPA route as a normal route, which does not introduce new failures compared to the absence of UPA.

4. BGP UPA Route Advertisement

4.1. Generation of BGP UPA Routes

When an ASBR receives an IGP UPA route via IGP, it MUST generate a corresponding BGP route following these rules:

1. *Prefix*: The NLRI of the BGP route MUST match the IGP UPA prefix (e.g., an IPv6 /64 address).
2. *Next Hop*: The BGP Next Hop attribute MUST be set to the ASBR's own address.
3. *Origin*: The Origin attribute MUST be set to INCOMPLETE (value 2).
4. *AS Path*: The AS Path MUST include the local AS at least once. To ensure proper route preference, it MAY include additional repetitions of the local AS number.
5. *UPA Attribute*: The BGP UPDATE MUST include the UPA Path Attribute.
6. *Aggregation Information*: If the UPA prefix is part of an aggregated route (e.g., /64 UPA prefix belongs to a /48 aggregate), the UPA Attribute MUST include the aggregate prefix information.

The ASBR MUST NOT generate BGP UPA routes for IGP UPA prefixes that it does not have reachability information for the aggregate route.

When the IGP UPA is withdrawn, the ASBR MUST withdraw the corresponding BGP UPA route by sending a BGP UPDATE message with the NLRI included in the Withdrawn Routes field. Upon receipt, the receiving BGP speaker MUST delete the BGP UPA route, terminate the associated unreachability processing, and restore the state to that of before receiving the BGP UPA route.

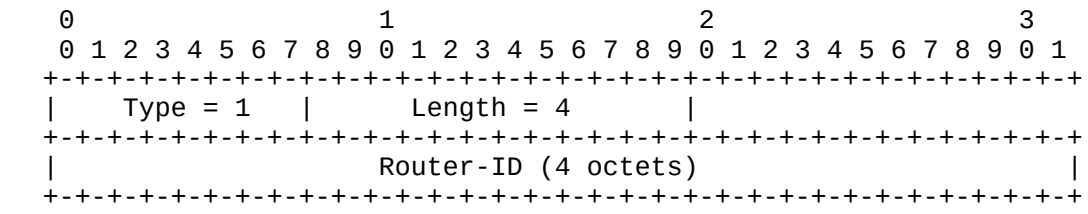
4.2. BGP UPA Path Attributes

4.2.1. UPA Optional Transitive Attribute

This document defines a new optional transitive BGP path attribute called the UPA Attribute (Type Code to be assigned by IANA). The UPA Attribute carries information necessary for validating and processing the UPA route.

4.2.1.1. UPA Origin TLV (Type=1)

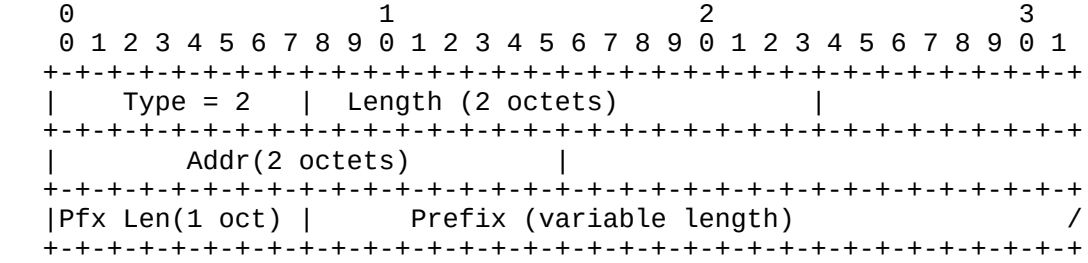
The UPA Origin TLV identifies the BGP speaker that generated the UPA route.



* *Type*: 1 (UPA Origin TLV)
* *Length*: 2 octets, Length = 4
* *Router-ID*: The 4-octets Router ID of the BGP speaker that generated this UPA route. This is used for identification and debugging purposes.

4.2.1.2. Aggregate Prefix TLV (Type=2)

The Aggregate Prefix TLV indicates the aggregate route to which the UPA prefix belongs.



* *Type*: 2 (Aggregate Prefix TLV)
* *Length*: Variable, including the 6-octet header. Length = 3 + ceil(PrefixLength / 8).
* *Addr Family*: 2-octets Address Family Identifier, taken from the registry defined in [RFC4760]. This MUST match the address family of the NLRI.
* *Prefix Length*: 1 octet, the length of the prefix in bits.
* *Prefix*: Variable length, the aggregate prefix. The prefix MUST be a proper supernet of the UPA NLRI prefix on a bit-wise basis.

