BESS Working Group
Internet-Draft
Intended status: Standards Track
Expires: 21 October 2026

J. Tantsura
V. Venkatraman
K. Muppalla
P. Doijode
Nvidia
M. Rzehak
CoreWeave

19 April 2026

EVPN Unreachability Signaling
draft-tantsura-bess-evpn-unreachability-00

Abstract

This document defines a new EVPN Route Type for signaling prefix unreachability information without affecting the forwarding plane. The route type reuses the Route Type 5 (IP Prefix Advertisement) field order defined for EVPN IP prefix routes, adds an Address Family octet for unambiguous IPv4/IPv6 parsing, and appends Reporter TLVs that allow aggregation of unreachability reports from multiple network vantage points.

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 21 October 2026.

Copyright Notice

Table of Contents

1. Introduction

EVPN (Ethernet VPN) [RFC7432] provides a flexible framework for Layer 2 and Layer 3 VPN services. While EVPN includes mechanisms for advertising reachable prefixes via Route Type 5 (IP Prefix Advertisement Route) [RFC9136], there is no standard way to signal unreachability information for monitoring and troubleshooting purposes without affecting the forwarding plane.

Similar to the challenges in standard BGP, EVPN withdrawals are only propagated for prefixes that have been previously announced. This behavior limits the ability of operators to share information about prefix unreachability for prefixes that were never announced or to correlate unreachability reports from multiple PE (Provider Edge) routers.

Use cases for EVPN unreachability signaling include but not limited to:

* Multi-tenant network debugging and troubleshooting

* Security event coordination across EVPN instances

* DDoS attack target information sharing without null-routing

* Monitoring tenant prefix health across multiple data centers

* Correlating unreachability from multiple PE vantage points

The goal of this mechanism is to provide comprehensive information about unreachability events:

* Where the event has happened: Reporter Identifier, Reporter AS Number, and optionally EVPN Instance (EVI)

* Why the event has happened: Reason Code indicating the cause of unreachability

* When the event has happened: Timestamp of the unreachability detection

This document defines a new EVPN Route Type that creates a parallel information channel for unreachability data, maintaining complete separation from the forwarding plane. The encoding follows the architecture and terminology of [RFC9136], applying similar concepts to EVPN as defined for standard BGP in [I-D.tantsura-idr-unreachability-safi].

1.1. 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.

1.2. Terminology

EVPN: Ethernet VPN

NVE: Network Virtualization Edge, as defined in [RFC7365] and [RFC9136]. An NVE is a network entity that provides virtualization functions. In this document, NVE and PE are used interchangeably.

PE: Provider Edge router

RT-5: EVPN Route Type 5 (IP Prefix Advertisement Route), as defined in [RFC9136]

Address Family (unreachability NLRI): 1-octet field in the IP Prefix Unreachability Route NLRI, immediately after the Ethernet Tag ID, indicating IPv4 (value 1) or IPv6 (value 2). The values are the existing BGP Address Family Identifier (AFI) numeric values assigned by IANA in the "Address Family Numbers" registry [RFC4760]; this document defines no new values and creates no new registry. See Section 3.3.

UI-RIB: Unreachability Information RIB

NLRI: Network Layer Reachability Information

TLV: Type-Length-Value

Reporter TLV: A nested TLV structure containing information about one Reporting PE and its associated unreachability details

Aggregation: The process of combining multiple Reporter TLVs from different paths into a single NLRI

Advertising PE: The PE router that sends the BGP UPDATE message containing the unreachability information

Reporting PE: The PE router that originally generated the unreachability information (identified within a Reporter TLV)

This document also assumes familiarity with the terminology of [RFC7365], [RFC7432], [RFC8365], and [RFC9136].

2. EVPN Unreachability Route Type

2.1. Route Type Definition

This document defines a new EVPN Route Type:

* Route Type: TBD1 (to be assigned by IANA)

* Name: IP Prefix Unreachability Route Type

* Based on: Route Type 5 field order and definitions from [RFC9136], plus an Address Family octet after the Ethernet Tag ID (not present in reachability RT-5)

This route type is carried in BGP UPDATE messages using the Multiprotocol Extensions for BGP-4 [RFC4760], with AFI = 25 (L2VPN) and SAFI = 70 (EVPN), following the same procedures as other EVPN route types defined in [RFC7432] and [RFC9136].

This route type creates a parallel information plane for unreachability signaling. It MUST NOT affect the EVPN Loc-RIB or forwarding plane in any way. PE routers receiving this route type MUST maintain it in a separate Unreachability Information RIB (UI-RIB) and MUST NOT install or remove routes in the Loc-RIB or forwarding table based on these advertisements.

2.2. Route Distinguisher and Route Targets

The Route Distinguisher (RD) field and Route Target (RT) extended communities operate as defined in [RFC7432] and [RFC9136]. Unreachability routes for an EVPN instance SHOULD use the same RD and RTs as the corresponding reachability routes (Route Type 5), ensuring correlation with the same EVPN instance and following established EVPN patterns.

Implementations MAY use distinct RTs for unreachability routes to limit distribution to specific PEs (e.g., monitoring systems) or to prevent distribution to legacy systems during incremental deployment.

3. NLRI Encoding

3.1. Design Philosophy

The IP Prefix Unreachability Route encoding uses the Route Type 5 field order and definitions from [RFC9136] to maximize consistency with existing EVPN implementations. Implementations can reuse much of existing RT-5 parsing logic but MUST insert handling for the Address Family octet (immediately after the Ethernet Tag ID) before reading the IP Prefix field, because reachability RT-5 has no such field.

This approach provides:

* Structural consistency with Route Type 5 (with the Address Family extension noted above)

* Simplified implementation by following familiar RT-5 field order and semantics aside from that extension

* Extensibility via TLV-based reporter information

* Future-proof design if any currently unused fields become relevant

3.2. Approach to Multiple Reporters

When multiple PE routers report unreachability for the same prefix, implementers have several options:

1. Single Reporter: Do nothing and allow the Reporter Identifier of the best route to be used as the only Reporter. This is the simplest approach but loses information from other reporters. This approach is fully compatible with all EVPN implementations.

2. Nested TLV Aggregation (Recommended): Implement the nested TLV aggregation approach described in this specification to preserve all reporter perspectives in a single NLRI. This provides the most comprehensive view while maintaining a single EVPN route per prefix. This approach is designed specifically for EVPN and does not require additional protocol extensions beyond this specification.

3. BGP ADD-PATH: Use BGP ADD-PATH [RFC7911] to maintain multiple paths, each carrying its own Reporter TLV. ADD-PATH is defined for BGP-4 and can be applied to EVPN route types by prepending a Path Identifier to each NLRI. However, ADD-PATH support for EVPN varies by implementation, and this approach would require both ADD-PATH capability negotiation and proper handling of Path Identifiers with EVPN Route Type structures. This option preserves full BGP path attributes per reporter but has higher complexity and deployment requirements.

This specification focuses on the nested TLV aggregation approach (option 2) as the preferred mechanism, providing detailed procedures and encodings for this method throughout the remainder of this document. Option 2 is recommended because it provides comprehensive multi-reporter visibility while maintaining compatibility with standard EVPN processing and minimizing implementation complexity.

