# L3 Neighbor Discovery for BGP (and Scaled EVPN)

draft-ymbk-idr-l3nd

draft-ymbk-idr-l3nd-ulpc

IETF 113 IDR

2022.03.22

randy@psg.com , housley@vigilsec.com, sra@hactrn.net,

shares@ndzh.com, keyur@arrcus.com

( in some order )

# Reliable, Boring, Predictable, Measurable

# Do Not Run a DataCenter with 10,000 Devices on Probabilistic Protocols

# TL;DR

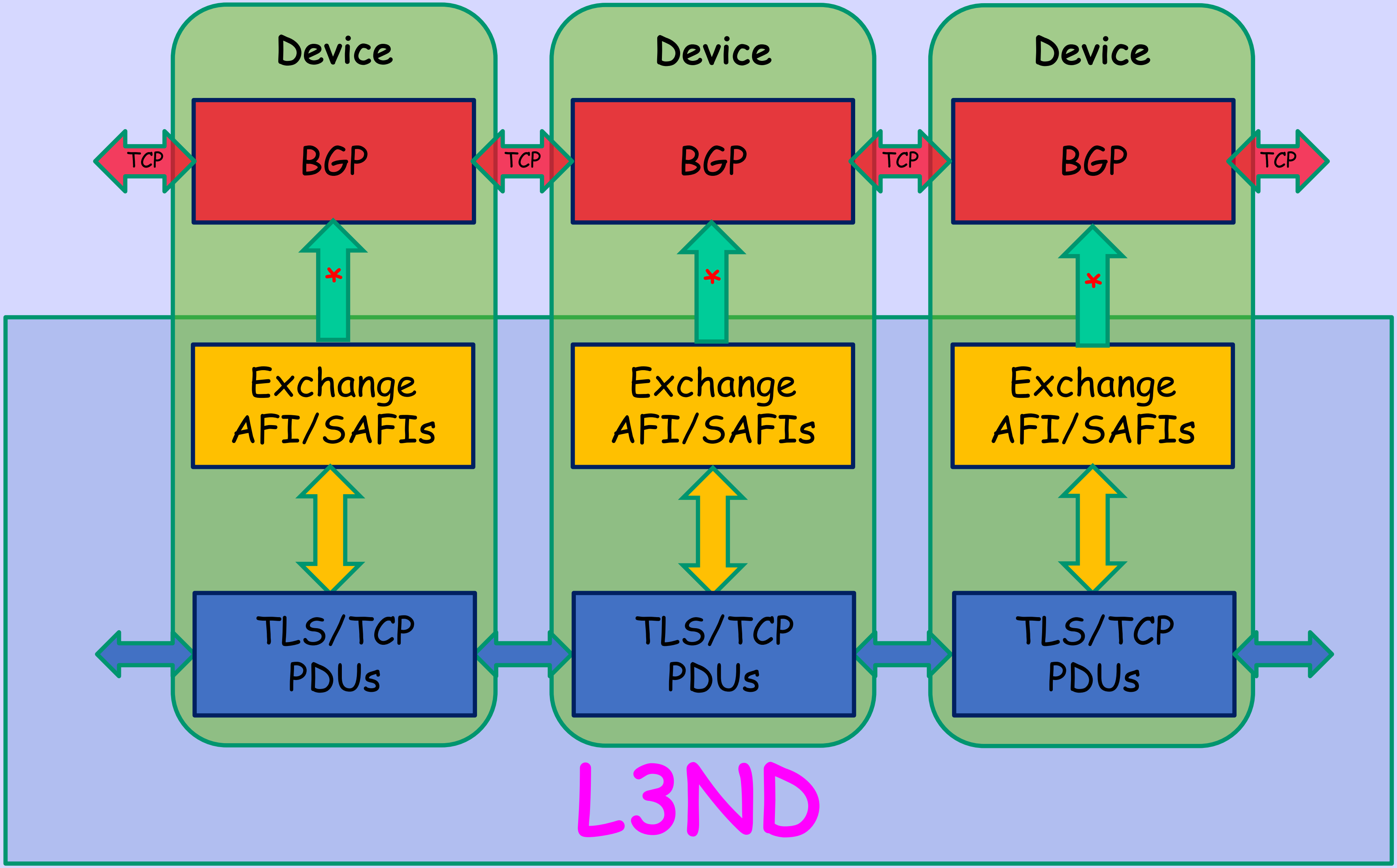
- L3ND like L3DL except it's Layer-3 not Layer-2
- Very similar Payload PDUs with same very large data (more for EVPN than BGP discovery)
- Multicast UDP HELLO for Initial Discovery
- Session Oriented and Resumable a la L3DL
- No Retransmission, Minimal Needed State Kept
- Guaranteed, Reliable, In-Order Delivery
- Transport over TCP, but TLS preferred (L3DL needed custom reliable transport)

Find Neighbor(s)

