

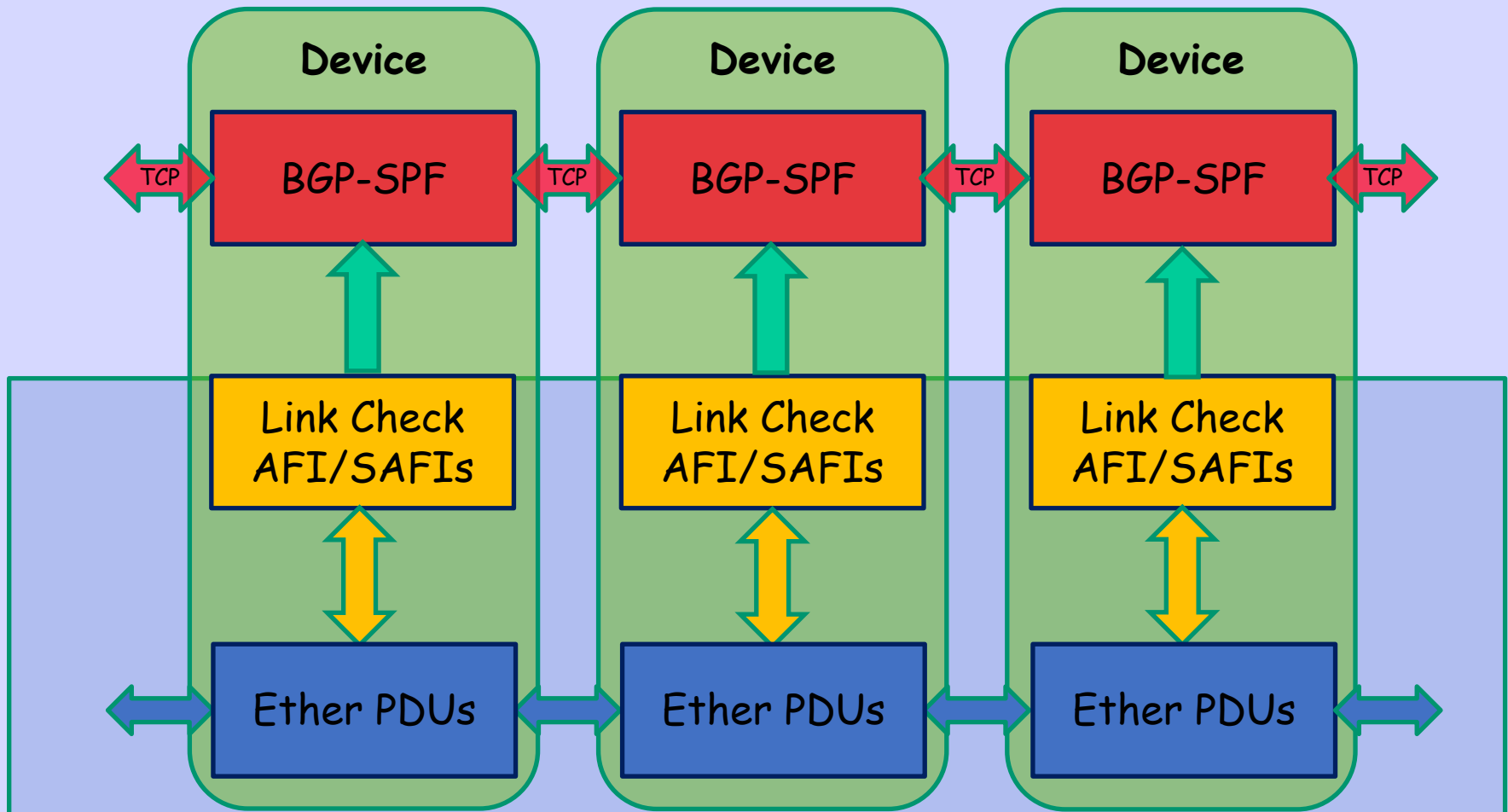
Link State Over Ether

What's Changed

Randy Bush Arrcus & IIJ Lab

Rob Austein, Arrcus

Keyur Patel, Arrcus



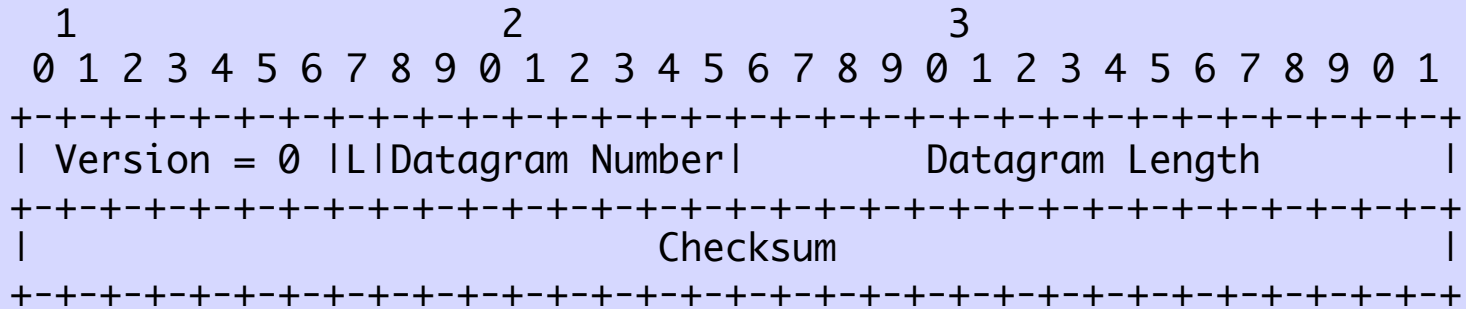
East West Protocol

New Data Framing

PDU

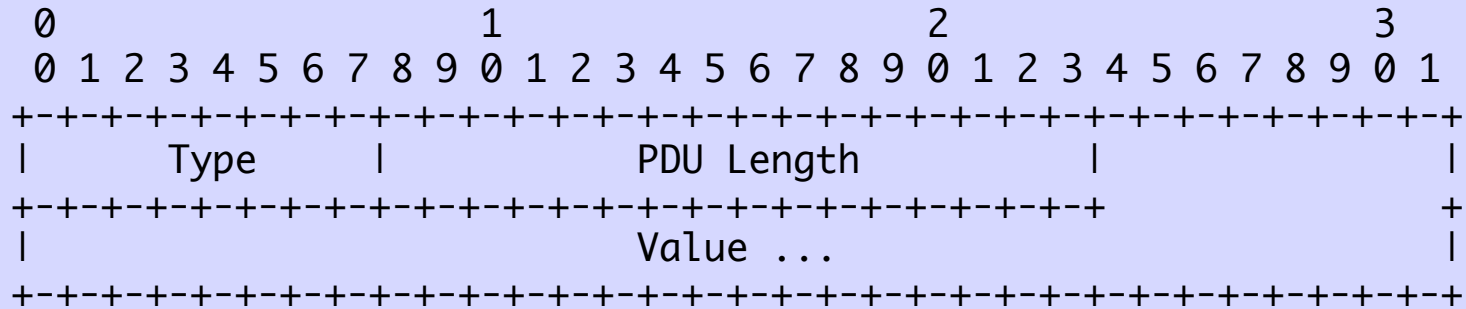
- A *PDU* (Protocol Data Unit) is an application layer message
- It may occupy multiple *Datagrams*
- Datagrams are one per Ethernet Frame

Datagram



- A Datagram is one Ethernet Frame
- A Datagram has *Number*, *Length*, and *Checksum*
- The *L* flag is set on the last datagram of an application layer PDU
- This Transport Layer assembles PDUs

