Hewlett Packard Enterprise

HPE Aruba Networking

LLDP Between LSVR Routers

Paul Bottorff

04/28/2025

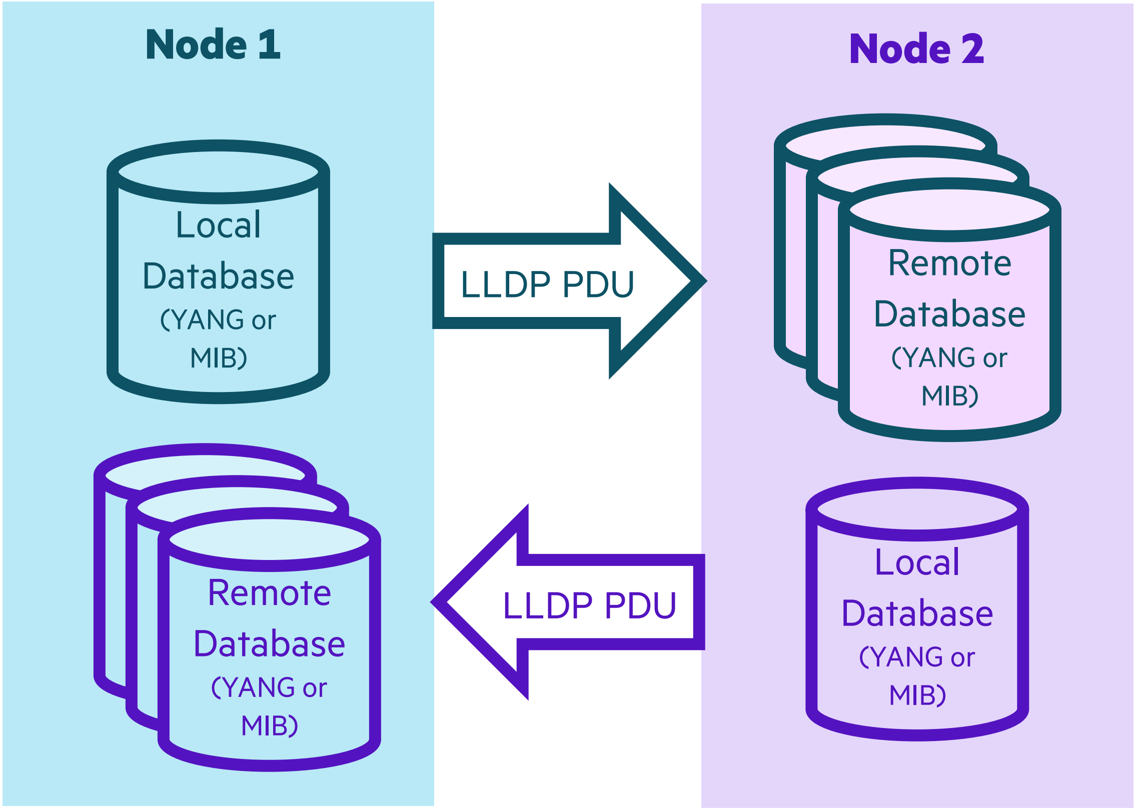
Recap

- The IEEE 802.1 WG has sent Liaison Statements (LSs) to the IETF LSVR WG
  - https://datatracker.ietf.org/liaison/1985/ (2025-03-16)
  - https://datatracker.ietf.org/liaison/1637/ (2019-03-28)
    - in response to LSVR WG's LS: https://datatracker.ietf.org/liaison/1621/ (2019-01-10)
  - The IEEE 802.1 LSs suggest the use of decades old, widely deployed Layer 2 (Link Layer) standards specified by IEEE 802.1 instead of the IETF specifying new Layer 2 protocols for the same purpose
    - The Link Layer Discovery Protocol (LLDP) for discovery
    - Continuity Check Messages (CCMs) of Connectivity Faut Management (CFM) for Layer 2 liveliness

- The IEEE 802.1 WG has amended the LLDP standard for the requirements of the IETF LSVR WG
  - **IEEE 802.1ABdh-2021**: IEEE Standard for Local and metropolitan area networks – Station and Media Access Control Connectivity Discovery Amendment 2: **Support for Multiframe Protocol Data Units**
  - https://standards.ieee.org/ieee/802.1ABdh/7667/
    - free access vie the IEEE Get Program (https://ieeexplore.ieee.org/browse/standards/get-program/page)