If the UPA Attribute does not contain an Aggregate Prefix TLV, the receiver MUST consider the validation rule based on aggregate matching to be failed (see Section 5.1).

This document reserves TLV Type values 3-127 for future use. TLV Type values 128-255 are for private use.

4.3. Route Preference Handling

To ensure that BGP UPA routes do not interfere with normal, reachable routes for the same prefix, routes carrying the UPA Attribute MUST be made less preferred than regular routes for the same or more specific prefixes. The advertising ASBR SHOULD apply one or more of the following techniques:

1. *AS Path Preparation*: Repeat the local AS number multiple times in the AS Path. This leverages the AS Path length tie-breaker in BGP path selection.
2. *LOCAL_PREF Reduction*: Set the LOCAL_PREF attribute to a lower value (e.g., 50 or 100) compared to regular routes (typically 100).
3. *MED Increase*: Set the MULTI_EXIT_DISCRIMINATOR (MED) attribute to a higher value (e.g., 200) to influence egress path selection toward preferring non-UPA routes.

These adjustments ensure that when both a UPA route and a normal route for the same prefix exist, BGP will select the normal route as the best path. The UPA route remains available for the specific purpose of indicating unreachability during recursion for VPN routes.

If only a UPA route is received for a prefix, it will be installed as the best path (subject to validation rules) and marked as unreachable in the IP RIB.

5. Processing of BGP UPA Routes

5.1. Validation Rules

Upon receiving a BGP UPDATE containing a route with the UPA Attribute, the receiving router MUST perform the following validation steps before processing it as a UPA route:

1. *Attribute Presence*: The route MUST include the UPA Attribute (as defined in Section 4.2). If the attribute is missing, the route is treated as a normal BGP route.

2. *UPA Origin TLV Presence*: The UPA Attribute MUST contain a UPA Origin TLV (Type=1) with a valid Router-ID.

3. *Aggregate Prefix Matching*:

    * The UPA Attribute MUST contain an Aggregate Prefix TLV (Type=2).

    * The aggregate prefix MUST be a proper supernet of the route NLRI prefix (i.e., Network Address of the aggregate must match the NLRI's network address for the aggregate's prefix length).
    * The Security Principal Identification (SPI) algorithm identifies a specific validation pattern for aggregate prefix processing, ensuring precise network prefix verification.

4. *No Intermediate Specific Routes*: The BGP speaker MUST NOT have any other BGP routes (excluding this UPA route) whose NLRI is more specific than the aggregate prefix but less specific than the UPA prefix. In other words, for an aggregate /48 and UPA /64, there must be no existing routes in the range /49 to /65. This ensures that the UPA route is the most precise route for its destination within the aggregate.

5. *Aggregate Route Consistency (Optional)*: For enhanced validation, the BGP speaker MAY check that the aggregate prefix corresponds to an existing BGP route (which may or may not carry the UPA Attribute). The next hop of that aggregate route MAY be used as part of consistency checks (see Section 5.1.2).

If any of these validation checks fail, the route MUST NOT be installed as a UPA route in the IP RIB. It MAY still be propagated to neighbors based on BGP best path selection and local policy, but the route MUST NOT generate a special "unreachable" entry in the IP RIB.

5.1.1. Aggregate Route Next Hop Consistency

As an additional consistency check, if the BGP speaker has a route for the aggregate prefix (advertised by the same neighbor or a different neighbor), it MAY verify that the next hop of this UPA route matches the next hop of the aggregate route. If the next hops differ, this indicates that the aggregate route is being advertised by a different ASBR that is not advertising a UPA route for the UPA prefix. In this case:

* The UPA route MUST NOT be installed as a UPA route in the IP RIB.
* The UPA route MUST NOT be further propagated.
* The route MAY still be stored in BGP's Adj-RIB-In for policy evaluation.

This check prevents inconsistent states where partial UPA information spreads through the network.

5.2. IP RIB Installation

Once a BGP UPA route successfully passes validation, the receiving router MUST install it in its IP RIB with special handling:

1. *Flagged Installation*: The route entry in the IP RIB MUST be flagged to indicate it is an "unreachable" route. This flag signifies that the prefix should be considered unreachable for the purposes of packet forwarding.

2. *Non-Forwarding*: The route MUST NOT be installed into the FIB (Forwarding Information Base) or control actual packet forwarding. It is solely for BGP recursion and route selection purposes.