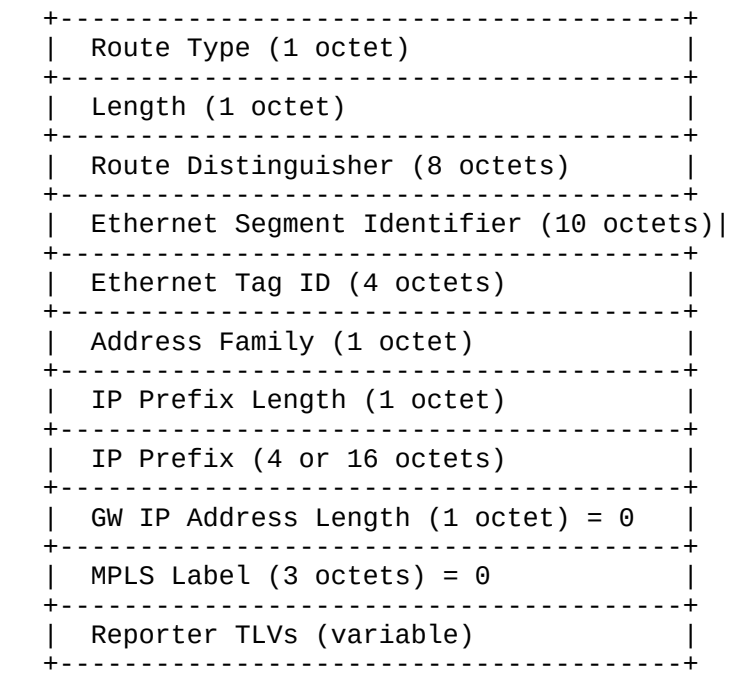
3.3. IP Prefix Unreachability Route NLRI

The IP Prefix Unreachability Route uses the Route Type 5 field order and definitions from [RFC9136], extended with Reporter TLVs and one additional field for address-family disambiguation (see below).

The NLRI is uniquely identified by the combination of Route Distinguisher, Ethernet Tag ID, Address Family, IP Prefix Length, and IP Prefix. Reporter TLVs are NOT part of the NLRI key but provide information about each Reporting PE. The presence of an Unreachability Route for a prefix signifies that one or more PEs report the prefix as unreachable. The withdrawal of such a route indicates that all reporters have cleared their unreachability reports for that prefix.

For reachability RT-5 routes, [RFC9136] fixes the size of the route-type-specific NLRI (34 octets for IPv4 or 58 octets for IPv6), which allows a receiver to infer whether the IP Prefix field is 4 or 16

octets. Because unreachability routes append a variable number of Reporter TLVs, the route-type-specific length is no longer sufficient to infer the address family: for example, an IPv4 prefix with a large set of Reporter TLVs can yield the same total size as a shorter IPv6 encoding. Furthermore, IP Prefix Length alone is ambiguous (e.g., a /24 can be valid for both IPv4 and IPv6). Therefore, this route type includes an explicit Address Family field immediately after the Ethernet Tag ID. Receivers MUST use this field to determine the width of the IP Prefix field before parsing Reporter TLVs.



Where:

* Route Type: TBD1 (IP Prefix Unreachability Route Type)

* Length: As defined in [RFC7432], the number of octets in the Route Type specific field (from Route Distinguisher through the end of the last Reporter TLV). This differs from reachability RT-5 in [RFC9136], where Length is 34 (IPv4) or 58 (IPv6) and does not include Reporter TLVs.

* Route Distinguisher: As defined in [RFC7432] and [RFC9136], used to identify the EVPN instance. The RD MUST be used as specified in those documents.

* Ethernet Segment Identifier: As defined in [RFC7432]. MUST be set to 0 (all bytes zero) for unreachability routes.

* Ethernet Tag ID: As defined in [RFC7432] and [RFC9136], identifies the broadcast domain. For unreachability routes, this SHOULD be set to 0 unless VLAN-aware bundle service is used, following the same conventions as reachability RT-5 routes.

* Address Family: 1 octet. The values are the existing BGP Address Family Identifier (AFI) numeric values assigned by IANA in the "Address Family Numbers" registry [RFC4760]; this document defines no new values and creates no new registry. Value 1 (IPv4, BGP AFI 1) requires an IP Prefix field of 4 octets and IP Prefix Length in the range 0-32 inclusive. Value 2 (IPv6, BGP AFI 2) requires an IP Prefix field of 16 octets and IP Prefix Length in the range 0-128 inclusive. Handling of any other Address Family value is specified in Section 4.10.1.

* IP Prefix Length: Length of the IP prefix in bits. Constraints depend on Address Family as described above.

* IP Prefix: IPv4 (4 octets) or IPv6 (16 octets) prefix being reported as unreachable, consistent with Address Family.

* GW IP Address Length: MUST be set to 0. The GW IP Address field is not used for unreachability routes.

* GW IP Address: MUST be zero octets (GW IP Address Length = 0). Unreachability routes do not use overlay index resolution.

* MPLS Label: 3-octet field where the high-order 20 bits contain the MPLS label value, as specified in [RFC9136]. MUST be set to 0 (reserved). Since unreachability routes are information-only and do not establish forwarding state, and do not use overlay index resolution, the label field has no semantic meaning and MUST be zero.

* Reporter TLVs: One or more Reporter TLVs as defined in Section 3.5

3.4. Field Usage for Information-Only Routes

Since unreachability routes are information-only and do not use overlay indexes for recursive resolution, the following constraints apply to fields inherited from Route Type 5:

* ESI: MUST be 0 (all bytes zero).

* GW IP Address Length: MUST be 0; GW IP Address: zero octets.

* MPLS Label: MUST be 0.

* Address Family: Used only to determine IP Prefix width for NLRI parsing.

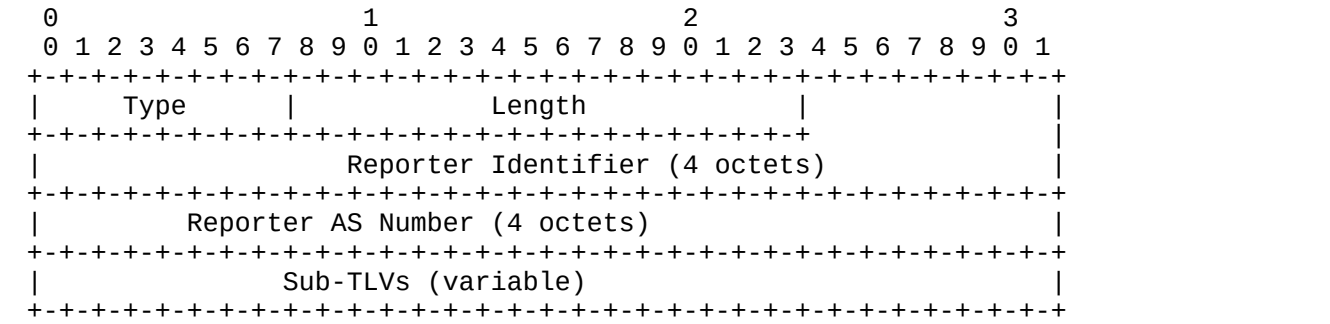
Receiving PEs MUST NOT use any of these fields for forwarding decisions or recursive resolution. ESI, GW IP Address, and MPLS Label are maintained for structural consistency with [RFC9136]; Address Family is an extension defined in this document.