- LLDP-based contributions to the LSVR WG so far:
  - https://datatracker.ietf.org/doc/draft-bottorff-lsvr-lldp-tlvs/00/ (2025-04-30)
  - https://datatracker.ietf.org/doc/draft-congdon-lsvr-lldp-tlvs/00/ (2019-10-22)
  - https://datatracker.ietf.org/meeting/104/materials/slides-104-lsvr-3-update-ieee-ietf-liaison-response-and-intro-to-lldpv2-proposal-01
  - https://datatracker.ietf.org/meeting/105/materials/slides-105-lsvr-2-ieee-lldpv2-update-00
  - https://datatracker.ietf.org/meeting/106/materials/slides-106-lsvr-3-update-lsvr-ietf-organizationally-specific-tlvs-for-ieee-std-8021ab-00
  - https://datatracker.ietf.org/meeting/106/materials/slides-106-lsvr-4-status-update-lsvr-ieee-p8021abdh-lldpv2-00

LLDP is an extremely simple database exchange protocol



- Nodes periodically advertise information from its local database to all of its neighbors in LLDP Protocol Data Units (PDUs)
- Nodes extract the information from the received LLDP PDUs and store it in a remote database for each neighbor
- The databases can be retrieved by a management entity or an application to gather a view on the network topology and the details of network nodes
- LLDP is designed to be easily extendible
  - LLDP PDUs carry TLVs
  - New TLVs can be defined to carry nodal information not yet covered by existing TLVs
  - For instance, IETF can specify new TLVs under IETF IANA OUI

Applying LLDP to LSVR Discovery and Configuration

- Discovery can be achieved by running LLDP at each LSVR Logical Link Endpoint (LLE)
  - Enabled by using a (to be assigned) Nearest Router Group Address for the LLDP Scope MAC Address
  - The Logical Link Endpoint Identifier (LLEI) is then encoded in the LLDP ChassisID + PortID
- Discovery of L3 and L2.5 encapsulations and addressing can be achieved by using LLDP to advertise a set of LLDP Organizationally Specific TLVs which encode the encapsulation and addressing available at each Logical Link Endpoint
  - LLDP builds a database at each LLE containing all the remote information from each discovered LLE
  - An LSVR LLDP application then determines available Logical Links, compatible encapsulations, and addresses by comparing the information in the local and remote databases at each LLE
- The LSVR LLDP application can use any interface desired to present the information to BGP-SPF
- L3 liveness can be enabled using Organizationally Specific TLVs
- L2 liveness can be supported using both the LLDP time-to-live and Connectivity Fault Management (CFM)
  - LLDP always provides time-to-live for each element of the remote database using the LLDP Time-To-Live TLV
  - CFM (IEEE Std 802.1Q cl 18-22) Connectivity Check Messages (CCMs) can be used for fast link failure detection
- LLDP can communicate the parameters for BGP (and other) protocols using Organizationally Specific TLVs
  - I-D.acee-idr-lldp-peer-discovery specifies LLDP configuration TLVs for BGP
- LLDP can be secured using the MACsec (IEEE Std 802.1AE)

LLDP IETF Organizationally Specific TLVs for LSVR

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=127    |   Length          |         OUI = 00-00-5E        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |  Subtype = xx | LsvrType = e |  LSVR Info      ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                             more …                              ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Type = 127: designates this as an Organizationally Specific TLV
- Length: is the total length starting after the length field. The length is from 5-511.
- Organizationally Unique Identifier (OUI): IETF IANA OUI from RFC 7042. Value 00-00-5E
- Subtype = xx: To be determined by IANA for LSVR (refer to I-D.acee-idr-lldp-peer-discovery)
- LsvrType = e: The LSVR TLV type
- LSVR Infor: The LSVR information field

Summary

- LLDP (IEEE Std 802.1AB) is a mature, tested, and widely implemented protocol

- The IEEE has enhanced Classic LLDP through amendments supporting:
  - Large database sizes (IEEE Std 802.1ABdh-2021)
  - Modern YANG data models (IEEE Std 802.1ABcu-2021)

- The LLDP protocol is capable of addressing the LSVR discovery and configuration requirements

