

IDR
Internet-Draft
Updates: 4684 (if approved)
Intended status: Standards Track
Expires: July 18, 2019

J. Heitz
A. Sajassi
Cisco
I. Bagdonas
Equinix
January 14, 2019

BGP Extra Extended Community
draft-heitz-idr-extra-extended-community-02

Abstract

Auto-derived identifiers are used by applications to enable zero configuration features. Such identifiers are often a combination of primitive identifiers and are thus longer. In addition, existing identifiers have grown longer. IP addresses have grown from 4 octets to 16. AS numbers have grown from 2 octets to 4. In order to accommodate such longer identifiers in BGP extended communities, this document defines a new BGP path attribute, the Extra Extended Communities attribute. It is similar to the Extended Community, but is 24 octets long. Communities are mostly used within ASes under a single administration or between neighboring ASes. Limiting the spread of communities beyond their intended reach and polluting the internet at large is complex and error prone. To simplify this, enhanced transitivity options are provided.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <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 July 18, 2019.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

1. Introduction	3
2. BGP Extra Extended Community Attribute	3
3. Transitivity	4
4. Capability	5
5. Constrained Route Distribution	5
6. IPv6-Address-Specific Extra Extended Community type	7
7. IPv4-Address-Specific Extra Extended Community type	7
8. AS-Specific Extra Extended Community type	8
9. Route Target Extra Extended Community Sub-Type	9
10. EVPN Extra Extended Community type	10
11. EVPN Route Target Extra Extended Community sub-types	10
12. EVPN ES-Import Route Target Extra Extended Community sub-type	11
13. EVPN ESI-EVI Route Target Extra Extended Community sub-type	11
14. EVPN Overlay Route Target Extra Extended Community sub-type	12
15. Duplicate XXC	14
16. Error Handling	14
17. Security Considerations	15
18. IANA Considerations	15
18.1. Registry: BGP Extra Extended Community Types	15
18.2. Registry: IPv6-Address-Specific Extra Extended Community Sub-Types	15
18.3. Registry: IPv4-Address-Specific Extra Extended Community Sub-Types	16
18.4. Registry: AS-Specific Extra Extended Community Sub-Types	16
18.5. Registry: EVPN Extra Extended Community Sub-Types	16
19. References	17
19.1. Normative References	17
19.2. Informative References	18
Authors' Addresses	18

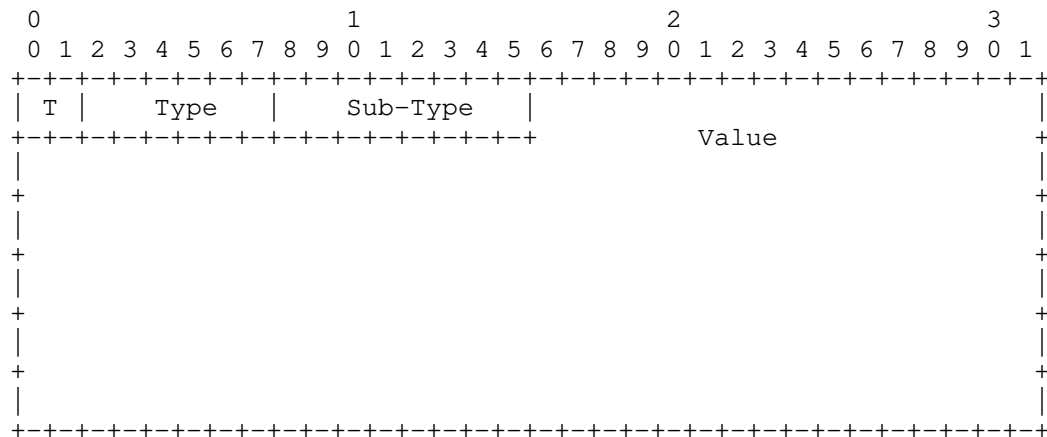
1. Introduction

A BGP Extended Community attribute is defined that encodes 24 octet communities. It is similar to the Extended Communities attribute defined in [RFC4360], but larger. To simplify the IANA registries, the transitivity of a Extra Extended Community is not part of the IANA registered type. Any type can be encoded with any transitivity. BGP autonomous system (AS) relationships have become more complex. Several contiguous ASes may be under a common administration. A transitivity is defined that allows a XXC to be sent among these ASes only. Some XXCs may be required to be transferred only between neighboring ASes, even though they are under a different administration. A transitivity type to allow this is defined. Up to now, the range of ASes among which a community is distributed is enforced by routing policies. These policies are sometimes executed in the receiving AS, not under the control of the sending AS. The enhanced transitivity options offered in this document will simplify policies that are used to distribute communities.