3. *BGP Control Plane Only*: The route remains visible to BGP’s control plane processes for the purpose of adjacency resolution (next-hop resolution) for other BGP routes (e.g., VPN routes).

4. *De-installation*: When the BGP UPA route is withdrawn (e.g., upon receipt of a BGP UPDATE with the NLRI withdrawn, or replacement with a validated non-UPA route), the "unreachable" flag MUST be cleared. If the route was previously preventing installation of a more specific route, normal BGP best path selection proceeds for that prefix.

If a non-UPA route for the same prefix or a more specific prefix is subsequently received and becomes the best path, the UPA route may cease to be the best path. The IP RIB entry may be replaced or remain depending on local implementation, but the flag indicates unreachability should be associated with whatever route represents the UPA intention (i.e., the route carrying the UPA Attribute that passed validation).

5.3. BGP VPN Route Recursion

The primary purpose of installing a flagged UPA route in the IP RIB is to influence BGP VPN route selection and forwarding behavior. When a BGP VPN route (e.g., a VPNv4 or VPNv6 route) is evaluated by local BGP:

1. *Next-Hop Resolution*: BGP attempts to resolve the VPN route’s BGP Next Hop by finding a matching route in the IP RIB.

2. *Matching the UPA Route*: If the IP RIB lookup results in matching the UPA route entry (or the most specific route containing the UPA route carries the "unreachable" flag), the recursion process recognizes that the next hop is unreachable.

3. *Trigger Re-evaluation*: Recognizing the unreachability triggers BGP to re-evaluate the best path for that VPN prefix. BGP considers alternative paths whose next hops resolve to reachable routes.

4. *Fast Failover*: This mechanism enables the VPN route to switch to an alternative path without waiting for the conventional BGP withdrawal or update to propagate from the source domain, achieving fast convergence.

This recursion-based failover is analogous to the behavior defined in [RFC9929] for IGP UPA routes, but it operates across AS boundaries and leverages BGP’s existing next-hop resolution mechanism.

6. Propagation Across AS Boundaries

6.1. Intra-AS Propagation via IGP

If the receiving ASBR in ASb has an IGP deployed internally (e.g., IS-IS or OSPFv3), it MAY redistribute the validated BGP UPA route into the IGP as a UPA route. The redistribution MUST follow the RFC 9929 procedures for UPA routes in IGP.

The IGP UPA route will then be propagated throughout ASb’s IGP domain. Interior PEs receiving the IGP UPA will process it according to [RFC9929], installing it in their IP RIB with the unreachable flag and using it for next-hop resolution for BGP VPN routes.

This approach is suitable when the ASb IGP domain is consistent and prefers IGP-based distribution of failure information.

6.2. Intra-AS Propagation via BGP

If the ASb border ASBR is directly connected to interior PEs via BGP (e.g., in an eBGP or iBGP session), it MAY advertise the BGP UPA route to those interior PEs. The advertisement follows these rules:

1. *Attribute Preservation*: The UPA Attribute MUST be included in the BGP UPDATE to the interior PEs. The UPA Origin TLV’s Router-ID MUST remain unchanged (identifying the original advertiser in ASa).

2. *Local Preference Adjustment*: The ASb border ASBR SHOULD adjust LOCAL_PREF, MED, or AS Path from internal peers to ensure the UPA route does not become preferred over any normal, reachable routes for the same prefix in ASb.

3. *Processing by Interior PEs*: Interior PEs receive the BGP UPA route and process it according to the validation rules in Section 5.1. When validated, they install it in their IP RIB with the unreachable flag.

This BGP-based intra-AS propagation is useful when an IGP is not deployed or when the network operator prefers to keep BGP-based reachability separate from IGP.

6.3. Inter-AS Propagation

If ASb is also peering with additional ASes (e.g., ASc), and the BGP UPA route passes validation at the ASb border router, the route MAY be further propagated from ASb to ASc. Propagation follows these principles:

1. *Regular BGP Advertisement*: From ASb's perspective, the UPA route is a regular BGP route (albeit with special attributes). It is subject to ASb's BGP export policies.

2. *Attribute Forwarding*: If ASb's export policy permits, the UPA Attribute MUST be included in the BGP UPDATE sent to ASc. The UPA Origin TLV information remains unchanged.

3. *Validation at ASc*: ASc's border router validates the received BGP UPA route using the same rules defined in Section 5.1. When validated, it installs the route in its IP RIB with the unreachable flag.

4. *Recursive Propagation*: This pattern continues across the appropriate AS boundaries, allowing the unreachable prefix information to reach PEs that may be multiple AS hops away from the failure location.