Every Datagram a TLV



0 - HELLO

1 - OPEN

2 - KEEPALIVE

3 - ACK

4 - IPv4 Announcement

5 - IPv6 Announcement

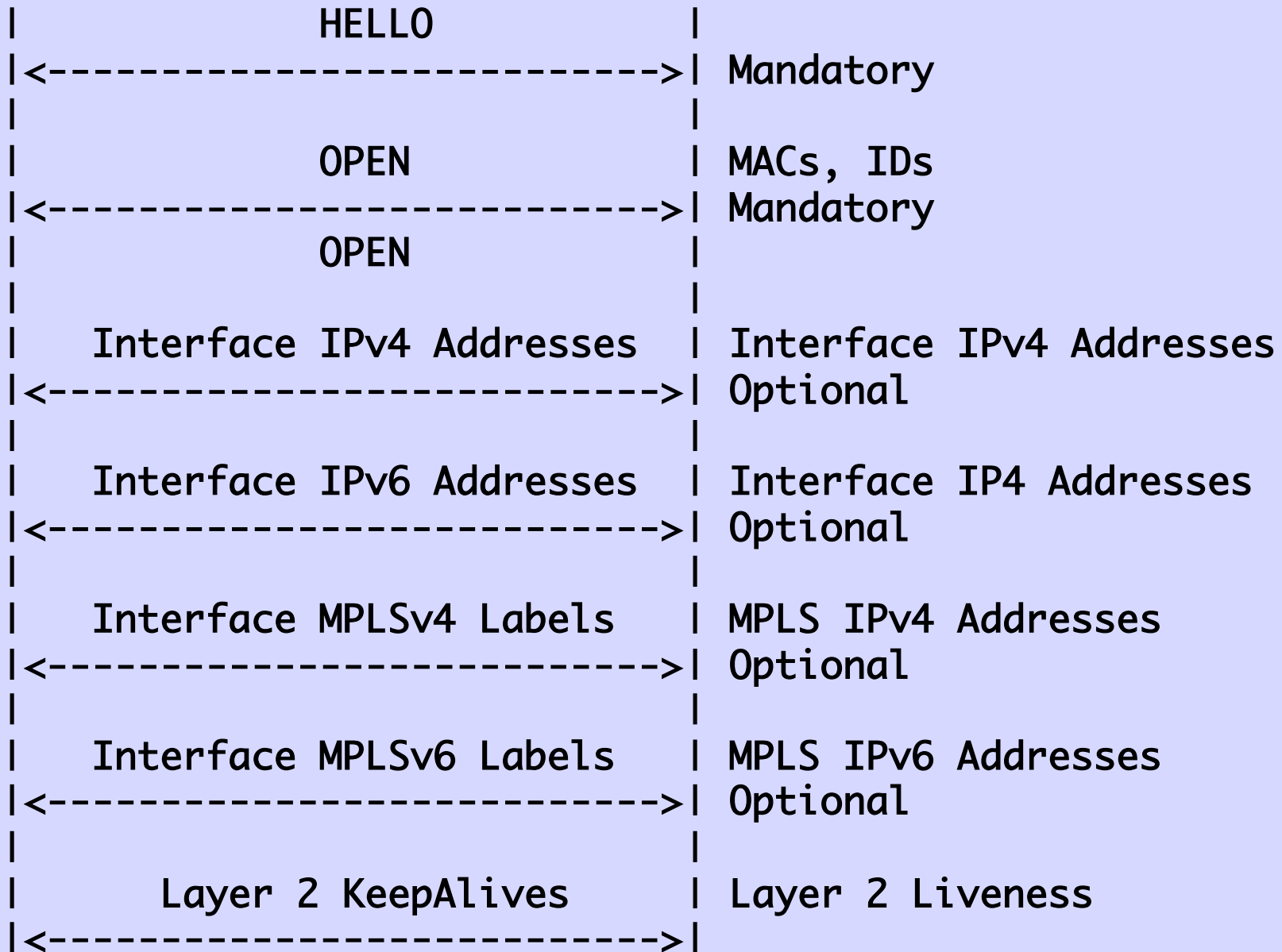
6 - MPLS IPv4 Announcement

7 - MPLS IPv6 Announcement

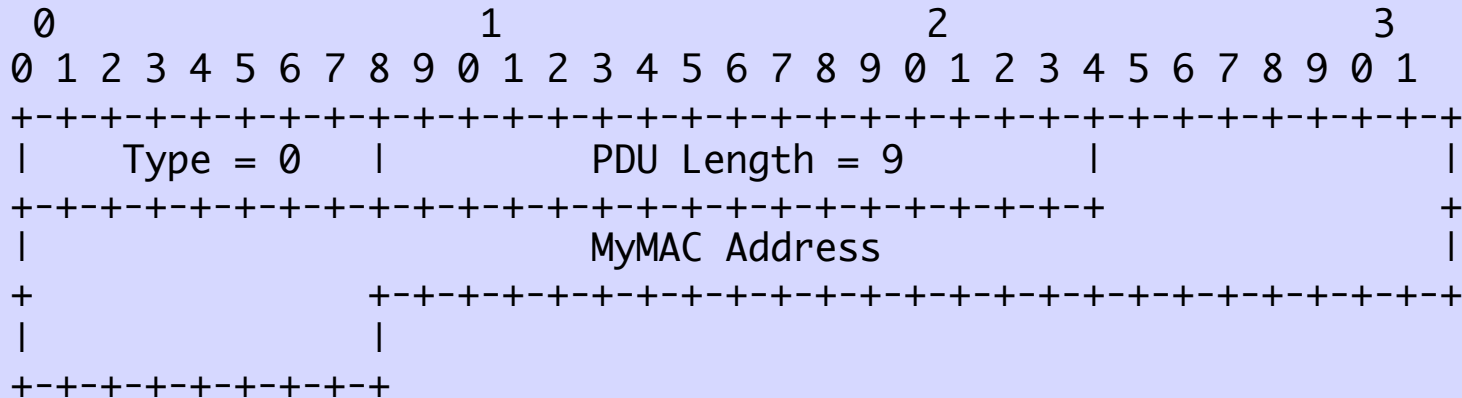
Sessions are Made Clear

OPEN is ACKed as well
as Encapsulation PDUs

Inter-Link Ether Protocol

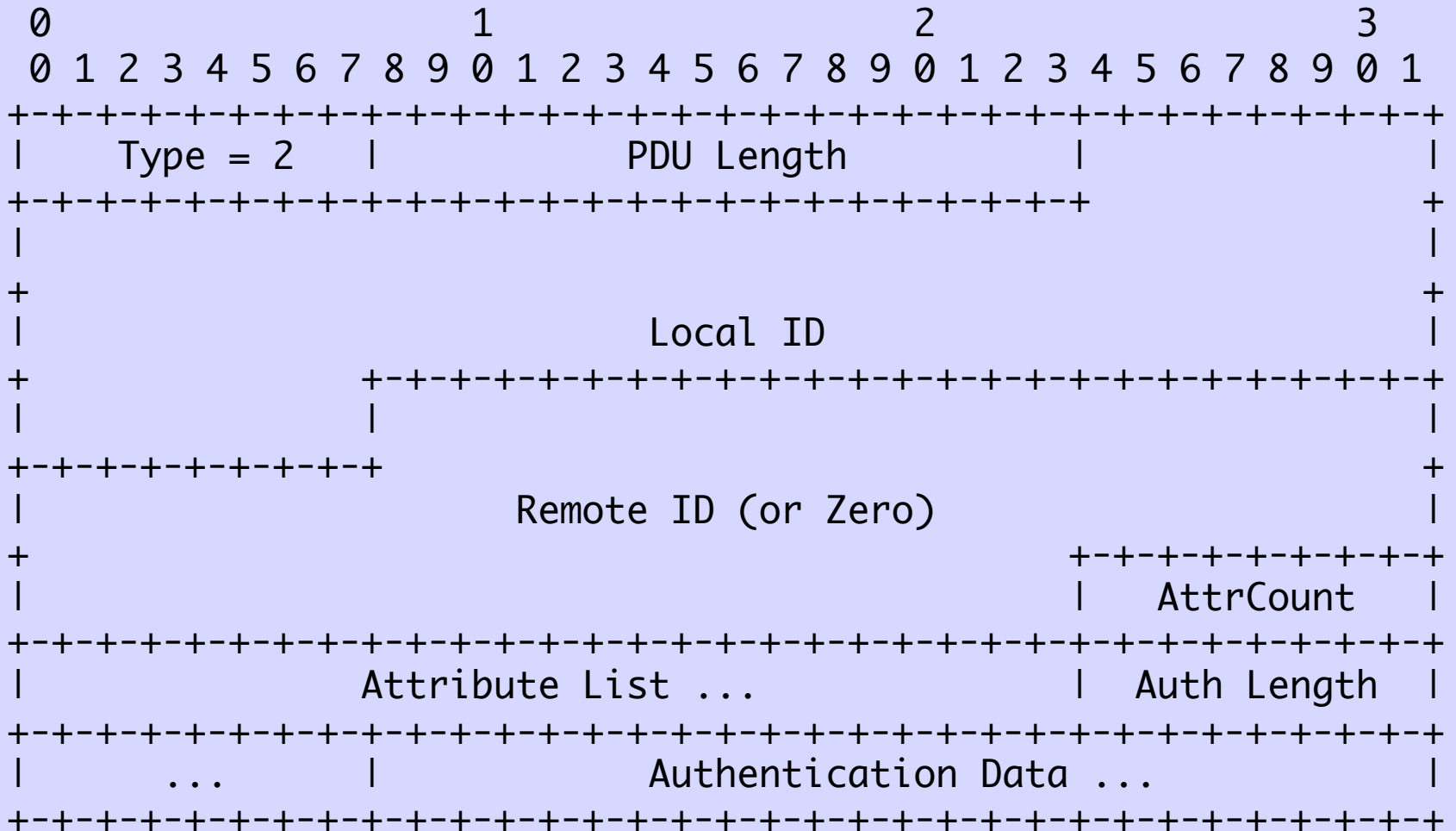


Link HELLO



- HELLO is Multicast, à la LLDP
- Each device learns the other's MAC from its HELLO whining. All devices on a wire/interface know each others MACs and learn each other's IDs
- Respond with OPEN