- Additional L2 liveliness support can be obtained by using Connectivity Fault Management (IEEE Std 802.1Q cl 18-22) Connectivity Check Messages (CCMs)

- LLDP-based contributions to the LSVR WG so far:
  - https://datatracker.ietf.org/doc/draft-bottorff-lsvr-lldp-tlvs/00/ (2025-04-30)
  - https://datatracker.ietf.org/doc/draft-congdon-lsvr-lldp-tlvs/00/ (2019-10-22)
  - https://datatracker.ietf.org/meeting/104/materials/slides-104-lsvr-3-update-ieee-ietf-liaison-response-and-intro-to-lldpv2-proposal-01
  - https://datatracker.ietf.org/meeting/105/materials/slides-105-lsvr-2-ieee-lldpv2-update-00
  - https://datatracker.ietf.org/meeting/106/materials/slides-106-lsvr-3-update-lsvr-ietf-organizationally-specific-tlvs-for-ieee-std-8021ab-00
  - https://datatracker.ietf.org/meeting/106/materials/slides-106-lsvr-4-status-update-lsvr-ieee-p8021abdh-lldpv2-00

- The IEEE 802.1 suggests the use of LLDP rather than IETF specifying a new Layer 2 protocols for the same purpose

Thank You

Backup

Classic LLDP single frame format (IEEE Std 802.1AB-2016)

DA Scope MAC Address	SA	LLDP Ethertype	ChassisID TLV (source)	PortID TLV (source)	TTL TLV	Optional TLV	. . .	Optional TLV	End of LLDPDU TLV
			M	M	M				
Ethernet Header			LLDP PDU						

LLDP TLV Format

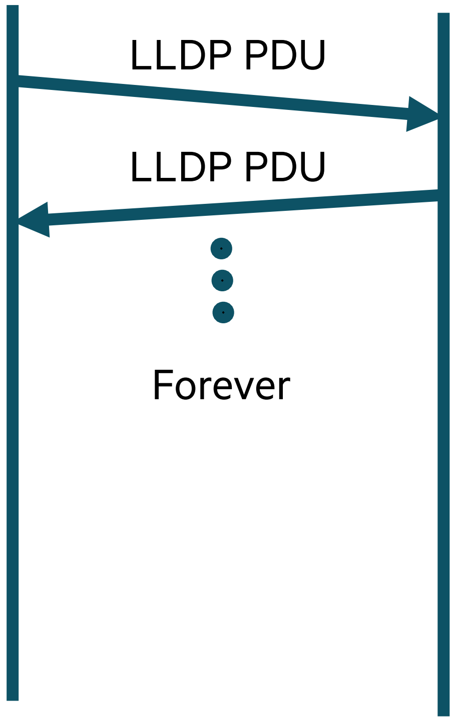
TLV type (7 bits)	TLV information String length (9 bits)	TLV information

LLDP Organizationally Specific TLV Format

TLV type = 127 (7 bits)	TLV information String length (9 bits)	OUI/CID (3 octets)	Subtype (1 octet)	TLV information

