Workgroup: Internet Engineering Task Force
Internet-Draft:
draft-scudder-idr-entropy-label-01
Published: 18 August 2022
Intended Status: Standards Track
Expires: 19 February 2023
Authors: J. G. Scudder, Ed. K. Kompella
Juniper Networks Juniper Networks
S. Mohanty J. Uttaro B. Wen
Cisco Systems AT&T Comcast

BGP Entropy Label Capability, Version 3

Abstract

This specification defines the Entropy Label Capability Attribute version 3 (ELCv3), a BGP attribute that can be used to inform an LSP ingress router about an LSP egress router's ability to process entropy labels. This version of the attribute corrects a specification error in the first version, and an improper code point reuse in the second.

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 19 February 2023.

Table of Contents

1. Introduction

[RFC6790] defines the Entropy Label Capability attribute (ELC), an optional, transitive BGP path attribute. For correct operation, it is necessary that any intermediate node modifying the next hop of a route must remove the ELC unless the node so doing is able to process entropy labels. Sadly, these requirements cannot be fulfilled with the ELC as specified, because it is an optional, transitive attribute: by definition, a node that does not support the ELC will propagate the attribute. But such a node might be exactly the one that we desire to remove it. For this reason, [RFC7447] deprecated the attribute.

Roughly concurrently with the development and advancement of RFC 7447, Juniper Networks began shipping routing code that implements what is documented in [I-D.scudder-bgp-entropy-label] and dubbed Entropy Label Capability version 2 (ELCv2). That implementation uses the code point that was assigned by RFC 6790 and deprecated by RFC 7447. At time of writing, the functionality is in use in operational networks.

The present specification is based on ELCv2 but moves to a new, previously unallocated, code point.

A related solution to the problem of signaling entropy label capability is [I-D.ietf-idr-next-hop-capability]. That specification is based on an optional, non-transitive path attribute. In contrast,

ELCv3 (and ELCv2) is based on an optional, transitive path attribute. This expands the deployment options available -- in many cases (for example, route reflectors) it's fine that an intermediate node does propagate an ELCv3 even if it doesn't itself have the ability to process entropy labels.

In order to prevent use of the signaled information beyond the intended perimeter (the problem that led to the deprecation of ELC, and which is inherently solved by [I-D.ietf-idr-next-hop-capability]'s use of a non-transitive attribute), in this specification we take the approach of carrying a copy of the next hop information in the ELCv3. This allows the node processing it to know if it can rely on the information carried therein, while still allowing it to be propagated by all intermediate nodes.

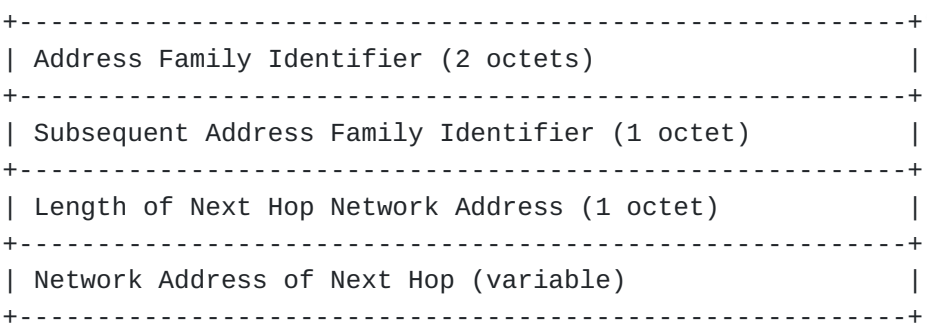
1.1. Requirements Language

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.

2. Entropy Label Capability Path Attribute, Version 3

The Entropy Label Capability Path Attribute, Version 3 (ELCv3) is an optional, transitive BGP attribute with type code TBD1. The ELCv3 has as its data a network layer address, representing the next hop of the route the ELCv3 accompanies. The ELCv3 signals a useful optimization, so it is desirable to make it transitive; the next hop data is to ensure correctness if it traverses BGP speakers that do not understand the ELCv3.