- A multi-point topology is a set of point-to-point links

OPEN



Local/Remote IDs

Might be

- an ASN with high order bits zero
- a classic RouterID with high order bits zero
- a catenation of the two
- a 80-bit ISO System-ID
- or any other identifier unique to a single device in the current routing space

Attributes

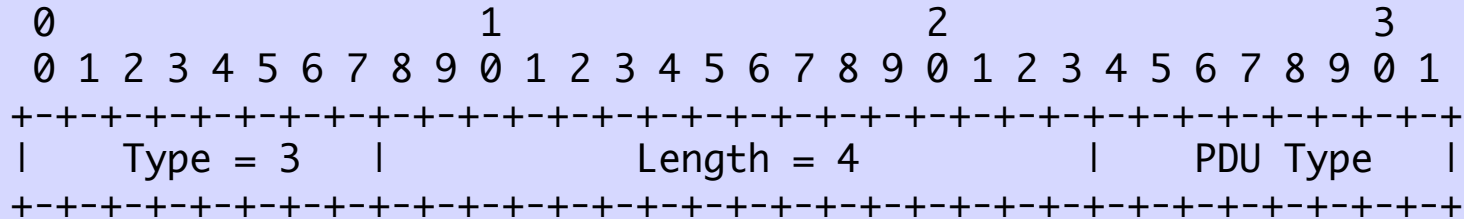
A node may have zero or more user-defined attributes, e.g. spine, leaf, backbone, route reflector, arabica, ...

Nodes exchange their attributes only in the OPEN message

Authentication Data

- Specific to the Operational Environment
- Might be Certificate derived from Op's CA