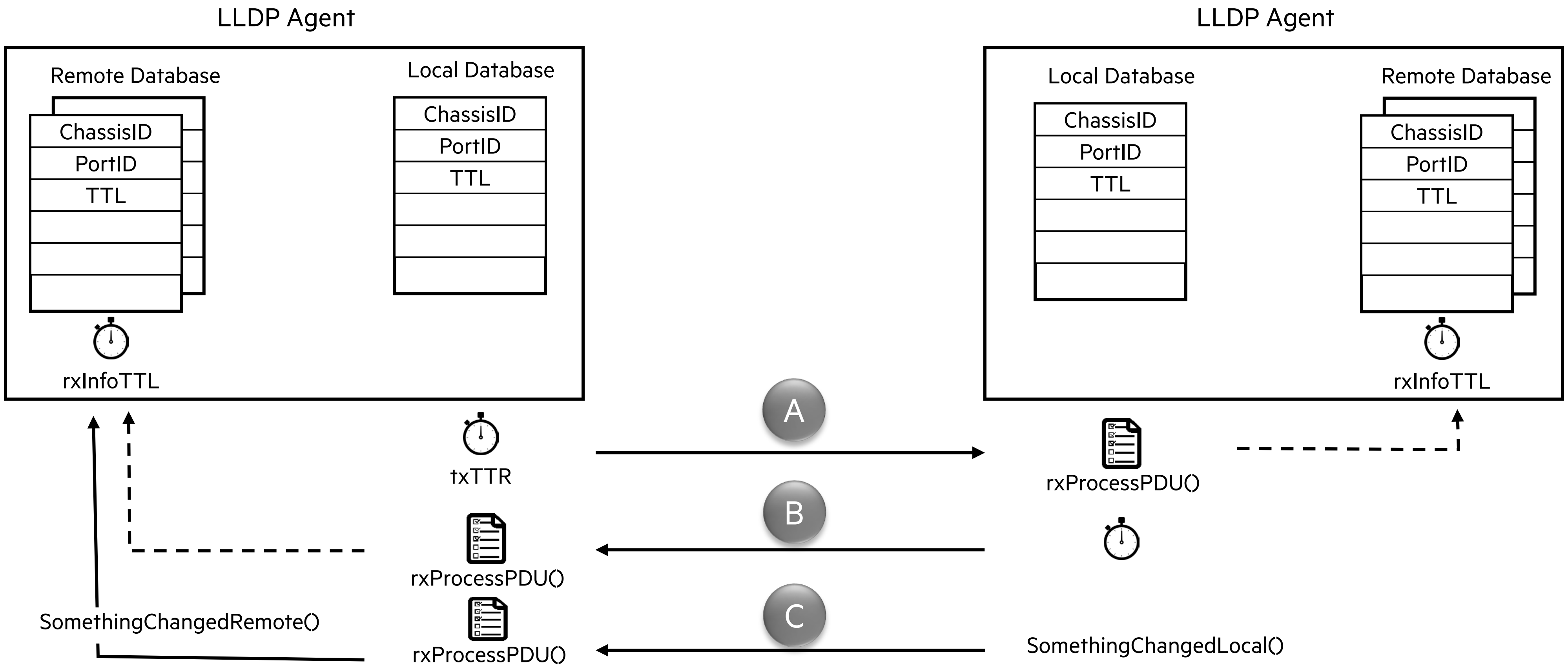
- The Scope MAC Address of the single frame Normal LLDPDU determines the limits of propagation of LLDPDUs. It can be the:
  - Nearest bridge group address, nearest customer bridge group address, nearest non-TPMR bridge group address
  - Also, the standard allows any group MAC address or any individual MAC address
- Each LLDPDU includes three mandatory TLVs which are the ChassisID, PortID, and Time To Live TLV
  - ChassisID provides a unique system identifier, i.e. MAC Address, IP Address, other management identifier
  - PortID provides a unique identifier for the interface (or sub-interface) within the system, i.e. IfIndex, Port Number
  - Time To Live provide duration the LLDPDU is valid
- Optional TLVs contain further information
  - An 802.1ABdh Manifest TLV can be used to extend the database size up adding up to 83 additional frames of TLVs

Classic LLDP ladder: (802.1AB-2016)



- Each LLDP PDU is transmitted to a destination MAC address called the Scope MAC Address, typically a group address that determines were the LLDP data bases reside.
- Each LLDP PDU is a single frame containing the TLVs which encode the LLDP information from the sender
- Every LLDPDU repeats all the sender's current information (in TLVs)
- LLDP PDU transmissions repeat at an average rate (default 30 seconds)
- Whenever a new device enters the network or when any change occurs in the content of the LLDPDU a series of fast LLDPDU transmissions occurs (default 4 LLDPDU spaced every second)

Classic LLDP single frame operation (IEEE Std 802.1AB-2016)



A B Each port periodically advertises its local LLDP database
A B On receipt of an LLDP advertisement, LLDP updates the remote database for the sender
C When something changes in the local database LLDP immediately transmits and update

A Manifest TLV is used to extend the LLDP database (IEEE Std 802.1ABdh)

DA Scope MAC Address	SA	LLDP Ethertype	ChassisID TLV (source)	PortID TLV (source)	TTL TLV	Manifest TLV	. . .	Optional TLV	End of LLDPDU TLV