The Attribute Data field of the ELCv3 path attribute is encoded as shown below:



The meanings of the fields are as given in Section 3 of [RFC4760].

When BGP [RFC4271] is used for distributing labeled Network Layer Reachability Information (NLRI) as described in, for example,

[RFC8277], the route may include the ELCv3 as part of the Path Attributes. The inclusion of this attribute with a route indicates that the egress of the associated Label Switched Path (LSP) can process entropy labels as an egress Label Switched Router (LSR) for that route -- see Section 4.2 of [RFC6790]. Below, we refer to this for brevity as being "EL-capable."

2.1. Sending the ELCv3

When a BGP speaker S has a route R it wishes to advertise with next hop N to its peer, it MUST NOT include the ELCv3 attribute except if it knows that the egress of the associated LSP L is EL-capable. Specifically, this will be true if S:

- Is itself the egress, and knows itself to be EL-capable, or
- Is re-advertising a BGP route it received with a valid ELCv3 attribute, and is not changing the value of N, or
- Is re-advertising a BGP route it received with a valid ELCv3 attribute, and is changing the value of N, and knows (for example, through configuration) that the router represented by N is either the LSP egress and is EL-capable, or that it will process the outer label(s) without processing the entropy label below, as with a transit LSR, or
- Is redistributing a route learned from another protocol, and that other protocol conveyed the knowledge that the egress of L was EL-capable (for example, this might be known through the LDP ELC TLV, Section 5.1 of [RFC6790]).

In any event, when sending an ELCv3, S MUST set the data portion of the ELCv3 to be equal to N, using the encoding given in Section 2.

The ELCv3 MAY be advertised with routes that are labeled, such as those using SAFI 4 [RFC8277]. It MUST NOT be advertised with unlabeled routes.

We note that due to the nature of BGP optional transitive path attributes, any BGP speaker that does not implement this specification will propagate the ELCv3, the requirements of this section notwithstanding. However, such a speaker will not update the data part of the ELCv3.

2.2. Receiving the ELCv3

When a BGP speaker receives an unlabeled route that includes the ELCv3, it MUST discard the ELCv3.

When a BGP speaker receives a labeled route that includes the ELCv3, it MUST compare the address given in the ELCv3's data portion to the next hop of the route. If the two are equal, the egress of the LSP supports entropy labels, which implies that the receiving BGP speaker, if acting as ingress, MAY insert an entropy label below the advertised label, as per Section 4.2 of [RFC6790]. If the two are not equal, either some intermediate router that does not implement this specification modified the next hop, or some router on the path had an incorrect implementation. In either case, the action taken is the same: the ELCv3 MUST be discarded. The Partial bit MAY be inspected -- if it is equal to zero, then the mismatch must have been caused by an incorrect implementation, and the error MAY be logged.

When a BGP speaker receives a route that includes an ELCv3 whose Attribute Length is less than 4, whose Attribute Length is not greater than or equal to 4 plus the value encoded in the Length of Next Hop Network Address carried in the Attribute Data, or whose Attribute Data is otherwise inconsistent with the encoding specified in Section 2, it MUST discard the ELCv3.

If an ELCv3 includes data beyond the Network Address of Next Hop field, such data MUST be disregarded. If the ELCv3 is propagated, the unknown data MUST be included with it.

3. IANA Considerations

IANA is requested to make a new allocation in the BGP Path Attributes registry:

*Value = TBD1

*Code = ELCv3

*Reference = (this document)

4. Security Considerations

Insertion of an ELCv3 by an attacker could cause forwarding to fail. Deletion of an ELCv3 by an attacker could cause one path in the network to be overutilized and another to be underutilized. However, we note that an attacker able to accomplish either of these (below, an "on-path attacker") could equally insert or remove any other BGP path attribute or message. The former attack described above denies service for a given route, which can be accomplished by an on-path attacker in any number of ways even absent ELCv3. The latter attack defeats an optimization but nothing more; it seems dubious that an attacker would go to the trouble of doing so rather than launching some more damaging attack.