- Failure to authenticate is a failure to start the LSOE association, and HELLOs MUST BE restarted.

ACK

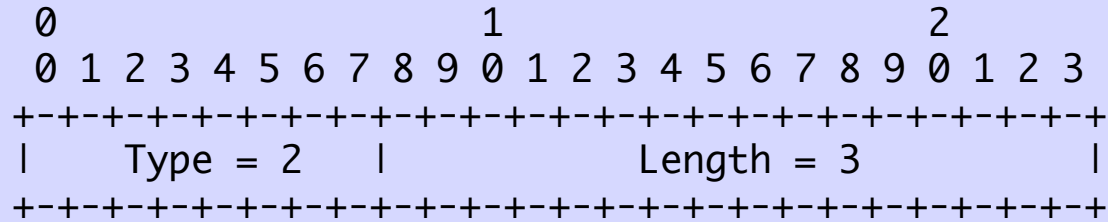


- The Receiver ACKs with a Type=3
- PDU Type is the Type of the PDU being ACKed
- If the Sender does not receive an ACK in one second, they retransmit. Operator configured failure count.

Once We Know
Each Other's MACs

Layer Two KeepAlives
May be Started

L2 KEEPALIVE



This is in addition to L3 BFD etc.

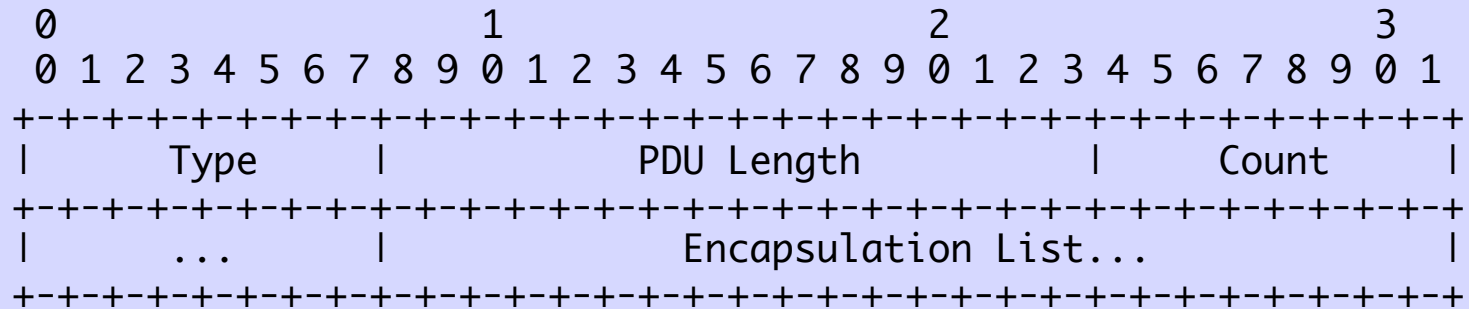
We assume that one or more Encapsulation addresses will be used to ping, BFD, or whatever the operator configures