2. BGP Extra Extended Community Attribute

The Extra Extended Communities Attribute is a transitive optional BGP attribute, with the Type Code [to be assigned by IANA]. The attribute consists of a set of "Extra Extended Communities" (XXC). The attribute SHOULD contain at least one XXC.

Each XXC is encoded as a 24-octet quantity, as follows:



The fields are as shown below:

T - Transitivity field (2 bits). This is further described below.

- Type - 6 bits. IANA will maintain a registry of types. Four types are described in this document.
- Sub-type - 8 bits. IANA will maintain a registry of sub-types for each registered type.
- Value - The actual information according to the type and sub-type.

Two XXCs are considered equal when all the 24 octets of the XXCs are equal, except the Transitivity field..

3. Transitivity

The transitivity field determines how BGP speakers transfer the XXC across real Autonomous System (AS) boundaries. The XXC is always transitive between Member-ASes in an AS confederation [RFC5065]. The values are:

- 0 - Transitive: The XXC is transitive across ASes.
- 1 - Non-transitive: The XXC is not transitive across ASes.
- 2 - Administration Transitive: The XXC is transitive across ASes under the same administration only.
- 3 - One-time Transitive: The XXC is transitive across ASes under the same administration and into an AS under the neighboring administration, but not into an AS under a further administration.

To be not transitive means that a sending speaker MUST not send the community. A speaker that receives an XXC across an AS boundary not allowed by the transitivity of that XXC MUST treat the containing UPDATE message as malformed.

A single administration may own a multitude of contiguous ASes. XXCs with transitivity types 2 and 3 are transitive between these ASes. By default, an EBGp neighbor is assumed to be under a different administration. If an EBGp neighbor session is to a speaker in the same administration, then it needs to be configured as such. No Administration identifiers are required. The configuration just needs to specify "Same Administration". There is no new field in the BGP OPEN message to indicate administration membership.

A BGP speaker that receives a XXC with transitivity 3 from a neighbor in an AS under a different administration SHOULD change the transitivity field of the XXC to 2.

The transitivity of an XXC is intended to limit the travel of the XXC in default conditions. It prevents an XXC from traveling far beyond its intended reach if nothing else is done. That means a speaker is free to change the transitivity of an XXC according to local policy to support specific use cases. As an example, a route server as defined in [RFC7947] MAY choose not to change transitivity from 3 to 2. The definition of an XXC type MAY include a specification of the default transitivity for the type.

The Transitivity field is not implicitly associated with the Type and Sub-Type fields the way they are in Extended Communities. The Transitivity field should be set by the originator based upon individual circumstances at the originator. The transitivity is not assigned by IANA.

4. Capability

BGP speakers that do not implement Extra Extended Communities will transfer XXCs even though they may not be transitive across their AS boundaries. To prevent this, a BGP capability as defined in [RFC5492] is required. The length of the capability is 0. A BGP speaker SHOULD NOT send a XXC, the transitivity of which is not 0, to a speaker from which it has not received the Extra Extended Community Capability in its OPEN message. A BGP speaker SHOULD withdraw a route from a neighbor if that neighbor does not advertise this capability and the route contains an XXC. These rules are intended to prevent XXCs from leaking outside their intended range. There may be cases where such leaking causes no ill effects and the rules can be safely ignored. Such cases are beyond the scope of this document.

A transitive XXC may be sent to a neighbor that has not sent the capability.

5. Constrained Route Distribution

[RFC4684] defines Constrained Route Distribution. That document is updated as follows:

The Extra Constrained Route Distribution SAFI is defined, the NLRI of which is as follows:

AFI	(2 octets)
SAFI	(1 octets)
origin AS	(4 octets)
XXC value	(24 octets)

The fields are as shown below:

- AFI - The filter specified in this NLRI applies only to prefixes with this AFI.
- SAFI - The filter specified in this NLRI applies only to prefixes with this SAFI.
- origin AS - Routes that do not have this origin AS will be blocked by the filter.
- XXC value - Routes that do not have this XXC will be blocked by the filter.