The [RFC9136] Section 3.1 rule that a route with a zero MPLS Label and no Overlay Index MUST be treated as withdrawn, and the Section 3.2 rule that a route with ESI, GW IP, Router's MAC, and MPLS Label all zero SHOULD be treated as withdrawn, apply only to Route Type 5. The IP Prefix Unreachability Route Type defined here uses a distinct EVPN Route Type (Section 3.3) and is not subject to those rules.

Unreachability Routes carry ESI=0 and encode no Ethernet-Segment membership. They signal IP-prefix unreachability from the Reporting PE's local perspective. DF election, Ethernet-Segment failure, and non-DF transitions do not in themselves trigger Unreachability Routes; the multi-homing procedures of [RFC7432] and [RFC9136] are unchanged.

3.5. Reporter TLV Format

The Reporter TLV encapsulates information about a single Reporting PE router. Multiple Reporter TLVs may be included in a single NLRI to support aggregation of reports from different network vantage points.



Reporter TLV Fields:

Type: 1 octet. Value: 1 (Reporter)

Length: 2 octets. Total length of Reporter Identifier, Reporter AS

Number, and Sub-TLVs fields in octets (minimum 8 octets when no Sub-TLVs are present)

Reporter Identifier: 4 octets. BGP Identifier (Router ID) of the Reporting PE in network byte order

Reporter AS Number: 4 octets. 4-octet AS number of the Reporting PE in network byte order. If the AS number is less than 65536, the upper 2 octets are set to 0.

Sub-TLVs: Variable length. Contains zero or more Sub-TLVs providing additional context about the unreachability event. A Reporter TLV carrying no Sub-TLVs is valid: the presence of the Reporter TLV itself conveys the fact of unreachability, with the Reporter Identifier and Reporter AS Number identifying the PE that observed it. When the Unreachability Reason Code Sub-TLV (Section 3.6.1) is absent, the reason is treated as Unspecified (code 0).

The combination of Reporter Identifier and Reporter AS Number uniquely identifies the Reporting PE. Multiple Reporter TLVs with the same Reporter Identifier and AS Number MUST NOT appear in the same NLRI. If such duplication occurs, only the first occurrence SHOULD be processed.

Except that the first Reporter TLV in an NLRI corresponds to the best path (see Section 4.4), the order of subsequent Reporter TLVs is not significant; receivers MUST NOT derive meaning from their relative ordering. Implementations MUST tolerate any ordering of Reporter TLVs past the first position.

3.6. Sub-TLV Types

3.6.1. Unreachability Reason Code Sub-TLV

* Sub-Type: 1
* Sub-Length: 2 octets
* Sub-Value: 2-octet reason code in network byte order

Defined Reason Codes:

* 0: Unspecified
* 1: Policy Blocked
* 2: Security Filtered

* 3: RPKI Invalid
* 4: No Export Policy
* 5: Martian Address
* 6: Bogon Prefix
* 7: Route Dampening
* 8: Local Administrative Action
* 9: Local Link Down
* 10: MAC Mobility Limit Exceeded (EVPN-specific)
* 11: Tenant Isolation Violation (EVPN-specific)
* 12: VTEP Unreachable (EVPN-specific)
* 13-64535: Reserved for Future Use (IANA allocation required)
* 64536-65535: Reserved for Private/Experimental Use

Reason codes 0-9 align with the BGP Unreachability Information SAFI [I-D.tantsura-idr-unreachability-safi] for consistency across standard BGP and EVPN unreachability signaling. Reason codes 10-12 are EVPN-specific extensions.

3.6.2. Timestamp Sub-TLV

* Sub-Type: 2
* Sub-Length: 8 octets
* Sub-Value: Unix timestamp (seconds since epoch) in network byte order, indicates when the unreachability event occurred or was detected by this reporter

3.6.3. EVI Sub-TLV

* Sub-Type: 3
* Sub-Length: 4 octets
* Sub-Value: EVPN Instance (EVI) identifier where the unreachability was observed

Since the Route Distinguisher in the NLRI already identifies the EVPN instance in most deployments, this Sub-TLV SHOULD only be included when additional EVI-specific information is necessary that cannot be derived from the RD.

Examples of when EVI Sub-TLV may be useful:

* Multiple EVIs share the same RD (non-recommended configuration)

* Correlation with locally-significant EVI identifiers

* Debugging scenarios requiring explicit EVI identification

Omitting this Sub-TLV when not needed saves 7 octets (3 octets TLV overhead + 4 octets EVI value) per Reporter TLV.

3.6.4. Sub-TLV Processing Rules

Implementations MUST be prepared to receive Sub-TLVs in any order. Unknown Sub-TLV types MUST be silently ignored to allow for future extensibility.

The Reason Code Sub-TLV SHOULD be included in all Reporter TLVs. If absent, implementations SHOULD treat it as Reason Code 0 (Unspecified).

4. Operation

4.1. Forwarding Plane Separation

This route type creates a parallel information plane that operates independently of the EVPN forwarding plane. Implementations MUST maintain strict separation between unreachability information and forwarding decisions.

Specifically, implementations MUST:

* NOT install or remove any routes in the Loc-RIB or forwarding table based on unreachability information

* Maintain unreachability routes in a separate Unreachability Information RIB (UI-RIB) that is not consulted for forwarding decisions

* NOT modify forwarding table entries (including next-hop, label, or other attributes) based on unreachability information

* Process unreachability routes with lower priority than forwarding routes to prevent resource contention

* Implement separate rate limiting for unreachability routes

Violation of these separation requirements could lead to incorrect forwarding behavior, traffic blackholing, or routing instability. Implementers MUST ensure proper separation through careful software architecture and testing.

4.2. Generating Unreachability Routes

A PE MAY generate an IP Prefix Unreachability Route when local processing determines, for any reason, that a prefix is to be reported as unreachable. The triggering condition is conveyed by the Reason Code Sub-TLV (Section 3.6.1). This document does not mandate the set of local conditions that cause generation; that set is implementation- and deployment-specific, constrained only by the Reason Codes defined in this document or subsequently registered.

The Reporting PE MUST set Address Family to 1 (IPv4) or 2 (IPv6) consistent with the IP Prefix field. The Reporting PE MUST populate the Reporter TLV with its own BGP Identifier and AS Number. The PE SHOULD include a Reason Code Sub-TLV and SHOULD include a Timestamp Sub-TLV to facilitate temporal correlation.

Implementations SHOULD provide configuration options to control:

* Which events trigger unreachability route generation

* Rate limiting on route generation per prefix and globally

* Filtering criteria for which prefixes can be reported

* Default Reason Codes for different trigger conditions

4.3. Processing Unreachability Routes

When a PE router receives an IP Prefix Unreachability Route:

1. It MUST validate that the route type is recognized

2. It MUST parse the NLRI using the Address Family field to determine whether the IP Prefix is 4 or 16 octets, then parse Reporter TLVs according to Section 3. Error handling for invalid Address Family or IP Prefix Length values is specified in Section 4.10.1.

