Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 29 November 2023

R. Bush
Arrcus & IIJ
K. Patel
Arrcus

28 May 2023

L3ND Upper-Layer Protocol Configuration
draft-ymbk-idr-l3nd-ulpc-07

Abstract

This document adds PDUs to the Layer-3 Neighbor Discovery protocol to communicate the parameters needed to exchange inter-device Upper Layer Protocol Configuration for upper-layer protocols such as the BGP family.

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.

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 29 November 2023.

Copyright Notice

Copyright (c) 2023 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction

Massive Data Centers (MDCs) which use upper-layer protocols such as BGP4 and other routing protocols may use the Layer-3 Neighbor Discovery Protocol, L3ND, [I-D.ymbk-idr-l3nd] to reveal the inter-device links of the topology. It is desirable for devices to facilitate the configuration parameters of those upper layer protocols to enable more hands-free configuration. This document defines a new L3ND PDU to communicate these Upper-Layer Protocol Configuration parameters.

2. Reading and Terminology

The reader is assumed to have read Layer-3 Neighbor Discovery [I-D.ymbk-idr-l3nd]. The terminology and PDUs there are assumed here.

Familiarity with the BGP4 Protocol [RFC4271] is assumed.

3. Upper-Layer Protocol Configuration PDU

To communicate parameters required to configure peering and operation of Upper-Layer Protocols at IP layer-3 and above, e.g., BGP sessions on a link, a neutral sub-TLV based Upper-Layer Protocol PDU is defined as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Version = 0  |    Type = 8   |         Payload Length        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               |   ULPC Type   |   AttrCount   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Serial Number                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Attribute List ...                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Version, Type, and Payload Length as defined in [I-D.ymbk-idr-l3nd] apply to this PDU.

The BGP Authentication sub-TLV provides for provisioning MD5, which is a quite weak hash, horribly out of fashion, and kills puppies. But, like it or not, it has been sufficient against the kinds of attacks BGP TCP sessions have endured. So it is what BGP deployments use.

As the ULPC PDU may contain keying material, e.g. [RFC2385], it SHOULD BE over TLS.

ULPC Type: A one byte integer denoting the type of the upper-layer protocol

0 :	Reserved
1 :	BGP
2-255 :	Reserved

The one octet AttrCount is the number of attribute sub-TLVs in the Attribute List.

The Attribute List is a, possibly null, set of sub-TLVs describing the configuration attributes of the specific upper-layer protocol.

An Attribute consists of a one octet Attribute Type, a one octet Attribute Length of the number of octets in the Attribute, and a Payload of arbitrary length up to 253 octets.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Attr Type = 1 |    Attr Len   |            Payload            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.1. ULPC BGP Attribute sub-TLVs

The parameters needed for BGP peering on a link are exchanged in sub-TLVs within an Upper-Layer Protocol PDU. The following describe the various sub-TLVs for BGP.

The goal is to provide the minimal set of configuration parameters needed by BGP OPEN to successfully start a BGP peering. The goal is specifically not to replace or conflict with data exchanged during BGP OPEN. Multiple sources of truth are a recipe for complexity and hence pain.

If there are multiple BGP sessions on a link, e.g., IPv4 and IPv6, then separate BGP ULPC PDUs should be sent, one for each address family.

A peer receiving BGP ULPC PDUs has only one active BGP ULPC PDU for an particular address family on a specific link at any point in time; receipt of a new BGP ULPC PDU for a particular address family replaces the data any previous one; but does not actually affect the session.

If there are one or more open BGP sessions, receipt of a new BGP ULPC PDU SHOULD not affect these sessions. The received data are stored for a future session restart.

As a link may have multiple encapsulations and multiple addresses for an IP encapsulation, which address of which encapsulation is to be used for the BGP session MUST be specified.

For each BGP peering on a link here MUST be one agreed encapsulation, and the addresses used MUST be in the corresponding L3ND IPv4/IPv6 Encapsulation PDUs. If the choice is ambiguous, an Attribute may be used to signal preferences.

If a peering address has been announced as a loopback, i.e. MUST BE flagged as such in the L3ND Encapsulation PDU (see [I-D.ymbk-idr-l3nd] Sec. 10.2), a two or three hop BGP session MUST be established as needed. Otherwise a direct one hop session is used. The BGP session to a loopback will forward to the peer’s address which was marked as Primary in the L3ND Encapsulation Flags, iff it is in a subnet which is shared with both BGP speakers. If the primary is not in a common subnet, then the BGP speaker MAY pick a forwarding next hop that is in a subnet they share. If there are multiple choices, the BGP speaker SHOULD have signaled which subnet to choose in an Upper-Layer Protocol Configuration PDU Attribute.

Attributes MUST be unique in the Attribute List. I.e. a particular Attr Type MUST NOT occur more than once in the Attribute List. If a ULPC PDU is received with more than one occurrence of a particular Attr Type, an Error ACK MUST be returned.

As there are separate PDU Attr Types for IPv4 and IPv6 peering addresses, separate sessions for the two AFIs MAY be created for the same ASN in one ULPC PDU.