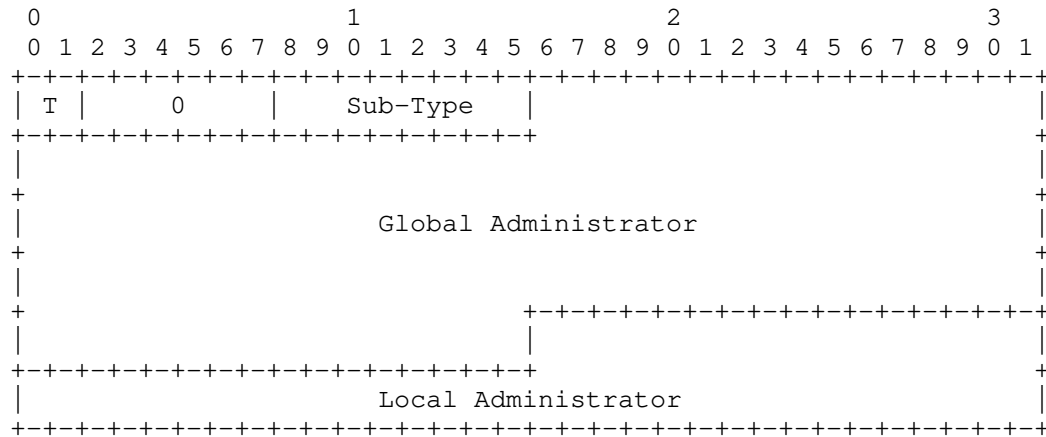
This works the same way as [RFC4684] with the following differences:

- The maximum prefix size of this NLRI is 248 bits.
- The minimum prefix size of this NLRI is 56 bits except for the default route, which is 0 bits long.
- The filter applies only to prefixes that have the specified AFI/SAFI.
- The filter applies to any XXC values. The filtered XXC does not need to be designated a "Route Target".

Route targets can be expressed as prefixes, where, for instance, a prefix would encompass all route target extended communities assigned by a given Global Administrator. Route Target prefixes can be aggregated. However if done so, then AFI, SAFI, Transitivity, Route Target Type, Sub-Type and the Global Administrator Route Target fields MUST NOT be aggregated.

The address family "Extra Constrained Route Distribution" cannot filter itself.

6. IPv6-Address-Specific Extra Extended Community type



The Type field is 0. The Sub-Type is to be assigned by IANA as detailed in the IANA Considerations section.

The Value field consists of 2 sub-fields:

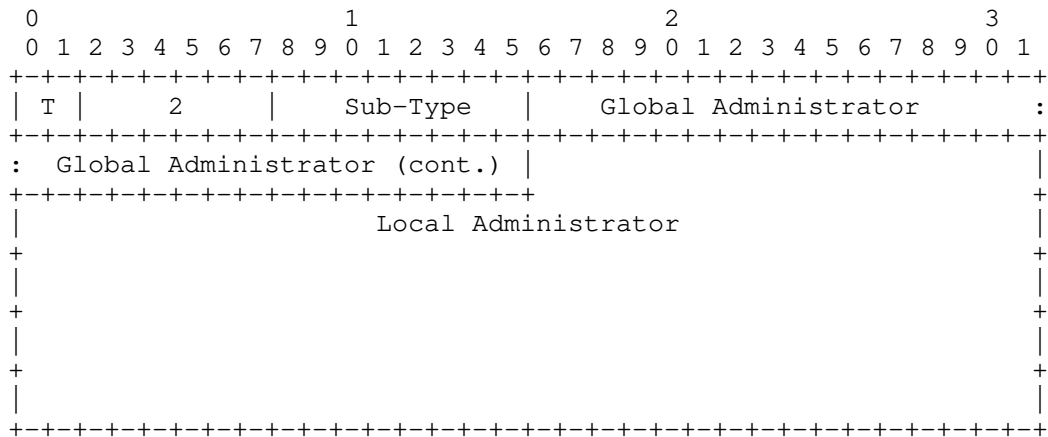
Global Administrator sub-field: 16 octets

This sub-field contains an IPv6 unicast address assigned by one of the Internet registries to the administration of the service using the XXC.

Local Administrator sub-field: 6 octets

The organization identified by the IP address in the Global Administrator sub-field can encode any information in this sub-field. The format and meaning of the value encoded in this sub-field should be defined by the sub-type of the XXC.

7. IPv4-Address-Specific Extra Extended Community type



The Type field is 2. The Sub-Type is to be assigned by IANA for individual functions.

The Value field consists of 2 sub-fields:

Global Administrator sub-field: 4 octets

This sub-field contains a 4-octet Autonomous System number assigned by IANA. Note that an ASN that is less than 65536 in value is represented in 4 octets by setting the higher two octets to 0.