3. It MUST NOT install or remove any routes in the Loc-RIB or forwarding table based on this route

4. It MUST maintain a separate UI-RIB for unreachability routes

5. It SHOULD apply standard BGP path selection to UI-RIB entries for consistency

6. It MAY propagate the route according to standard EVPN rules and local policy

7. It MAY aggregate Reporter TLVs as described in Section 4.4

8. It SHOULD make unreachability information available to management systems and monitoring tools

Unknown Sub-TLV types within Reporter TLVs MUST be silently ignored to allow for future extensibility.

4.4. Aggregation Procedures

When multiple routes arrive for the same prefix (identified by RD, Ethernet Tag ID, Address Family, IP Prefix Length, and IP Prefix), a PE supporting aggregation SHOULD combine Reporter TLVs from multiple paths into a single advertisement.

Aggregation procedure:

1. Perform standard BGP path selection based on BGP attributes (NOT Reporter TLV content) to select the best path

2. Extract Reporter TLVs from the best path

3. For each non-selected feasible path, extract Reporter TLVs and add unique reporters (by Reporter Identifier and Reporter AS) to the aggregated set. If a reporter already exists, keep the entry with the most recent timestamp (if present).

4. Create a new NLRI with all unique Reporter TLVs

5. Advertise the aggregated NLRI using BGP attributes from the best path

A speaker performing Reporter TLV aggregation MUST place the Reporter TLV corresponding to the best path in the first position of the resulting NLRI. Reporter TLVs drawn from non-selected feasible paths MAY follow in any order. Pinning the first position provides a deterministic fallback for speakers that do not perform aggregation (see below).

A receiver MUST parse all Reporter TLVs present in a received NLRI, up to the implementation limit defined in this section (RECOMMENDED 50). How the received Reporter TLVs are consumed locally -- stored in the UI-RIB, exported via BMP, displayed to operators -- is an implementation matter and does not affect wire behavior.

A speaker that does not perform Reporter TLV aggregation, when re-advertising an IP Prefix Unreachability Route to its peers, MUST include only the first Reporter TLV from the received NLRI and MUST NOT append Reporter TLVs drawn from other paths. Because EVPN does not negotiate per-Route-Type capabilities, this rule -- together with the sender ordering rule above -- constitutes the interoperability contract between aggregating and non-aggregating speakers: at the first non-aggregating hop, the propagated view degrades to the best-path Reporter TLV only, with no loss of correctness.

The maximum number of Reporter TLVs per route SHOULD be limited to prevent excessive route sizes. RECOMMENDED maximum: 50 Reporter TLVs per route.

If the maximum is reached and a new reporter must be added, implementations SHOULD remove the oldest Reporter TLV based on Timestamp Sub-TLV (if present). The reporter from the best path MUST NOT be removed; if it is the oldest, remove the second-oldest instead.

4.5. Withdrawal Procedures

Withdrawal of unreachability information operates at two levels:

4.5.1. Individual Reporter Withdrawal

When a PE determines that a specific reporter no longer considers a prefix unreachable (e.g., receives an UPDATE from that reporter’s PE that does not include the unreachability NLRI, or local policy determines the report is stale), it SHOULD:

1. Remove the corresponding Reporter TLV from the NLRI

2. If other Reporter TLVs remain, re-advertise the NLRI with the remaining Reporter TLVs

3. If no Reporter TLVs remain, withdraw the entire NLRI as described below

To facilitate individual reporter withdrawal, implementations MUST track the source of each Reporter TLV (which BGP neighbor or local process it came from).

4.5.2. Complete NLRI Withdrawal

A PE MUST withdraw an Unreachability Route (send the NLRI key fields in MP_UNREACH_NLRI) when:

* All Reporter TLVs have been removed
* The prefix is explicitly withdrawn by all upstream sources
* Local policy dictates the information should no longer be propagated

The MP_UNREACH_NLRI contains the NLRI fields (Route Distinguisher, Ethernet Segment Identifier, Ethernet Tag ID, Address Family, IP Prefix Length, IP Prefix, GW IP Address Length, and MPLS Label) without any Reporter TLVs.

4.5.3. Stale Reporter Detection

Implementations MAY implement aging mechanisms to remove stale Reporter TLVs:

* If a Timestamp Sub-TLV is present and indicates the report is older than a configurable threshold (RECOMMENDED default: 24 hours), the Reporter TLV MAY be removed
* If the BGP session to the neighbor that provided a Reporter TLV goes down, implementations MAY mark associated Reporter TLVs as potentially stale and MAY remove them after a grace period

4.6. Interaction with Graceful Restart

BGP Graceful Restart [RFC4724] applies to the EVPN SAFI (AFI=25, SAFI=70) and thus to Unreachability Routes. GR procedures for other EVPN route types ([RFC7432], [RFC9136]) are unchanged.

Unreachability Routes follow the same GR procedures as RT-5 [RFC9136]: one GR Capability for SAFI=70, one End-of-RIB (EoR) marker, one Restart Time. They are marked stale, retained, refreshed, and removed identically. The sole departure is the "Forwarding State" (F) bit. Unreachability Routes install no

forwarding state (Section 4.1); the F bit has no meaning for this Route Type. Its interpretation for forwarding-state-bearing route types is unchanged.

4.6.1. Graceful Restart Capability

No new capability is defined. GR is negotiated using the EVPN AFI/SAFI (25/70).

Implementations MUST NOT treat any F bit value as indicating forwarding-state preservation for Unreachability Routes. The F bit advertised for SAFI=70 is determined by the preservation properties of the other route types carried in the SAFI.

4.6.2. Restarting Speaker Behavior

A PE that has negotiated GR for SAFI=70:

1. SHOULD re-advertise its preserved Unreachability Information RIB (UI-RIB) as soon as practicable; gradual re-advertisement is permitted to limit burstiness.

2. If the UI-RIB was not preserved, SHOULD rebuild it from local sources (link-down state, policy decisions) before re-advertising.

3. MUST send the SAFI=70 EoR marker [RFC4724] after re-advertisement completes. This marker covers all EVPN route types in the SAFI; no Route-Type-specific EoR is defined.

4.6.3. Receiving Speaker Behavior

On detection of peer restart:

1. All Unreachability Routes from the restarting peer MUST be marked stale, irrespective of the F bit.

2. Stale routes MUST NOT be withdrawn. They MUST be retained until the SAFI=70 EoR is received or the peer’s Restart Time expires, whichever occurs first.

3. While stale, the routes MAY be used for monitoring and correlation, MAY be distinguished in display and APIs, and SHOULD NOT be propagated to other peers absent explicit configuration.

4. On receipt of the EoR:

    * unrefreshed stale routes MUST be removed;

* Reporter TLVs from the restarted peer within aggregated NLRIs MUST be removed if not refreshed;

* if Reporter TLVs from other sources remain for the same NLRI key, the route SHOULD be re-advertised with those remaining TLVs (Section 4.4).