We Know MAC/Ether Link State
of This Device & Neighbor

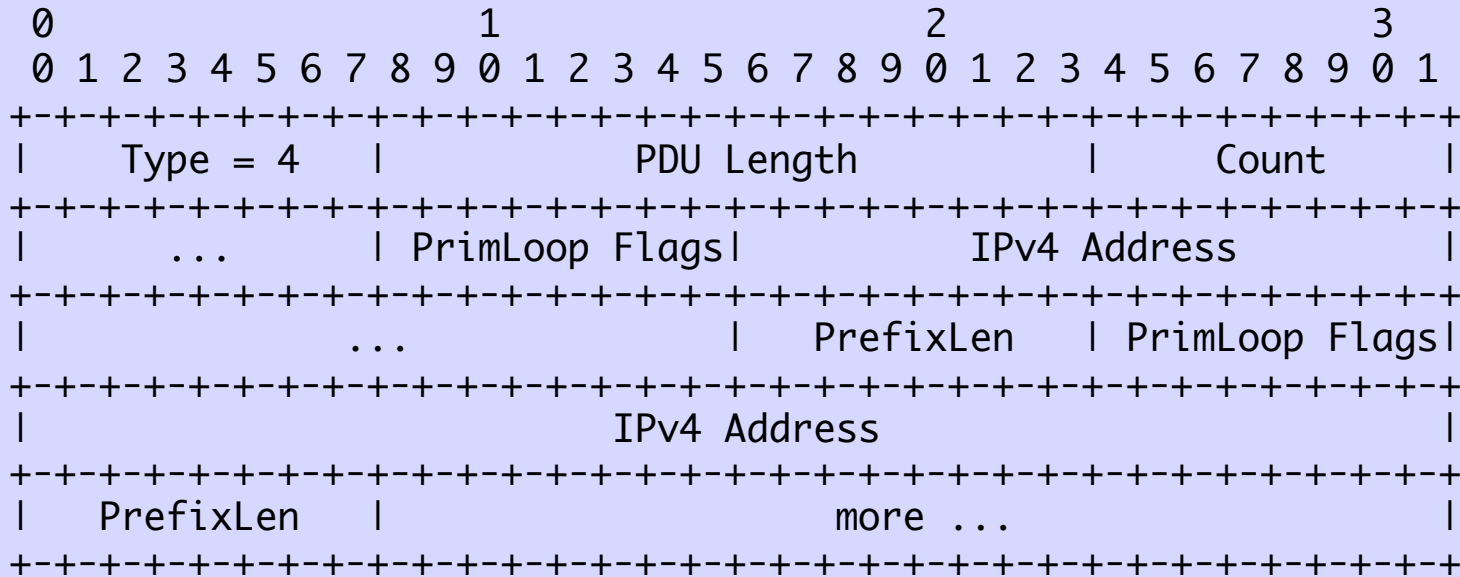
And Node IDs (often ASNs)