Local Administrator sub-field: 18 octets

The organization identified by the Autonomous System number in the Global Administrator sub-field can encode any information in this sub-field. The format and meaning of the value encoded in this sub-field should be defined by the sub-type of the XXC.

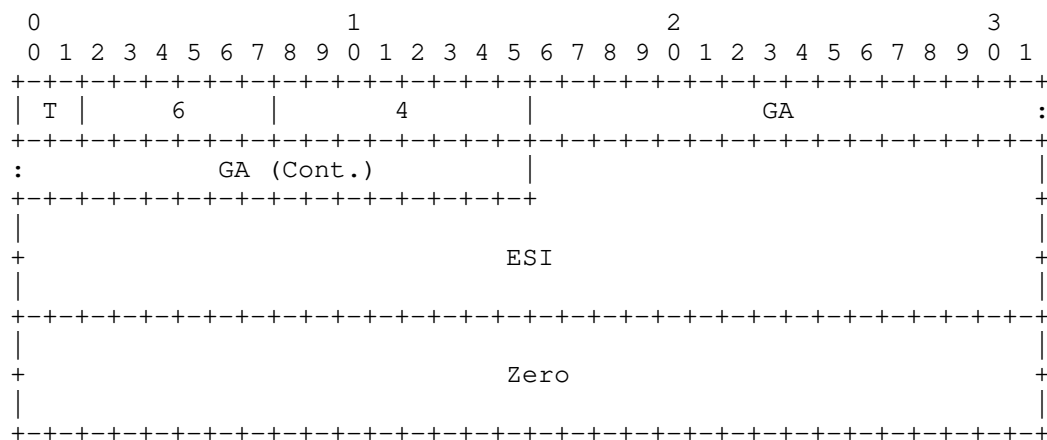
9. Route Target Extra Extended Community Sub-Type

For each of the Types of XXC: IPv4-Address-Specific, IPv6-Address-Specific and AS-Specific, the Sub-Type 2 denotes a Route-Target. It has the same use as the Route Target defined in [RFC4360] Sec. 4. The XXC route targets are independent of the RFC4360 Route Targets. There is no way to convert between the two. Either or both may be used in any deployment.

The Ethernet Tag ID is placed into the least significant octets of the Value field. The remaining octets are 0.

12. EVPN ES-Import Route Target Extra Extended Community sub-type

The ES-Import Route Target as specified in [RFC7432] Sec. 7.6 limits the ESI to 6 octets. Thus it cannot be automatically derived for all ESI types. This ES-Import RT-XXC allows the use of the full 10 octets of ESI. The ES-Import Route Target XXC and the ES-Import Route Target extended community are independent. Either or both may be used in any deployment. There is no conversion from one to the other.

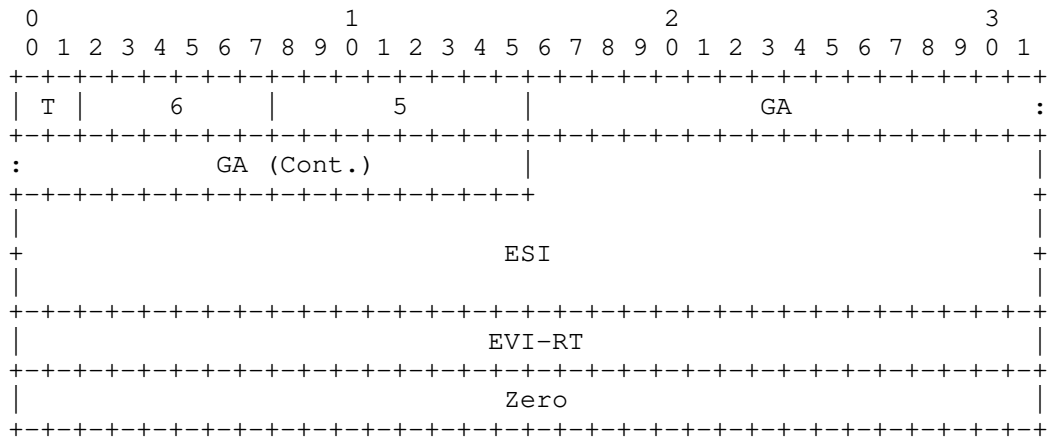


The Type field is 6. The Sub-Type is 4. The fields are as follows:

- GA - 4 octets. Global Administrator. This is the Autonomous System Number of the AS where this RT is assigned.
- ESI - 10 octets. Ethernet Segment Identifier.
- Zero - 8 octets filled with 0. Must be ignored by the receiver.