5. If the Restart Time expires before the EoR arrives, all stale routes MUST be removed.

4.6.4. Route Reflector Considerations

A Route Reflector participating in GR for SAFI=70:

* MUST stale-mark Unreachability Routes from restarting clients identically to other EVPN route types;

* SHOULD NOT reflect stale Unreachability Routes absent an explicit reflection policy;

* MUST preserve ORIGINATOR_ID and CLUSTER_LIST across stale-to-refreshed transitions;

* SHOULD send the SAFI=70 EoR to clients after completing its own restart processing.

4.6.5. Implementation Recommendations

Restart Time is shared with the rest of SAFI=70. Operators SHOULD account for the additional UI-RIB re-advertisement volume when tuning it.

Implementations SHOULD expose:

* UI-RIB preservation, toggled independently of other EVPN route types;

* propagation of stale Unreachability Routes during GR;

* action on EoR timeout.

Implementations SHOULD log, per GR cycle: peer-restart detection affecting the UI-RIB, stale marking, EoR receipt, and stale removal.

4.7. Preventing State Explosion

To prevent unbounded growth of the UI-RIB, implementations SHOULD enforce the following limits:

* Maximum Reporter TLVs per route (RECOMMENDED: 50)

* Maximum total UI-RIB routes (SHOULD be configurable; RECOMMENDED default: 100,000)

* Rate limiting on accepting new unreachability routes

* Per-peer rate limiting

When limits are reached, implementations SHOULD log the event, apply aging policies to remove oldest entries, and continue accepting withdrawals to allow state to decrease.

4.8. Relationship to BGP Route Damping

Unreachability routes SHOULD NOT be subject to standard BGP route damping mechanisms since they do not affect forwarding and represent information that operators explicitly want to propagate.

However, implementations MAY implement rate limiting specific to unreachability routes to prevent:

* UI-RIB resource exhaustion

* Excessive BGP UPDATE message generation

* Processing overhead on Receiving PEs

* Malicious or misconfigured PEs flooding the network

Rate limiting should be applied at:

* Route generation (at Reporting PE)

* Route acceptance (at Receiving PE)

* Per-peer basis

* Global aggregate level

4.9. Path Selection for Aggregation

Path selection for Unreachability Routes follows standard BGP best path selection ([RFC7432] Section 15, incorporating [RFC9136]) with the following clarifications:

* Weight/Local Preference: Apply normally based on local policy.

* AS_PATH Length: Shorter AS_PATH is preferred. This represents the path the UPDATE message took.

* ORIGIN: IGP preferred over EGP over INCOMPLETE.

* MED: Apply if comparing paths from the same neighboring AS.

* BGP Identifier: Use for tie-breaking.

The content of Reporter TLVs (number of reporters, reason codes, timestamps, etc.) MUST NOT influence path selection. Path selection determines which UPDATE's BGP attributes are used for propagation, while aggregation combines Reporter TLVs from multiple paths.

4.10. Communities and Attributes

Standard BGP communities and attributes apply to the UPDATE message carrying Unreachability Routes:

* NO_EXPORT, NO_ADVERTISE, and NO_EXPORT_SUBCONFED work as defined in their respective specifications

* Large Communities [RFC8092] MAY be used for policy control of aggregation behavior

* AS_PATH is constructed normally for the UPDATE message path

* ORIGIN SHOULD be set to INCOMPLETE for locally generated unreachability information, reflecting that the information does not originate from routing protocol state

* The Router's MAC Extended Community has no effect on Unreachability Routes; overlay index semantics do not apply (Section 3.4). Senders SHOULD NOT attach it; receivers MUST ignore it if present.

These attributes represent the path taken by the UPDATE message itself, not the paths of individual reporters (which are preserved in Reporter TLVs).

4.11. Error Handling

Error handling for IP Prefix Unreachability Routes follows [RFC7606] and [I-D.ietf-bess-rfc7432bis] Section 7.14.1. Per-class actions are specified in Sections 4.10.1 through 4.10.4. Per [I-D.ietf-bess-rfc7432bis] Section 7.14, "session reset" MAY be replaced with "AFI/SAFI disable" behavior where supported. Checks in Section 4.10.1 are performed before those in Section 4.10.2; on first

error the corresponding action MUST be taken and further NLRI parsing MUST cease. All error conditions MUST be logged.

4.11.1. NLRI Structural Errors

A receiver MUST apply "session reset" per [I-D.ietf-bess-rfc7432bis] Section 7.14.1 on:

* Length below the minimum for this route type.

* Length inconsistent with the enclosing MP_REACH_NLRI or MP_UNREACH_NLRI attribute.

* Address Family other than 1 or 2: the IP Prefix field boundary cannot be determined, and subsequent Key fields cannot be parsed.

* IP Prefix Length outside the range permitted for the indicated Address Family: the prefix cannot be unambiguously constructed and is not suitable as a lookup key.

4.11.2. Non-Key Field Errors

A receiver MUST apply "treat-as-withdraw" per [I-D.ietf-bess-rfc7432bis] Section 7.14.1 on:

* Non-zero Ethernet Segment Identifier, GW IP Address Length, GW IP Address, or MPLS Label (Section 3.4 requires each to be zero).

* No Reporter TLVs present (Section 3.3 requires one or more).

4.11.3. Reporter TLV Errors

Per [RFC9552] Section 5.1, unknown or malformed Reporter TLVs MUST NOT cause the NLRI to be considered malformed.

* Unrecognized TLV Type: the TLV is ignored; parsing resumes at the next TLV boundary.

* Malformed TLV (Length inconsistent with remaining NLRI data or below the 8-octet minimum): the TLV MUST be discarded. If well-formed Reporter TLVs remain, they are processed; otherwise "treat-as-withdraw" MUST be applied.

* Duplicate Reporter TLVs (same Identifier and AS Number): only the first is processed (Section 3.5).

* Reporter TLV count exceeding the implementation limit (Section 4.6): excess TLVs MUST be discarded.

4.11.4. Sub-TLV Errors

An unrecognized Sub-TLV Type within a Reporter TLV is silently ignored (Section 3.6.4). A Sub-TLV whose Length is inconsistent with available data MUST be discarded; processing of the enclosing Reporter TLV and remaining Sub-TLVs continues.

5. Interoperability Considerations

5.1. Incremental Deployment

The IP Prefix Unreachability Route Type can be deployed incrementally without requiring network-wide upgrades:

* PEs that don’t support this route type will ignore it per standard BGP behavior (unknown route type handling)
* Mixed environments with supporting and non-supporting PEs work correctly
* The feature can be enabled on specific EVPN instances for testing before broader deployment
* Aggregation support is OPTIONAL; PEs that do not implement aggregation can still propagate single-reporter routes
* Distinct Route Targets allow control over which PEs receive unreachability information

5.2. Interaction with Route Reflectors

Route Reflectors process Unreachability Routes like any other EVPN route type:

* Apply standard route reflection rules
* ORIGINATOR_ID and CLUSTER_LIST attributes apply normally to the UPDATE message, not to individual reporters
* Route Reflectors SHOULD support aggregation to combine reports from multiple clients
* When reflecting to clients, include all aggregated Reporter TLVs

The distinction between the ORIGINATOR_ID BGP attribute and the Reporter Identifier field in Reporter TLVs is important:

* ORIGINATOR_ID identifies the originator of the BGP UPDATE message for loop prevention

* Reporter Identifier identifies the PE that observed and reported the unreachability condition

* These MAY be different in aggregated scenarios

Route Reflectors that do not support aggregation will still properly reflect unreachability routes using standard route reflection procedures. In this case, only the best path's Reporter TLV(s) will be visible to clients.