Now Announce Encapsulations
of the Link Interfaces

Encaps PDU Header



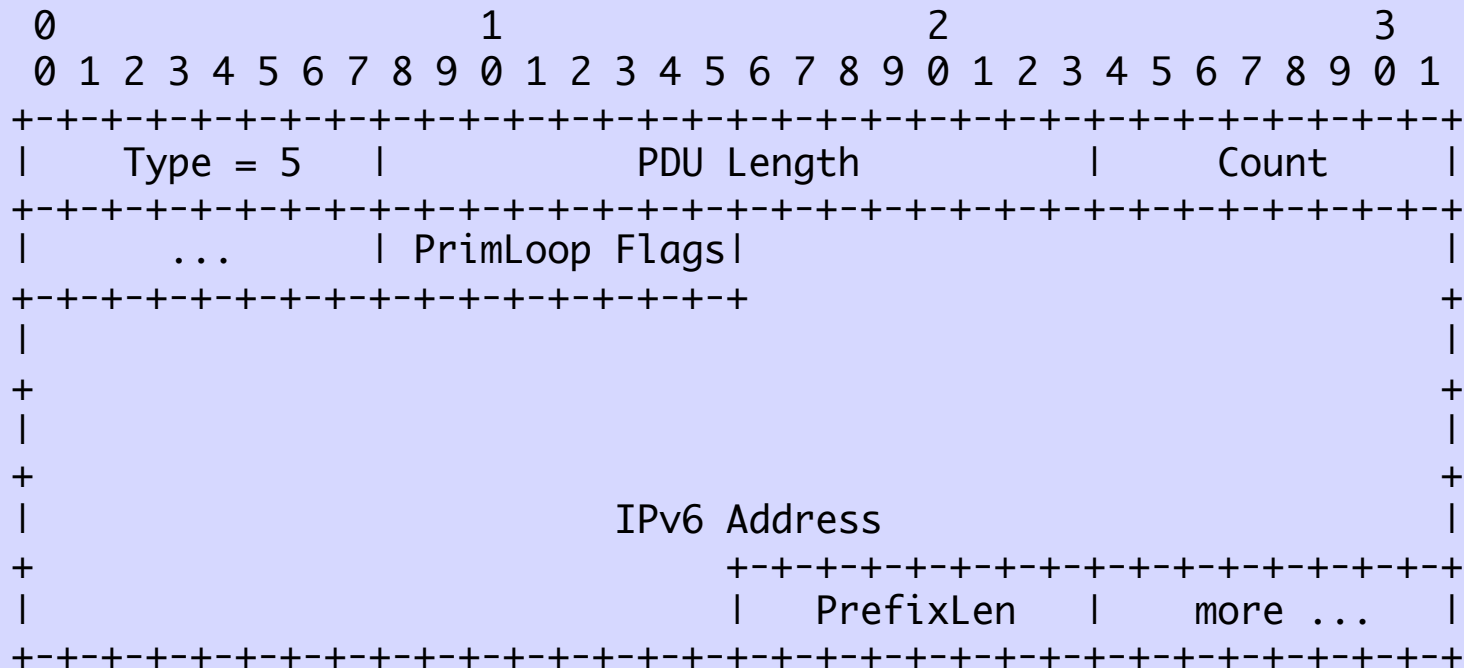
IPv4 Encapsulations



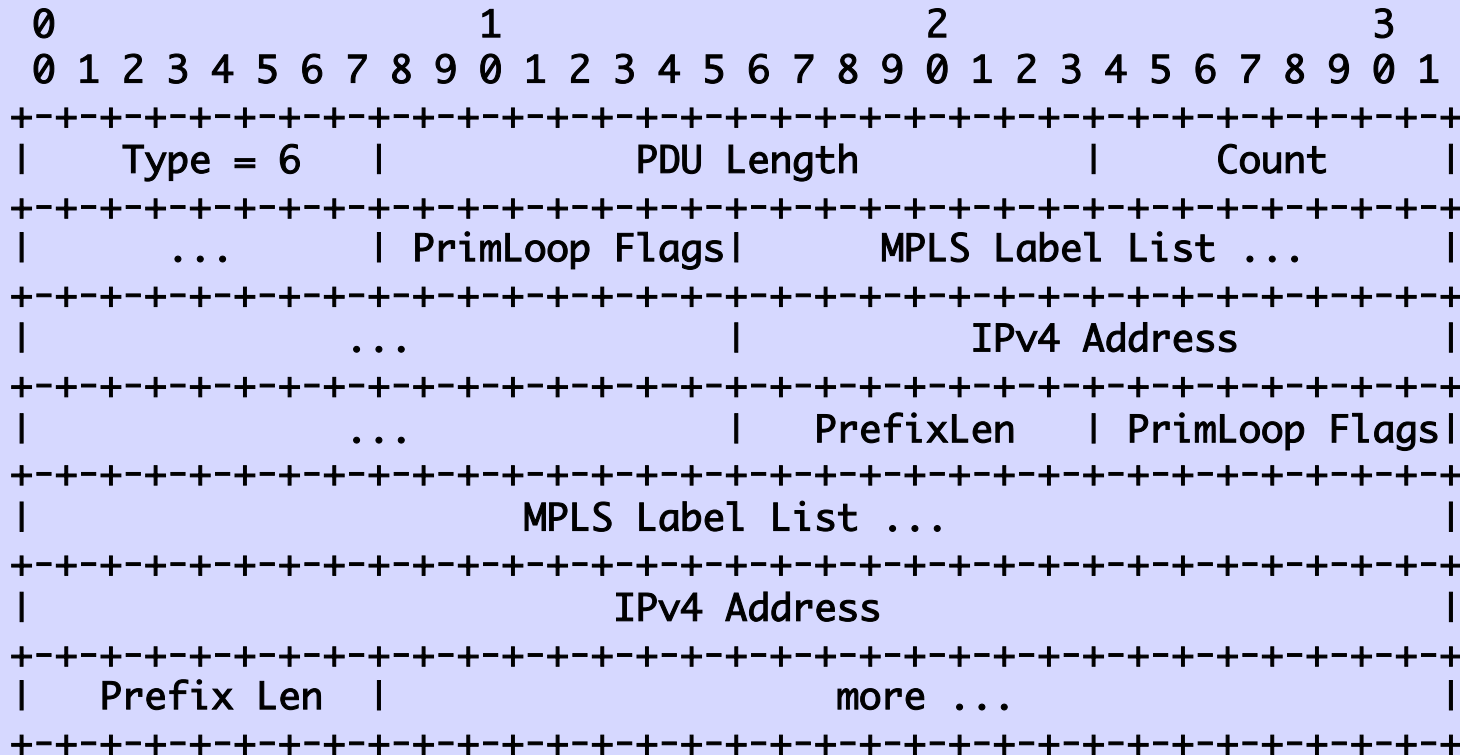
An Encapsulation message describes zero or more addresses of the encapsulation type.