3.1.1. BGP ASN

The four octet Autonomous System number MUST be specified. If the AS Number is less than 32 bits, it is padded with high order zeros.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Attr Type = 1 | Attr Len = 4  |            My ASN             ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   ~                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.1.2. BGP IPv4 Address

The BGP IPv4 Address sub-TLV announces the sender’s four octet IPv4 BGP peering source address and one octet Prefix Lenth to be used by the receiver. At least one of IPv4 or IPv6 BGP source addresses MUST be announced.

As usual, the BGP OPEN capability negotiation will determine the AFI/SAFIs to be transported over the peering, see [RFC4760].

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Attr Type = 2 | Attr Len = 5  |    My IPv4 Peering Address    ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   ~                               |   Prefix Len  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.1.3. BGP IPv6 Address

The BGP IPv6 Address sub-TLV announces the sender’s 16 octet IPv6 BGP peering source address and one octet Prefix Length to be used by the receiver. At least one of IPv4 or IPv6 BGP source addresses MUST be announced.

As usual, the BGP OPEN capability negotiation will determine the AFI/SAFIs to be transported over the peering, see [RFC4760].

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Attr Type = 3 | Attr Len = 17 |                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               +
   |                                                               |
   +                                                               +
   |                    My IPv6 Peering Address                    |
   +                                                               +
   |                                                               |
   +                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               |  Prefix Len   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.1.4. BGP Authentication sub-TLV

The BGP Authentication sub-TLV provides any authentication data needed to OPEN the BGP session. Depending on operator configuration of the environment, it might be a simple MD5 key (see [RFC2385]), the name of a key chain in a KARP database (see [RFC7210]), or one of multiple Authentication sub-TLVs to support hop[RFC4808].

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Attr Type = 4 |   Attr Len    |                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               ~
   ~                  BGP Authentication Data ...                  ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.1.5. BGP Miscellaneous Flags

The BGP session OPEN has extensive, and a bit complex, capability negotiation facilities. In case one or more extra attributes might be needed, the two octet BGP Miscellaneous Flags sub-TLV may be used.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Attr Type = 5 | Attr Len = 2  |           Misc Flags          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Misc Flags:

Bit 0: GTSM

Bit 1: BFD

Bit 2-15: Must be zero

The GTSM flag, when 1, indicates that the sender wishes to enable the [RFC5082] Generalized TTL Security Mechanism for the session.

The BFD flag, when 1, indicates that the sender wishes to enable the [RFC5880] Bidirectional Forwarding Detection for the session.

4. Security Considerations

All the Security considerations of [I-D.ymbk-idr-l3nd] apply to this PDU.

As the ULPC PDU may contain keying material, see Section 3.1.4, it SHOULD BE over TLS, not clear TCP.

Any keying material in the PDU SHOULD BE salted and hashed.

The BGP Authentication sub-TLV provides for provisioning MD5, which is a quite weak hash, horribly out of fashion, and kills puppies. But, like it or not, it has been sufficient against the kinds of attacks BGP TCP sessions have endured. So it is what BGP deployments use.

5. IANA Considerations

This document requests the IANA create a new entry in the L3ND PDU Type registry as follows:

PDU Code	PDU Name
9	ULPC

This document requests the IANA create a registry for L3ND ULPC Type, which may range from 0 to 255. The name of the registry should be L3ND-ULPC-Type. The policy for adding to the registry is RFC Required per [RFC5226], either standards track or experimental. The initial entries should be the following:

Value	Name
0	Reserved
1	BGP
2-255	Reserved

6. Acknowledgments

The authors thank Rob Austein, Sue Hares, and Russ Housley.

7. References

7.1. Normative References

[I-D.ymbk-idr-l3nd]
Bush, R., Housley, R., Austein, R., Hares, S., and K. Patel, "Layer-3 Neighbor Discovery", Work in Progress, Internet-Draft, draft-ymbk-idr-l3nd-05, 1 October 2022, <https://datatracker.ietf.org/doc/html/draft-ymbk-idr-l3nd-05>.

[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>.

[RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., Ed., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, DOI 10.17487/RFC5082, October 2007, <https://www.rfc-editor.org/info/rfc5082>.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <https://www.rfc-editor.org/info/rfc5226>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <https://www.rfc-editor.org/info/rfc5880>.

[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>.

7.2. Informative References

[RFC2385] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, DOI 10.17487/RFC2385, August 1998, <https://www.rfc-editor.org/info/rfc2385>.

[RFC4808] Bellovin, S., "Key Change Strategies for TCP-MD5", RFC 4808, DOI 10.17487/RFC4808, March 2007, <https://www.rfc-editor.org/info/rfc4808>.

[RFC7210] Housley, R., Polk, T., Hartman, S., and D. Zhang, "Database of Long-Lived Symmetric Cryptographic Keys", RFC 7210, DOI 10.17487/RFC7210, April 2014, <https://www.rfc-editor.org/info/rfc7210>.

Authors’ Addresses

Randy Bush
Arrcus & IIJ
5147 Crystal Springs
Bainbridge Island, WA 98110
United States of America
Email: randy@psg.com

Keyur Patel
Arrcus
2077 Gateway Place, Suite #400
San Jose, CA 95119
United States of America
Email: keyur@arrcus.com