Operators deploying this feature SHOULD enable aggregation on Route Reflectors to maximize the utility of multi-vantage-point reporting.

5.3. Interaction with Other EVPN Route Types

IP Prefix Unreachability Routes are completely independent from other EVPN route types. Specifically:

* A Route Type 5 (IP Prefix Advertisement) and an IP Prefix Unreachability Route for the same logical prefix (same Route Distinguisher, Ethernet Tag ID, IP Prefix Length, and IP Prefix value, with the unreachability Address Family value matching the IPv4 or IPv6 encoding of that RT-5 per [RFC9136]) are independent and may coexist

* Presence of an unreachability route does NOT imply absence of a reachability route

* Withdrawal of a reachability route does NOT automatically generate an unreachability route

* BGP path selection is performed independently for each route type

This independence is crucial for maintaining forwarding plane separation and allowing unreachability signaling for prefixes that were never advertised as reachable.

6. Deployment Considerations

6.1. Use Cases

Multi-Tenant Data Centers: Share unreachability information across tenant networks for coordinated security response without affecting tenant traffic forwarding

DCI (Data Center Interconnect): Correlate unreachability reports from multiple data centers to distinguish between local issues and global prefix problems

Security Monitoring: Track suspicious prefix patterns across EVPN instances, coordinate DDoS response, and share threat intelligence

Troubleshooting: Debug prefix reachability issues without impacting production forwarding, identify asymmetric reachability, and correlate with overlay network health

Compliance and Auditing: Maintain records of unreachability events for compliance purposes and SLA verification

Policy-Driven Actions: Trigger an action, as defined by a local policy, in response to received unreachability information (e.g., traffic engineering adjustments, alerting, or logging)

6.2. Operational Recommendations

* Enable aggregation on Route Reflectors to maximize visibility while minimizing route count

* Include Timestamp Sub-TLVs for temporal correlation

* Monitor UI-RIB size for capacity planning

* Test the feature on non-production EVPN instances before production deployment

Implementations SHOULD provide management interfaces to query the UI-RIB, display reporters per prefix, and export unreachability data to external monitoring systems.

7. Security Considerations

This route type SHOULD only be enabled between trusted BGP peers. The trust model is similar to that required for standard EVPN route types.

The following threats are specific to unreachability signaling:

1. Information Leakage: Unreachability information may reveal network topology or operational issues. Operators SHOULD use Route Target filtering to restrict distribution.

2. State Exhaustion: Malicious peers could exhaust UI-RIB memory. Implementations SHOULD enforce the limits described in Section 4.4 and the Preventing State Explosion section.

3. False Information: Peers could advertise false unreachability data. Since this SAFI does not affect forwarding, the impact is limited to monitoring systems.

4. Reporter Impersonation: A peer could include Reporter TLVs claiming to represent other PEs. Implementations SHOULD validate that Reporter AS Numbers are consistent with the AS_PATH of the UPDATE that introduced them.

5. Aggregation Amplification: A peer could send many UPDATEs with different Reporter TLVs for the same prefix. Reporter TLV limits and rate limiting mitigate this.

6. Tenant Isolation: Improper RT configuration could leak unreachability information between tenants. Use separate RDs and tenant-specific RTs.

Operators SHOULD use BGP session security (TCP-AO per [RFC5925]), validate Reporter Identifiers against known PE lists, configure per-peer rate limits, and maintain audit logs of unreachability route updates.

8. IANA Considerations

8.1. EVPN Route Type

IANA is requested to assign a new EVPN Route Type value from the "EVPN Route Types" registry within the "Border Gateway Protocol (BGP) Parameters" registry group:

* Value: TBD1

* Description: IP Prefix Unreachability Route Type

* Reference: This document

8.2. EVPN Unreachability Reporter TLV Types

IANA is requested to create a new registry called "EVPN Unreachability Reporter TLV Types" under the "Border Gateway Protocol (BGP) Parameters" registry page.

Registration Procedure: Standards Action

Initial registrations:

Value	Description	Reference
0	Reserved	This document
1	Reporter TLV	This document
2-254	Unassigned	
255	Reserved	This document

8.3. EVPN Unreachability Sub-TLV Types

IANA is requested to create a new registry called "EVPN Unreachability Sub-TLV Types" under the "Border Gateway Protocol (BGP) Parameters" registry page.

Registration Procedure: RFC Required

Initial registrations:

Value	Description	Reference
0	Reserved	This document
1	Unreachability Reason Code	This document
2	Timestamp	This document
3	EVI	This document
4-254	Unassigned	
255	Reserved	This document

8.4. EVPN Unreachability Reason Codes

IANA is requested to create a new registry called "EVPN Unreachability Reason Codes" under the "Border Gateway Protocol (BGP) Parameters" registry page.

Registration Procedure: RFC Required for values 0-64535, Reserved for Private Use for values 64536-65535

Initial registrations:

Value	Description	Reference
0	Unspecified	This document
1	Policy Blocked	This document
2	Security Filtered	This document
3	RPKI Invalid	This document
4	No Export Policy	This document
5	Martian Address	This document
6	Bogon Prefix	This document
7	Route Dampening	This document
8	Local Administrative Action	This document
9	Local Link Down	This document
10	MAC Mobility Limit Exceeded	This document
11	Tenant Isolation Violation	This document
12	VTEP Unreachable	This document
13-64535	Unassigned	
64536-65535	Reserved for Private Use	This document

9. Acknowledgements

The authors would like to thank the BESS working group for their valuable feedback and suggestions on this proposal. Special thanks to the EVPN protocol designers whose work on RFC 7432, RFC 9136, and related specifications provided the foundation for this extension.

The aggregation mechanism in this specification draws inspiration from similar multi-reporter approaches in other monitoring and troubleshooting protocols.

10. References

10.1. Normative References

[I-D.ietf-bess-rfc7432bis]
Sajassi, A., Ed., Burdet, L., Ed., Drake, J., and J. Rabadan, "BGP MPLS-Based Ethernet VPN", Work in Progress, Internet-Draft, draft-ietf-bess-rfc7432bis-14, March 2026, <https://datatracker.ietf.org/doc/draft-ietf-bess-rfc7432bis/>.

[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>.

[RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y. Rekhter, "Graceful Restart Mechanism for BGP", RFC 4724, DOI 10.17487/RFC4724, January 2007, <https://www.rfc-editor.org/info/rfc4724>.

[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>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[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>.

[RFC9136] Rabadan, J., Ed., Henderickx, W., Drake, J., Lin, W., and A. Sajassi, "IP Prefix Advertisement in Ethernet VPN (EVPN)", RFC 9136, DOI 10.17487/RFC9136, October 2021, <https://www.rfc-editor.org/info/rfc9136>.

[RFC9552] Talaulikar, K., Ed., "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, January 2024, <https://www.rfc-editor.org/info/rfc9552>.