An Encapsulation message of Type T replaces all previous encapsulations of Type T

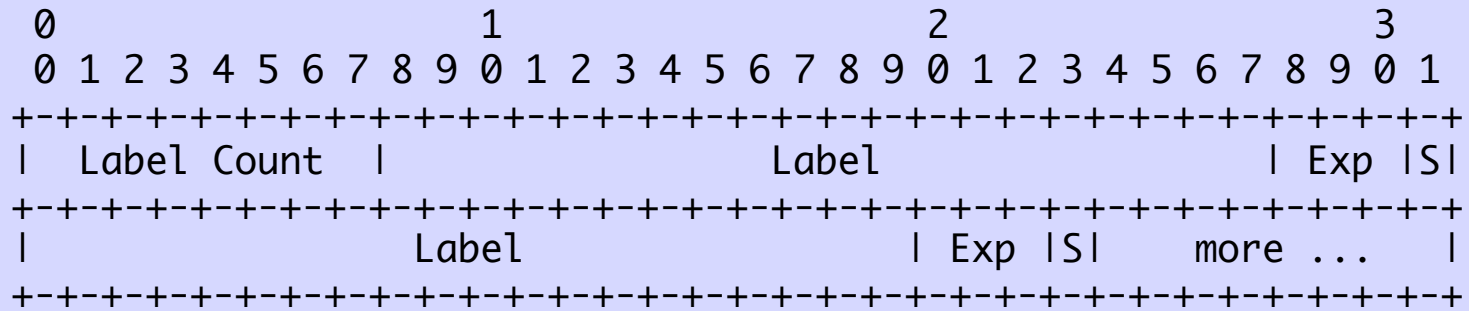
IPv6 Encapsulations



MPLS IPv4 Encapsulations

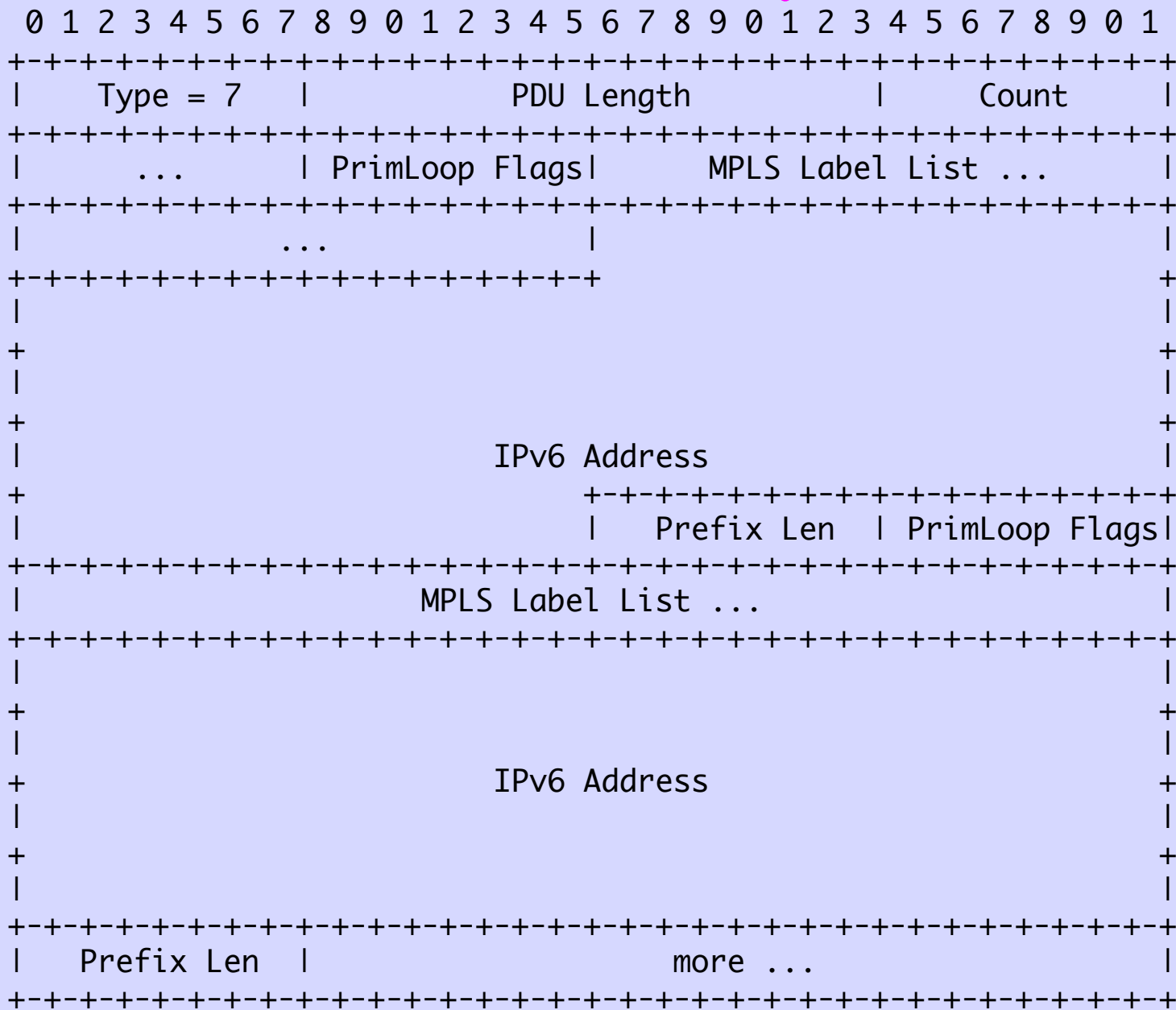


MPLS Label List



Use Multiple MPLS Label Encapsulations to Allow One Label to be Associated with Multiple AFI/SAFIs and/or Multiple IP Addresses

MPLS IPv6 Encapsulations



We're Looking at Security

Are you the Droid
I was talking to
Earlier?