13. EVPN ESI-EVI Route Target Extra Extended Community sub-type

The ESI-EVI Route Target is used in EVPN route types 7 and 8 to filter routes by both ESI and Ethernet Tag ID. More details are in [I-D.ietf-bess-evpn-igmp-mld-proxy].

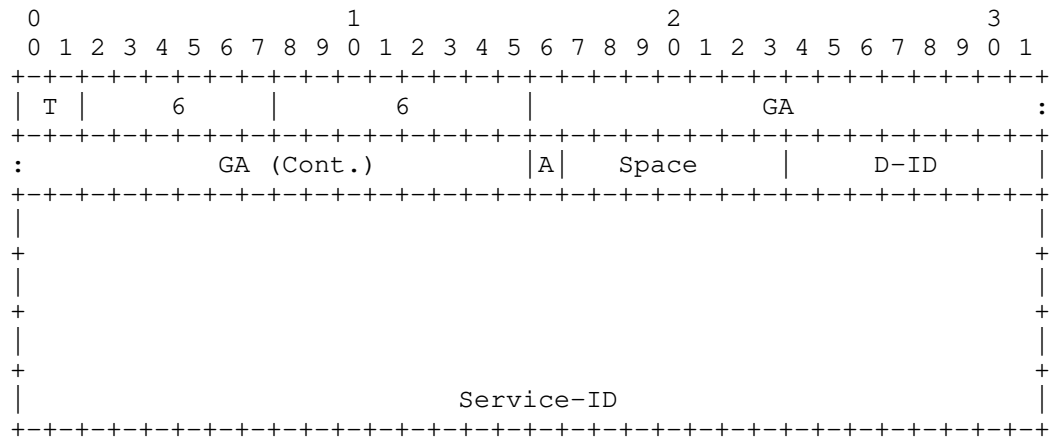


The Type field is 6. The Sub-Type is 5. The fields are as follows:

- GA - 4 octets. Global Administrator. This is the Autonomous System Number of the AS where this RT is assigned.
- ESI - 10 octets. Ethernet Segment Identifier.
- EVI-RT - 4 octets. The least significant 4 octets of the Local Administrator field of the route target associated with the EVI.
- Zero - 4 octets filled with 0. Must be ignored by the receiver.

14. EVPN Overlay Route Target Extra Extended Community sub-type

This EVPN Overlay Route Target Extra Extended Community type is used to filter routes based upon the identifier used in the specified overlay protocol. It works the same way as the RT described in [RFC8365] Sec. 5.1.2.1. It simply provides more room in the fields to allow auto-derivation for more cases. First, it allows a 4 octet AS number. Second, the Service ID is large enough to fit any of the selected IDs.



The Type field is 6. The Sub-Type is 6. The fields are as follows:

- GA - 4 octets. Global Administrator. This is the Autonomous System Number of the AS where this RT is assigned.
- A - A single bit indicating if this RT is auto-derived
- 0 : auto-derived
- 1 : manually derived
- Space - 7 bits. The identifier space appropriate to the service. The following spaces are defined:
- 0 : VID (802.1Q VLAN ID)
- 1 : VXLAN
- 2 : NVGRE
- 3 : I-SID
- 4 : EVI
- 5 : dual-VID (QinQ VLAN ID)
- D-ID - 1 octet. The default value of domain-id is zero indicating that only a single numbering space exists for a given technology. However, if there is more than one number space for a given technology (e.g., overlapping VXLAN spaces), then each of the number

spaces need to be identified by their corresponding domain-id starting from 1.

Service-ID - 16 octets. VNI, VSID, I-SID, VID or other identifier as appropriate for the service specified in the Space field. If the contained identifier is less than 16 octets long, then it is placed in the least significant octets of the Service-ID field with the higher octets being filled with 0.

15. Duplicate XXC

Two XXCs are considered duplicate if the values of each field except the Transitivity field match.

A BGP speaker SHOULD NOT send a duplicate XXC. However, this is not an error, but merely suboptimal. The duplication of a XXC has no meaning. A receiver of a duplicate XXC SHOULD silently discard the duplicate. The duplication of a XXC cannot be used to compound its effect. For example if one XXC causes the local preference to be incremented by 5, the presence of two of the same XXC will not increment the LP by 10. OTOH, if one XXC increments the LP by 5 and a different XXC increments it by 10, then the combination will cause an increment of 15.