10.2. Informative References

[I-D.tantsura-idr-unreachability-safi] Tantsura, J., Sharp, D., Venkatraman, V., Muppalla, K., and M. Rzehak, "BGP Unreachability Information SAFI", Work in Progress, Internet-Draft, draft-tantsura-idr-unreachability-safi-03, April 2026, <https://datatracker.ietf.org/doc/html/draft-tantsura-idr-unreachability-safi-03>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <https://www.rfc-editor.org/info/rfc5925>.

[RFC7365] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for Data Center (DC) Network Virtualization", RFC 7365, DOI 10.17487/RFC7365, October 2014, <https://www.rfc-editor.org/info/rfc7365>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <https://www.rfc-editor.org/info/rfc7911>.

[RFC8092] Heitz, J., Ed., Snijders, J., Ed., Patel, K., Bagdonas, I., and N. Hilliard, "BGP Large Communities Attribute", RFC 8092, DOI 10.17487/RFC8092, February 2017, <https://www.rfc-editor.org/info/rfc8092>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Isaac, A., Henderickx, W., and R. Shekhar, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

Appendix A: Encoding Examples

Example 1: Minimal Unreachability Route

IPv4 tenant prefix 192.0.2.0/24 reported unreachable by a single leaf PE because egress export policy suppresses the prefix (Reason Code 4, "No Export Policy"):

```
   Route Type: TBD1
   Length: 48 octets (route-type-specific)
   Route Distinguisher: 198.51.100.1:100 (RD Type 1, 8 octets)
   Ethernet Segment Identifier: 0 (10 octets)
   Ethernet Tag ID: 0 (4 octets)
   Address Family: 1 (IPv4)
   IP Prefix Length: 24 (1 octet)
   IP Prefix: 192.0.2.0 (4 octets)
   GW IP Address Length: 0 (1 octet)
   MPLS Label: 0 (3 octets)

   Reporter TLV (on-wire: 16 octets = 1 Type + 2 Length + 13 payload):
     Type: 1
     Length: 13 (payload octets)
     Reporter Identifier: 198.51.100.1 (4 octets)
     Reporter AS: 65001 (4 octets)
     Sub-TLV Reason Code:
       Sub-Type: 1
       Sub-Length: 2
       Value: 4 (No Export Policy)

   Route-type-specific = 8 + 10 + 4 + 1 + 1 + 4 + 1 + 3 + 16 = 48 octets
   NLRI (with Route Type + Length) = 2 + 48 = 50 octets

   Hexadecimal encoding (TT = TBD1 Route Type):
   TT 30 00 01 C6 33 64 01 00 64 00 00 00 00 00 00
   00 00 00 00 00 00 00 00 01 18 C0 00 02 00 00 00
   00 00 01 00 0D C6 33 64 01 00 00 FD E9 01 00 02
   00 04
```

   See Example 3 for the equivalent IPv6 encoding.

Example 2: Aggregated Route with Multiple Reporters

   IPv4 tenant prefix 198.51.100.0/24 reported by three leaf PEs in the same DC fabric following a coordinated administrative action (Reason Code 8, "Local Administrative Action"). Timestamps are spaced by a few seconds, consistent with propagation of the administrative event across the fabric:

```
   Route Type: TBD1
   Length: 113 octets (route-type-specific)
   Route Distinguisher: 198.51.100.1:100
   Ethernet Segment Identifier: 0
   Ethernet Tag ID: 0
   Address Family: 1 (IPv4)
   IP Prefix Length: 24
   IP Prefix: 198.51.100.0
```

```
   GW IP Address Length: 0
   MPLS Label: 0

   Reporter TLV #1 (on-wire: 27 octets):
     Type: 1, Length: 24 (payload)
     Reporter Identifier: 198.51.100.1
     Reporter AS: 65001
     Sub-TLVs:
       Reason Code (Type 1, Length 2): 8 (Local Administrative Action)
       Timestamp (Type 2, Length 8): 1704672000

   Reporter TLV #2 (on-wire: 27 octets):
     Type: 1, Length: 24
     Reporter Identifier: 198.51.100.2
     Reporter AS: 65001
     Sub-TLVs:
       Reason Code (Type 1, Length 2): 8
       Timestamp (Type 2, Length 8): 1704672005

   Reporter TLV #3 (on-wire: 27 octets):
     Type: 1, Length: 24
     Reporter Identifier: 198.51.100.3
     Reporter AS: 65001
     Sub-TLVs:
       Reason Code (Type 1, Length 2): 8
       Timestamp (Type 2, Length 8): 1704672008

   Fields through MPLS Label: 32 octets
   Reporter TLVs on wire: 3 x 27 = 81 octets
   Route-type-specific total: 32 + 81 = 113 octets
   NLRI (with Route Type + Length) = 2 + 113 = 115 octets

   Hexadecimal encoding (TT = TBD1 Route Type):
   TT 71 00 01 C6 33 64 01 00 64 00 00 00 00 00 00
   00 00 00 00 00 00 00 00 01 18 C6 33 64 00 00 00
   00 00 01 00 18 C6 33 64 01 00 00 FD E9 01 00 02
   00 08 02 00 08 00 00 00 00 65 9B 3B 00 01 00 18
   C6 33 64 02 00 00 FD E9 01 00 02 00 08 02 00 08
   00 00 00 00 65 9B 3B 05 01 00 18 C6 33 64 03 00
   00 FD E9 01 00 02 00 08 02 00 08 00 00 00 00 65
   9B 3B 08
```

Example 3: IPv6 Unreachability Route

IPv6 tenant prefix 2001:db8::/32 reported unreachable by a leaf PE following a local CE-facing link-down event (Reason Code 9, "Local Link Down"):

```
  Route Type: TBD1
  Length: 60 octets (route-type-specific)
  Route Distinguisher: 198.51.100.1:100 (RD Type 1, 8 octets)
  Ethernet Segment Identifier: 0 (10 octets)
  Ethernet Tag ID: 0 (4 octets)
  Address Family: 2 (IPv6)
  IP Prefix Length: 32 (1 octet)
  IP Prefix: 2001:db8:: (16 octets)
  GW IP Address Length: 0 (1 octet)
  MPLS Label: 0 (3 octets)

  Reporter TLV (on-wire: 16 octets = 1 Type + 2 Length + 13 payload):
    Type: 1
    Length: 13 (payload octets)
    Reporter Identifier: 198.51.100.1 (4 octets)
    Reporter AS: 65001 (4 octets)
    Sub-TLV Reason Code:
      Sub-Type: 1
      Sub-Length: 2
      Value: 9 (Local Link Down)

  Route-type-specific = 8 + 10 + 4 + 1 + 1 + 16 + 1 + 3 + 16 = 60 octets
  NLRI (with Route Type + Length) = 2 + 60 = 62 octets

  Hexadecimal encoding (TT = TBD1 Route Type):
  TT 3C 00 01 C6 33 64 01 00 64 00 00 00 00 00 00
  00 00 00 00 00 00 00 00 02 20 20 01 0D B8 00 00
  00 00 00 00 00 00 00 00 00 00 00 00 00 01 00
  0D C6 33 64 01 00 00 FD E9 01 00 02 00 09
```