Thinking of Security

- OPEN has public key plus ...
- Signed with private key, proves possession
- All PDUs signed with 512-bit suffix
- KEEPALIVE could get a sequence number to detect replay attack
- Maybe a later Proof of Possession Challenge / Response PDU pair

Layer-3 IP/Label Liveness Should Also be Tested

One or more Discovered
AFI/SAFI Addresses Are
Used to Ping, BFD, ... to
Assure Layer-3 Liveness

We now know all links, IDs,
Encapsulation Types, and
Addresses of this Device

Now Present an API to
Topology and Dijkstra Layers

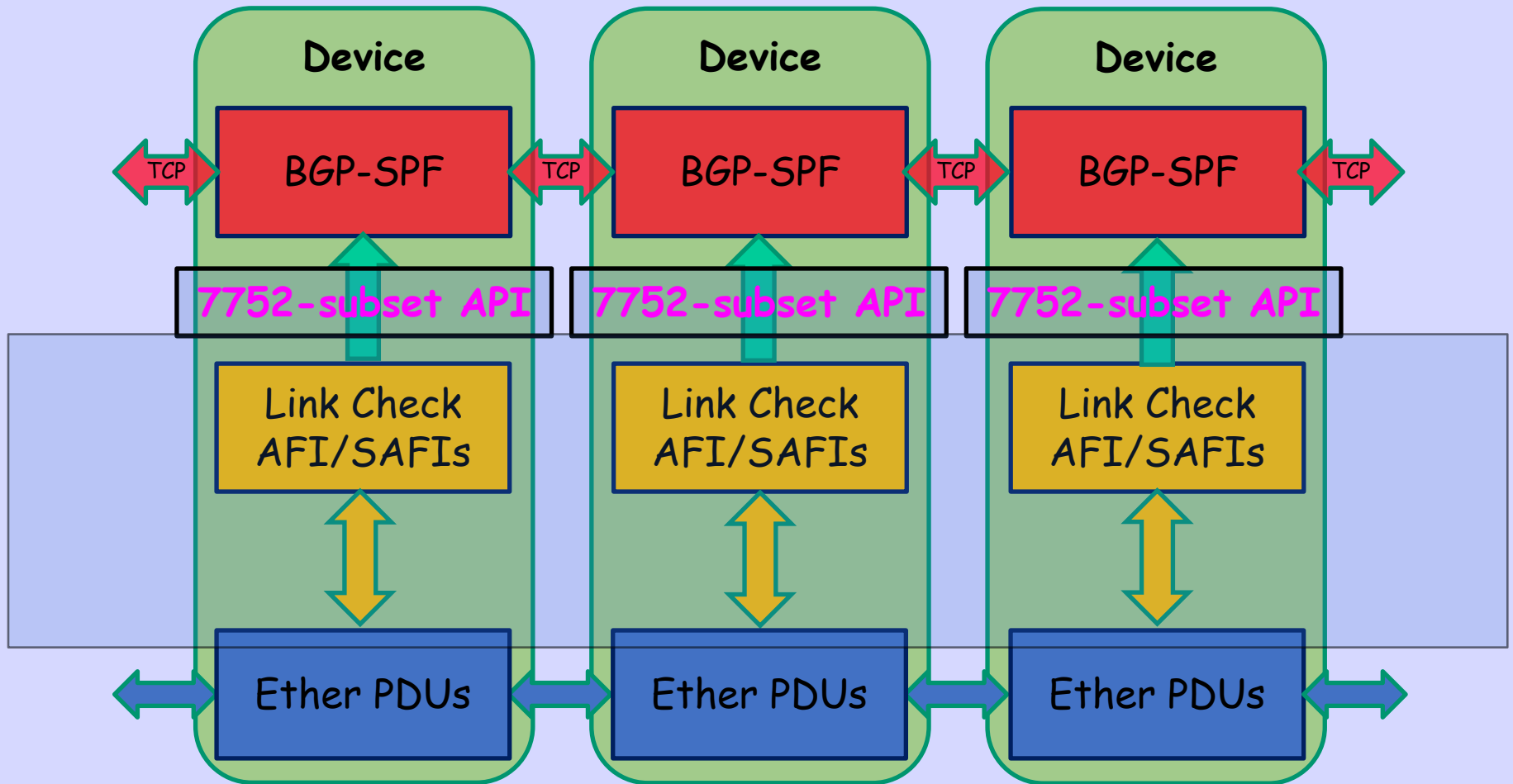
BGP-LS (RFC 7752)

an extension to BGP to

distribute the network's

link-state (LS) topology

North/South Protocol



Node Descriptors

- Similarly to BGP-SPF, the BGP protocol is used in the Protocol-ID field specified in table 1 of draft-ietf-idr-bgpls-segment-routing-epe.
- The local and remote node descriptors for all NLRI are the ID's described in Section 5.3.
- This is equivalent to an adjacency SID or a node SID if the address is a loopback address.

IPvX Links

TLVs 259 and 260 are used. And for IPv6 links, TLVs 261 and 262. If there are multiple addresses on a link, multiple TLV pairs are pushed North, having the same ID pairs.

MPLS Links

Label Sub-TLVs from draft-ietf-idr-bgp-ls-segment-routing-ext Section 2.1.1, are used to associate one or more MPLS Labels with a link.

And Bob's Your Uncle

Open Questions

Should HELLO go
Through
an intermediate
Layer Two Switch

Are HELLO and
KEEPALIVE
Redundant?

BTW,
There is No IPR