The propagation model transforms the originally intra-domain IGP UPA mechanism into a trans-AS capability, ensuring that failure information can travel across the global BGP topology as needed.

7. Backward Compatibility

The BGP UPA mechanism is designed to be backward compatible with BGP speakers that do not recognize the UPA Attribute. This is achieved through the optional transitive nature of the attribute and the defensive processing rules.

7.1. Unknown Attribute Handling

When a BGP speaker that does not recognize the UPA Attribute receives a BGP UPDATE carrying this attribute:

* Because the attribute is defined as optional transitive, the speaker MUST preserve the attribute when advertising the route to its peers (per RFC 4271).
* The speaker will treat the route as a normal BGP route without special UPA processing.

Implications:

* The route may be installed as a normal, reachable route in the IP RIB.
* Since the UPA route is generated based on an actual IGP UPA (indicating the original source node is unreachable), advertising this route as a normal reachable route may not lead to improved forwarding behavior. However, it does not introduce new failures compared to the scenario where the UPA mechanism is not deployed at all.
* If the route's preference is low (due to the techniques in Section 4.3), it may not be selected as the best path anyway.

7.2. Hybrid Deployment Scenarios

Consider a scenario where:

1. The ASa ASBR advertising the UPA route understands and supports the UPA Attribute.
2. The ASb border ASBR does not recognize the UPA Attribute but preserves it (optional transitive).
3. The ASb interior PE router does recognize the UPA Attribute.

In this case:

* The ASb border ASBR will treat the route as a normal BGP route. It may or may not install it in the IP RIB (depending on BGP best path selection).
* When ASb's border ASBR advertises the route to its interior PE (preserving the UPA Attribute), the interior PE, recognizing the attribute, will validate and process it as a UPA route.
* The interior PE can thus benefit from the fast failover capability, achieving partial deployment.

This partial deployment can still provide significant benefits for traffic that traverses the ASb interior PE.

8. Security Considerations

8.1. Spoofing of UPA Routes

Since the BGP UPA mechanism causes traffic to be rerouted away from the advertised prefix, a malicious or compromised BGP speaker could potentially generate fake UPA routes to cause denial of service or influence traffic paths. Countermeasures include:

* *Validation of Aggregate Prefix*: The requirement that the UPA prefix must be within a matching aggregate route limits the scope of spoofing to aggregates that the attacker can legitimately influence.
* *Router-ID Attribution*: The UPA Origin TLV includes the Router-ID of the originating speaker. This can be used for auditing and tracing.
* *BGP Security Mechanisms*: Deployments that use BGPsec [RFC8205] can verify that the UPA UPDATE is cryptographically signed by a legitimate AS. The UPA Attribute is transitive and can be part of the signed UPDATE message.

8.2. Information Leakage

The UPA Attribute contains detailed information (Router-ID, aggregate prefix) about failures in remote ASes. This is not significantly different from the information exposure inherent in BGP route advertisements. Operators should consider their policy for receiving and advertising routes with optional transitive attributes.

9. IANA Considerations

9.1. BGP Path Attribute Code Point

IANA is requested to allocate a new BGP path attribute code point from the "BGP Path Attributes" registry for the UPA Attribute. The attribute type code will be specified in the RFC version of this document.

9.2. UPA Attribute TLV Type Registry

IANA is requested to create a new registry named "BGP UPA Attribute TLV Types". The registration procedure is Standards Action. Initial registrations are:

Type	Description
1	UPA Origin TLV
2	Aggregate Prefix TLV
3-127	Unassigned
128-255	Reserved for Private/Experimental Use

Table 1

10. Acknowledgements

The authors would like to thank the authors of [RFC9929] for defining the foundational IGP-based UPA mechanism...

11. 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>.

[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/info/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/info/rfc8174>.

[RFC8205] Lepinski, M., Ed. and K. Sriram, Ed., "BGPsec Protocol Specification", RFC 8205, DOI 10.17487/RFC8205, September 2017, <https://www.rfc-editor.org/info/rfc8205>.

[RFC9929] Psenak, P., Ed., Filsfils, C., Voyer, D., Hegde, S., and G. Mishra, "IGP Unreachable Prefix Announcement", RFC 9929, DOI 10.17487/RFC9929, February 2026, <https://www.rfc-editor.org/info/rfc9929>.

Authors' Addresses

Haibo Wang
Huawei Technologies

Email: rainsword.wang@huawei.com

Jie Dong
Huawei Technologies
Email: jie.dong@huawei.com