The Attribute Data portion of the ELCv3 contains the next hop the attribute's originator included when sending it. This will typically be an IP address of the router in question. This may be an infrastructure address the network operator does not intend to announce beyond the border of its Autonomous System, and it may even be considered in some weak sense, confidential information. Although the desired operation of the protocol is for the attribute's propagation scope to be limited to the network operator's own Autonomous System, it will not always be so -- indeed, that is the reason this specification had to be written. So, sometimes this information could leak beyond its intended scope. (Note that it will only propagate as far as the first router that does support this specification, at which point it will be discarded per Section 2.2.)

5. References

5.1. 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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[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>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[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>.

5.2. Informative References

[I-D.ietf-idr-next-hop-capability] Decraene, B., Kompella, K., and W. Henderickx, "BGP Next-Hop dependent capabilities", Work in Progress, Internet-Draft, draft-ietf-idr-next-hop-capability-08, 8 June 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-next-hop-capability-08.txt>.

[I-D.scudder-bgp-entropy-label]
Scudder, J. G. and K. Kompella, "BGP Entropy Label Capability, Version 2", Work in Progress, Internet-Draft, draft-scudder-bgp-entropy-label-00, 28 April 2022, <https://www.ietf.org/archive/id/draft-scudder-bgp-entropy-label-00.txt>.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <https://www.rfc-editor.org/info/rfc4272>.

[RFC7447] Scudder, J. and K. Kompella, "Deprecation of BGP Entropy Label Capability Attribute", RFC 7447, DOI 10.17487/RFC7447, February 2015, <https://www.rfc-editor.org/info/rfc7447>.

[RFC8277] Rosen, E., "Using BGP to Bind MPLS Labels to Address Prefixes", RFC 8277, DOI 10.17487/RFC8277, October 2017, <https://www.rfc-editor.org/info/rfc8277>.

Appendix A. Other Means of Signaling EL Capability

A router that supports this specification could also have other means to know that an egress is EL-capable, for example it could support ELCv2 [I-D.scudder-bgp-entropy-label], or it could know through configuration. If a router learns through any means that an egress is EL-capable, it MAY treat the egress as EL-capable. For example, reception of a valid ELCv2 would be sufficient (even if a valid ELCv3 is not received), and similarly reception of a valid ELCv3 would be sufficient (even if a valid ELCv2 is not received). The details of which methods are accepted for signaling EL capability are beyond the scope of this specification, but SHOULD be configurable by the user.

A.1. Backward Compatibility with ELCv2

As was noted in Section 1, there are networks in which ELCv2 (documented in [I-D.scudder-bgp-entropy-label]) is already in use.

Any node that sends the ELCv2 format may also include an ELCv3 per Section 2.1, so that both formats are sent. The exact set of formats to send SHOULD be user-configurable.

As discussed above, a route received with either a valid ELCv2 or ELCv3 may be considered EL-capable.

Appendix B. Contributors

Serge Krier
Cisco Systems

Email: sekrier@cisco.com

Appendix C. Acknowledgements

Thanks to Swadesh Agrawal, Alia Atlas, Bruno Decraene, Martin Djernaes, John Drake, Adrian Farrell, Keyur Patel, Toby Rees, and Ravi Singh, for their discussion of this issue. Particular thanks to Kevin Wang for his many valuable contributions.

Authors' Addresses

John G. Scudder (editor)
Juniper Networks

Email: jgs@juniper.net

Kireeti Kompella
Juniper Networks

Email: kireeti@juniper.net

Satya Mohanty
Cisco Systems

Email: satyamoh@cisco.com

James Uttaro
AT&T

Email: ju1738@att.com

Bin Wen
Comcast

Email: Bin_Wen@comcast.com