A BGP speaker that receives duplicate XXCs that differ in the Transitivity MAY silently discard the XXC with the more restrictive transitivity.

16. Error Handling

A BGP Extra Extended Communities attribute SHALL be considered malformed if the length of the BGP Extra Extended Communities Attribute value, expressed in octets, is not a multiple of 24. The error SHALL be handled using the approach of "treat-as-withdraw" as described in Section 2 of [RFC7606]. Receipt of a zero length Extra Extended Communities attribute SHALL be treated as "attribute-discard".

The order in which the XXCs appear in the XXC attribute is not significant. It is not an error for a BGP speaker to propagate a set of XXCs in a different order than in which they were received.

If a field in a specific type of XXC is invalid in another setting, it is not necessarily to be considered invalid in a XXC. For example, 0 is an invalid AS number when used in an AS path attribute. That does not make it invalid as an ASN in the AS-Specific XXC. The

behavior and validity of fields in XXCs are to be defined by a specification of the specific type and sub-type of the XXC.

The receipt of an XXC that violates the transitivity rules SHALL be handled using the approach of "treat-as-withdraw".

17. Security Considerations

TBD

18. IANA Considerations

IANA is requested to assign a SAFI for the Extra Constrained Route Distribution address family.

IANA is requested to assign a BGP path attribute value for the Extra Extended Community attribute.

IANA is requested to create and maintain registries as detailed in the following sections. For each registry, the allocation policies as per [RFC8126] are stated for the ranges of values and some values are allocated by this document.

18.1. Registry: BGP Extra Extended Community Types

Range	Allocation Policy	
0-31	First Come First Served	
32-47	Experimental	
48-63	Standards Action	
Value	Description	Reference
0	IPv6-Address-Specific	This RFC
1	IPv4-Address-Specific	This RFC
2	AS-Specific	This RFC
6	EVPN	This RFC

18.2. Registry: IPv6-Address-Specific Extra Extended Community Sub-Types

Range	Allocation Policy	
-----	-----	
0-191	First Come First Served	
192-255	IETF Review	
Value	Description	Reference
-----	-----	-----
2	Route Target	RFC4360

18.3. Registry: IPv4-Address-Specific Extra Extended Community Sub-Types

Range	Allocation Policy	
-----	-----	
0-191	First Come First Served	
192-255	IETF Review	
Value	Description	Reference
-----	-----	-----
2	Route Target	RFC4360

18.4. Registry: AS-Specific Extra Extended Community Sub-Types

Range	Allocation Policy	
-----	-----	
0-191	First Come First Served	
192-255	IETF Review	
Value	Description	Reference
-----	-----	-----
2	Route Target	RFC4360

18.5. Registry: EVPN Extra Extended Community Sub-Types

Range	Allocation Policy	
-----	-----	
0-191	First Come First Served	
192-255	IETF Review	
Value	Description	Reference
-----	-----	-----
1	EVPN AS-Specific Route Target	This RFC
2	EVPN IPv4-Specific Route Target	This RFC
3	EVPN IPv6-Specific Route Target	This RFC
4	EVPN ES-Import Route Target	This RFC
5	EVPN ESI-EVI Route Target	This RFC
6	EVPN Overlay Route Target	This RFC

19. References

19.1. 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/info/rfc2119>>.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <<https://www.rfc-editor.org/info/rfc4360>>.
- [RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <<https://www.rfc-editor.org/info/rfc4684>>.
- [RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <<https://www.rfc-editor.org/info/rfc5065>>.
- [RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <<https://www.rfc-editor.org/info/rfc5492>>.
- [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>>.
- [RFC7947] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker, "Internet Exchange BGP Route Server", RFC 7947, DOI 10.17487/RFC7947, September 2016, <<https://www.rfc-editor.org/info/rfc7947>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.

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

19.2. Informative References

[I-D.ietf-bess-evpn-igmp-mld-proxy] Sajassi, A., Thoria, S., Patel, K., Yeung, D., Drake, J., and W. Lin, "IGMP and MLD Proxy for EVPN", draft-ietf-bess-evpn-igmp-mld-proxy-00 (work in progress), March 2017.

Authors' Addresses

Jakob Heitz
Cisco
170 West Tasman Drive
San Jose, CA, CA 95134
USA

Email: jheitz@cisco.com

Ali Sajassi
Cisco
170 West Tasman Drive
San Jose, CA, CA 95134
USA

Email: sajassi@cisco.com

Ignas Bagdonas
Equinix
London
UK

Email: ibagdona.ietf@gmail.com