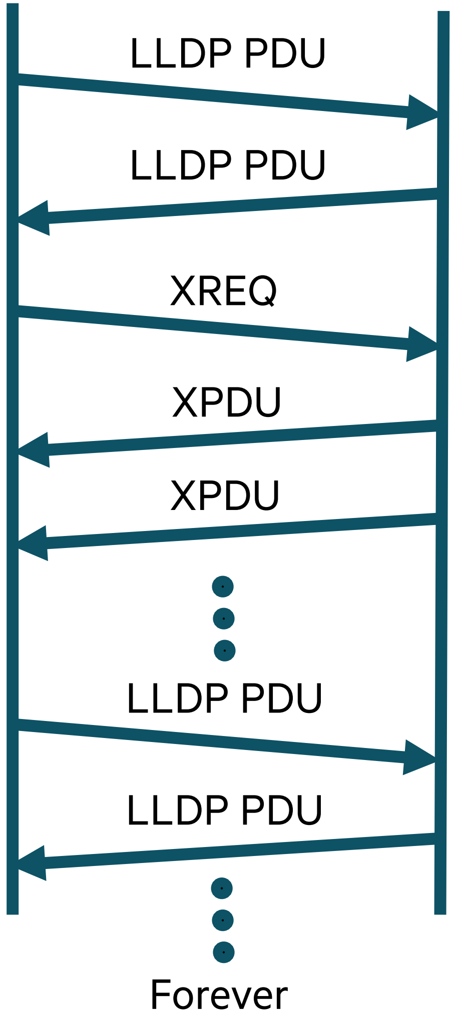
Manifest TLV Format

1	2	3	9	12	n*6+13	n*6+14	n*6+15 n*6+18
TLV type = 9 (7 bits)	TLV information String length (9 bits)	Return MAC Address (48 bits)	Total DB Size (24 bits)	Number of X-PDUs = k (8 bits)	X-PDU Number (8 bits)	X-PDU Revision (8 bits)	X-PDU Check Value (32 bits)

X-PDU Descriptor repeat n times (0<= n <= 82)

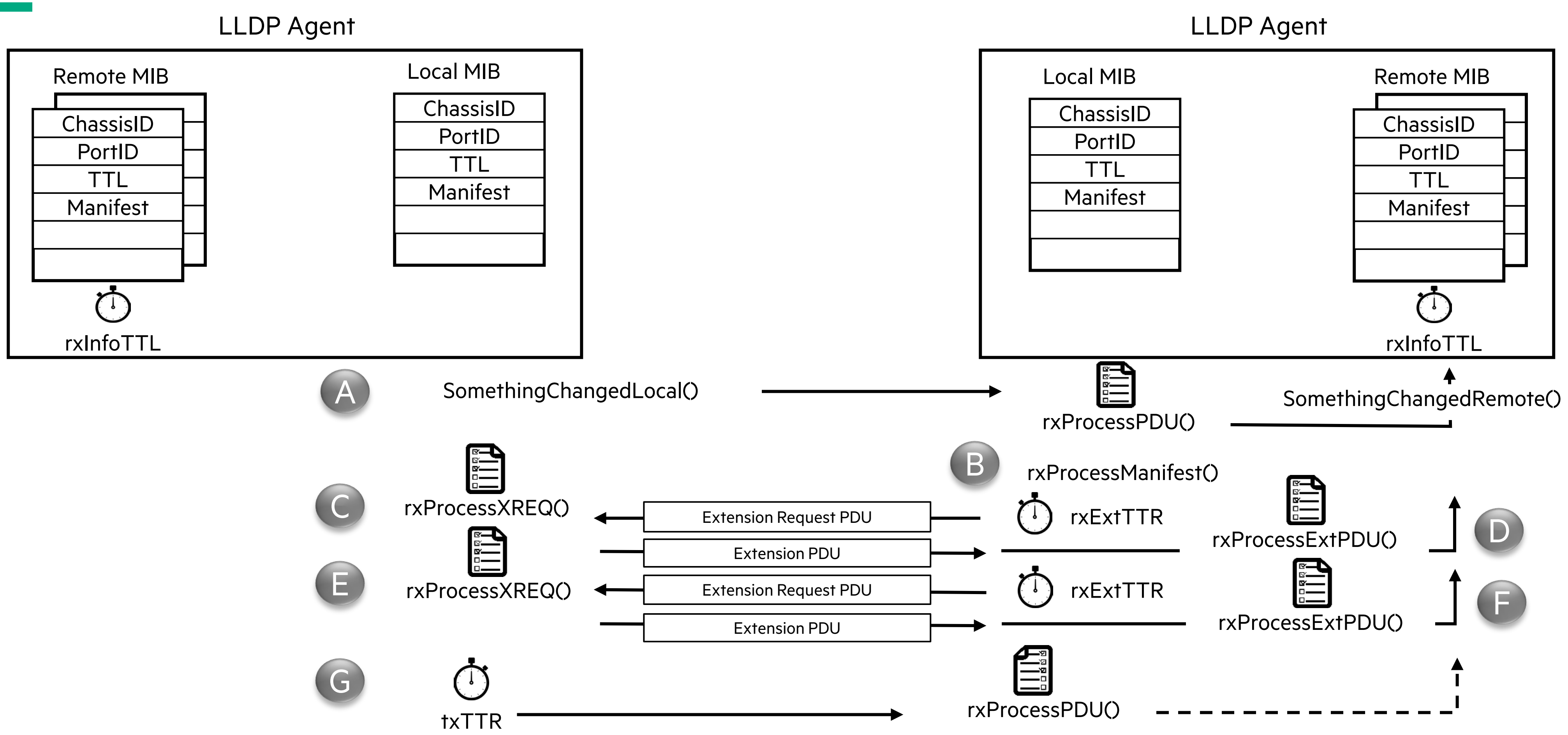
- Here the LLDPDU contains a Manifest TLV indicating an extended LLDP database
- The Manifest TLV specifies up to 83 extension XPDUs each containing LLDP TLVs
- Each of the extension XPDUs are requested using a unicast extension request XREQ
- Since the manifest contains a digest of each extension TLV the receiver can decide which if any extension TLV have changed. If none have changed then none need to be requested.
- Since the receiver controls the rate of XPDU transmissions it can manage congestion at input buffers.