Example 4: Complete BGP UPDATE Message

A PE (198.51.100.1, AS 65001) advertises that 192.0.2.0/24 is unreachable, with Reason RPKI Invalid and a detection timestamp:

```
   BGP UPDATE Message:
     Withdrawn Routes Length: 0
     Total Path Attribute Length: (calculated)

     Path Attributes:
       ORIGIN (Type 1):
         Value: INCOMPLETE (2)

       AS_PATH (Type 2):
         Segment Type: AS_SEQUENCE
         Segment Length: 1
         AS: 65001

       MP_REACH_NLRI (Type 14, Flags 0x90):
         AFI: 25 (L2VPN)
         SAFI: 70 (EVPN)
         Next Hop Length: 4
         Next Hop: 198.51.100.1
         Reserved: 0
         NLRI:
           Route Type: TBD1 (IP Prefix Unreachability)
           Length: 59
           Route Distinguisher: 198.51.100.1:100
           Ethernet Segment Identifier: 0
           Ethernet Tag ID: 0
           Address Family: 1 (IPv4)
           IP Prefix Length: 24
           IP Prefix: 192.0.2.0
           GW IP Address Length: 0
           MPLS Label: 0
           Reporter TLV:
             Type: 1
             Length: 24
             Reporter Identifier: 198.51.100.1
             Reporter AS: 65001
             Sub-TLV (Reason):
               Sub-Type: 1
               Sub-Length: 2
               Value: 3 (RPKI Invalid)
             Sub-TLV (Timestamp):
               Sub-Type: 2
               Sub-Length: 8
               Value: 1733789400

       EXTENDED_COMMUNITIES (Type 16):
         Route Target: 65001:100
```

Example 5: Aggregation at a Receiving PE

Router R1 receives two UPDATEs for the same Unreachability NLRI key (RD 198.51.100.1:100, Ethernet Tag 0, IPv4, 192.0.2.0/24) from different upstream neighbors. R1 selects a best path by standard BGP procedure, extracts Reporter TLVs from the best path plus feasible paths, de-duplicates by (Reporter Identifier, Reporter AS), and re-advertises a single aggregated NLRI.

```
   UPDATE 1 (from Neighbor N1, AS 65100):
     AFI: 25, SAFI: 70
     AS_PATH: 65100
     NLRI (192.0.2.0/24 Unreachability):
       Reporter TLV:
         Reporter ID: 198.51.100.1, AS: 65001
         Reason: RPKI Invalid (3)
         Timestamp: 1733789400

   UPDATE 2 (from Neighbor N2, AS 65200):
     AFI: 25, SAFI: 70
     AS_PATH: 65200
     NLRI (192.0.2.0/24 Unreachability):
       Reporter TLV:
         Reporter ID: 198.51.100.2, AS: 65002
         Reason: Policy Blocked (1)
         Timestamp: 1733789410

   R1 Path Selection:
     - Compare AS_PATH length: both length 1
     - Compare by BGP Identifier: UPDATE 1 wins

   R1 Aggregation:
     - Extract Reporter TLV from UPDATE 1 (best path)
     - Extract Reporter TLV from UPDATE 2 (feasible path)
     - No duplicate (distinct Reporter ID + AS)
     - Build NLRI with both Reporter TLVs

   R1 Advertisement to downstream:
     AFI: 25, SAFI: 70
     AS_PATH: 65100 (from best path)
     NLRI (192.0.2.0/24 Unreachability):
       Reporter TLV #1:
         Reporter ID: 198.51.100.1, AS: 65001
         Reason: 3 (RPKI Invalid)
         Timestamp: 1733789400
       Reporter TLV #2:
         Reporter ID: 198.51.100.2, AS: 65002
         Reason: 1 (Policy Blocked)
         Timestamp: 1733789410
```

Example 6: Withdrawal Procedures

Continuing from Example 5, Reporter 198.51.100.1 clears its report first (partial withdrawal), then Reporter 198.51.100.2 also clears (complete withdrawal).

```
   Initial State on R1:
     UI-RIB Entry for NLRI key
     (RD 198.51.100.1:100, ETag 0, IPv4, 192.0.2.0/24):
       Reporter TLV #1: 198.51.100.1 / AS 65001 (from N1)
       Reporter TLV #2: 198.51.100.2 / AS 65002 (from N2)

   Event: N1 sends MP_UNREACH_NLRI for the NLRI key.

   R1 Processing:
     1. Identify withdrawal source: N1
     2. Remove Reporter TLVs sourced from N1
        (Reporter 198.51.100.1 / AS 65001)
     3. Reporter TLV #2 remains
     4. Re-advertise with the remaining Reporter TLV

   R1 Advertisement to downstream:
     MP_REACH_NLRI (AFI 25, SAFI 70):
       NLRI (192.0.2.0/24 Unreachability):
         Reporter TLV:
           Reporter ID: 198.51.100.2, AS: 65002
           Reason: 1 (Policy Blocked)
           Timestamp: 1733789410

   Later Event: N2 also sends MP_UNREACH_NLRI for the NLRI key.

   R1 Processing:
     1. Remove Reporter TLVs sourced from N2
     2. No Reporter TLVs remain
     3. Withdraw the entire NLRI

   R1 Advertisement to downstream:
     MP_UNREACH_NLRI (AFI 25, SAFI 70):
       Withdrawn NLRI:
         Route Type: TBD1
         Length: 32
         Route Distinguisher: 198.51.100.1:100
         Ethernet Segment Identifier: 0
         Ethernet Tag ID: 0
         Address Family: 1 (IPv4)
         IP Prefix Length: 24
         IP Prefix: 192.0.2.0
         GW IP Address Length: 0
         MPLS Label: 0
```

Appendix B: Design Tradeoffs

This appendix summarizes encoding choices and their rationale.

Explicit Address Family vs inferring IPv4/IPv6: Reachability RT-5 uses a fixed route-type-specific size (34 or 58 octets), so receivers can infer prefix width. Unreachability NLRIs append variable-length Reporter TLVs, so total length no longer implies address family, and IP Prefix Length alone is ambiguous (e.g., /24 is valid for both). A 1-octet Address Family field resolves this.

One route type vs two (per address family): Separate route type values would remove the Address Family octet but duplicate specifications and IANA registrations. This document uses one type with an explicit family indicator.

RT-5-shaped NLRI vs minimal custom encoding: Reusing the RT-5 field order maximizes familiarity and parser reuse, at the cost of carrying unused fields (ESI, GW IP, MPLS Label) plus one new octet for Address Family. Reporter detail is concentrated in TLVs.

Route-type-specific Length limit: The Length field is one octet ([RFC7432]), so the route-type-specific portion cannot exceed 255 octets. Implementations MUST stay within that limit by bounding the number of Reporter TLVs per NLRI.

Authors' Addresses

Jeff Tantsura
Nvidia
Email: jefftant.ietf@gmail.com

Vivek Venkatraman
Nvidia
Email: vivek@nvidia.com

Karthikeya Venkat Muppalla
Nvidia
Email: kmuppalla@nvidia.com

Pooja Jagadeesh Doijode
Nvidia
Email: pdoijode@nvidia.com

Maciej Rzehak
CoreWeave
Email: mrzehak@coreweave.com