Learn L3 IP Addresses

Bootstrap BGP

# L2 Discover L3 Attrs



* see final slide

# This is NOT a Routing Protocol

Discovers the Layer 3 Addresses on PointToPoint or MultiPoint Links

# Basic PDU - TLV 101

```
 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  |    PDU Type   |         Payload Length        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               ~
~                          Payload ...                          ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

# L3 MultiCast UDP HELLO

```
 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  |  PDU Type = 0 |      Payload Length = 3       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |     Flags     |      Port     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |
+-+-+-+-+-+-+-+-+
```

Flags (bit):
   0 - 0 Raw TCP, 1 TLS
   1 - 0 Self-Signed Cert for TLS, 1 CA-based

Port is IANA Assigned TLS or TCP Server Port (Op may Override)

# TLS/TCP Session Open

- HELLO Sender knows its IP address
- HELLO Receiver knows Source Address of Sender
- Each has Sent and Received a HELLO
- Lowest IP Address provides TLS/TCP Server
- Highest IP Address acts as Client

# If TLS (recommended)

# The HELLO Specified CA-Based or Self-Signed Server Certificate

# Trust on First Use (TOFU)

- A Self-Signed Server TLS Certificate is generated on the TLS Server
- It is **Believed Without Question** by the TLS Client
- You do get Integrity and knowing your Peer (Attacker or otherwise) has not changed on Restart

# CA-Based PKI Keying

- A Server's Certificate is signed by the the operational environment's Trust Anchor
- The TLS Server MUST Use that Cert
- The TLS Client can be confident that the TLS Server is under control of the identified Trust Anchor for which the Client has the Public Certificate

The Choices of
TLS or Naked TCP
and, if TLS, of
TOFU or Trust Anchor
are for the Operator

# L3ND Session OPEN

```
 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  |  PDU Type = 1 |         Payload Length        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |          Session ID           ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |         Serial Number         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |   AttrCount   |               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+               +
~                      Attribute List ...                       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Session ID – Unique Nonce per Session to Allow Restart
- Serial Number PDU *timestamp* allows Session Restart
- Attributes such as Leaf, Spine, … are User Defined

# All PDUs are ACKed

```
 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  |  PDU Type = 3 |       Payload Length = 6      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               | ACKed PDU Type|     EType     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Error Code          |           Error Hint          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

EType
  0 - No Error, Error Code and Error Hint MUST be zero
  1 - Warning, something not too serious happened, continue
  2 - Session should not be continued, try to restart from HELLO
  3 - Restart is hopeless, call the operator

Error Code and Error Hint give details
Error Code is an IANA Table, see I-D

Fully Stateful
Session Per Peer

Graceful Restart

State May Be Resumed
á la BGP

# IPv4 Encapsulation PDU

```
 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  |  PDU Type = 4 |         Payload Length        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |             Count             ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |                 Serial Number                 ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |  Encaps Flags |          IPv4 Address         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |   PrefixLen   |    more ...   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


0           1            2            3            4  ...       7
+------------+------------+------------+------------+------------+
|  Ann/With  |   Primary  | Under/Over |  Loopback  | Reserved ..|
+------------+------------+------------+------------+------------+
```

# IPv6 Encapsulations and MPLSv4 and MPLSv6 as Expected

# L3ND-ULPC

## Upper Layer Protocol Configuration

draft-ymbk-idr-l3nd-ulpc

Meant to Allow Config of Arbitrary L3+ Protocols

So Far Only Defined for BGP

# L3ND PDU for ULPC

```
 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 Sub-TLV List ...                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

ULPC Type

| | |
|---|---|
| 0 | : Reserved |
| 1 | : BGP |
| 2-255 | : Reserved |

Provide the <u>Minimal</u> set of Configuration Parameters for BGP OPEN to Succeed

Not to replace or conflict with data exchanged by BGP OPEN

Multiple sources of truth are a recipe for complexity and pain

# AS and Peering IP

```
 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            ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
 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  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

# Auth Data and GSTM

```
 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 ...                 ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


 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 1-15:   Must be zero
```

# Yes, there is one for IPv6 ☺

Remember that the Base L3ND Protocol Provided and Marked Loopbacks etc.

# Features or Bugs?

- Stateful and Re-Startable
- Handshakes/ACKs; Provide Error Reporting, Pacing, and Solid Confirmation
- TCP/TLS; You Have BGP, You Have TCP
- You Want Security, Do You Roll Your Own or Just Use TLS?
- L3ND Provides Large Scale, Probably More Than BGP Needs; But it Only Costs a Few Bits in the Length Fields

# That's It

# But …

# Still Do Not Understand

- How Parameters (BGP, etc.) are Passed to Forwarding (for loopbacks), BGP, etc.?
- How is BGP Started, Restarted, Stopped?
- When is Discovery Finished and Should be Stopped?
- Does even Highly Scaled EVPN Need the Restartability Hacks?

[ Note that L3DL Uses a Minimal Bit of BGP-LS to Communicate with BGP ]