Extended LLDP (multi-frame) ladder: (802.1ABdh-2021)



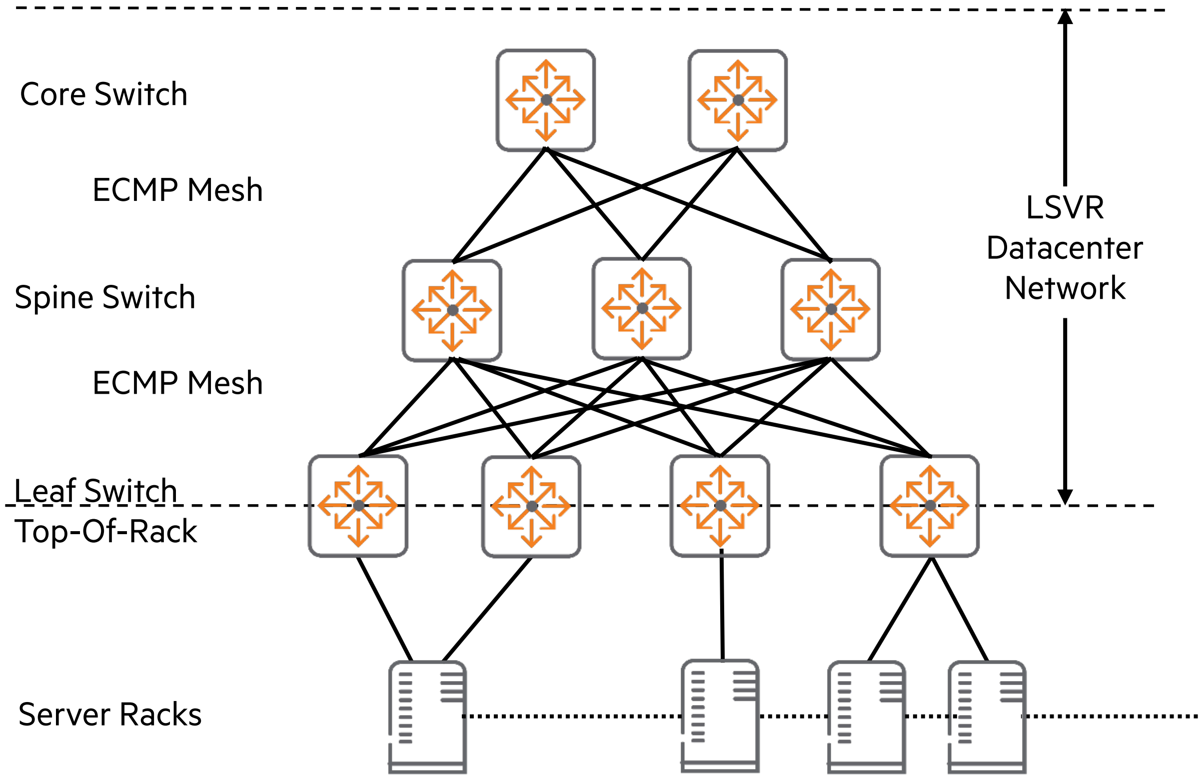
- Each Normal LLDP PDU is a single frame transmitted to the Scope MAC Address, exactly the same as classic LLDP
- The Normal LLDP PDU for extended LLDP is backward compatible with Classic LLDP (802.1AB-2016)
- LLDP is extended by adding a Manifest TLV to the Normal LLDP PDU describing all extension XPDUs
- Each XPDU contains LLDP TLVs which extend the size of the LLDP database
- The receiver of a Normal LLDPDU containing a Manifest TLV makes requests for those XPDUs that are out of date from the receiver's database.
- Requests for XPDUs and XPDU transmissions are unicast transmissions
- The receiver can control the rate of XPDU transmission based on it's processing rate and available resources.

Extended LLDP Operation (IEEE 802.1ABdh-2021)



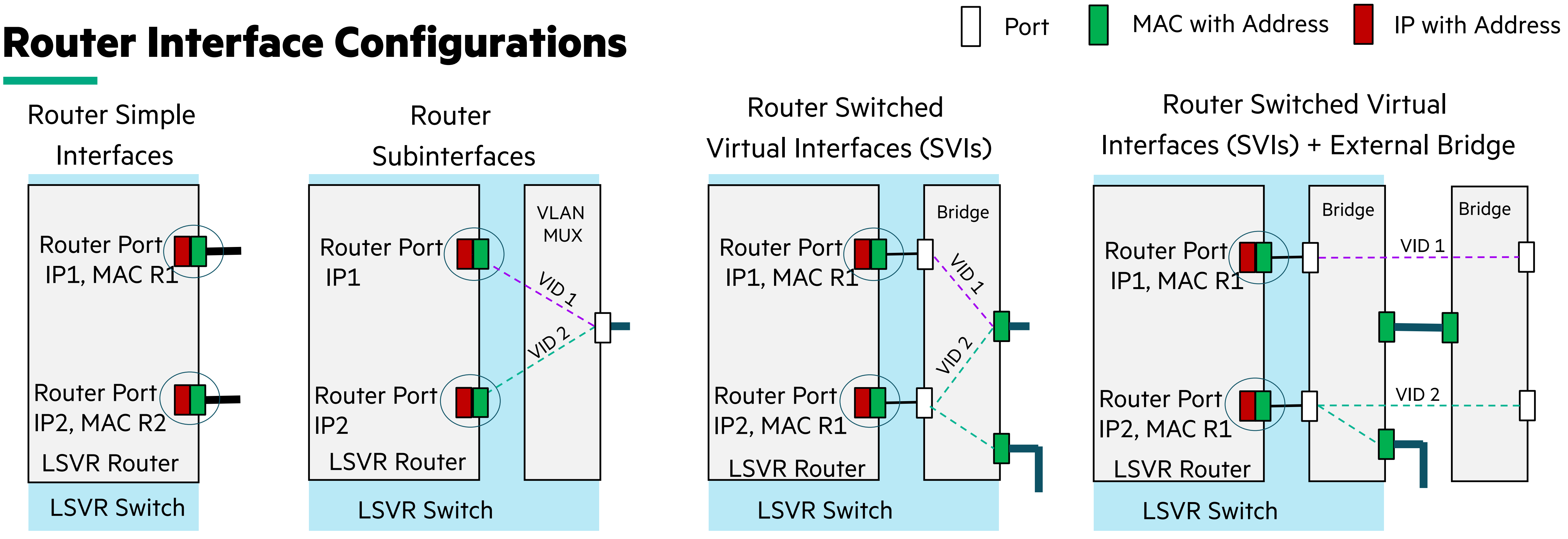
- Send LLDPDU as specified when periodically and whenever something changes
- Only send extension LLDPDU when explicitly requested by a XREQ
- Only issue XREQ when manifest shows the local copy is out of date

Datacenter Network Using LSVR



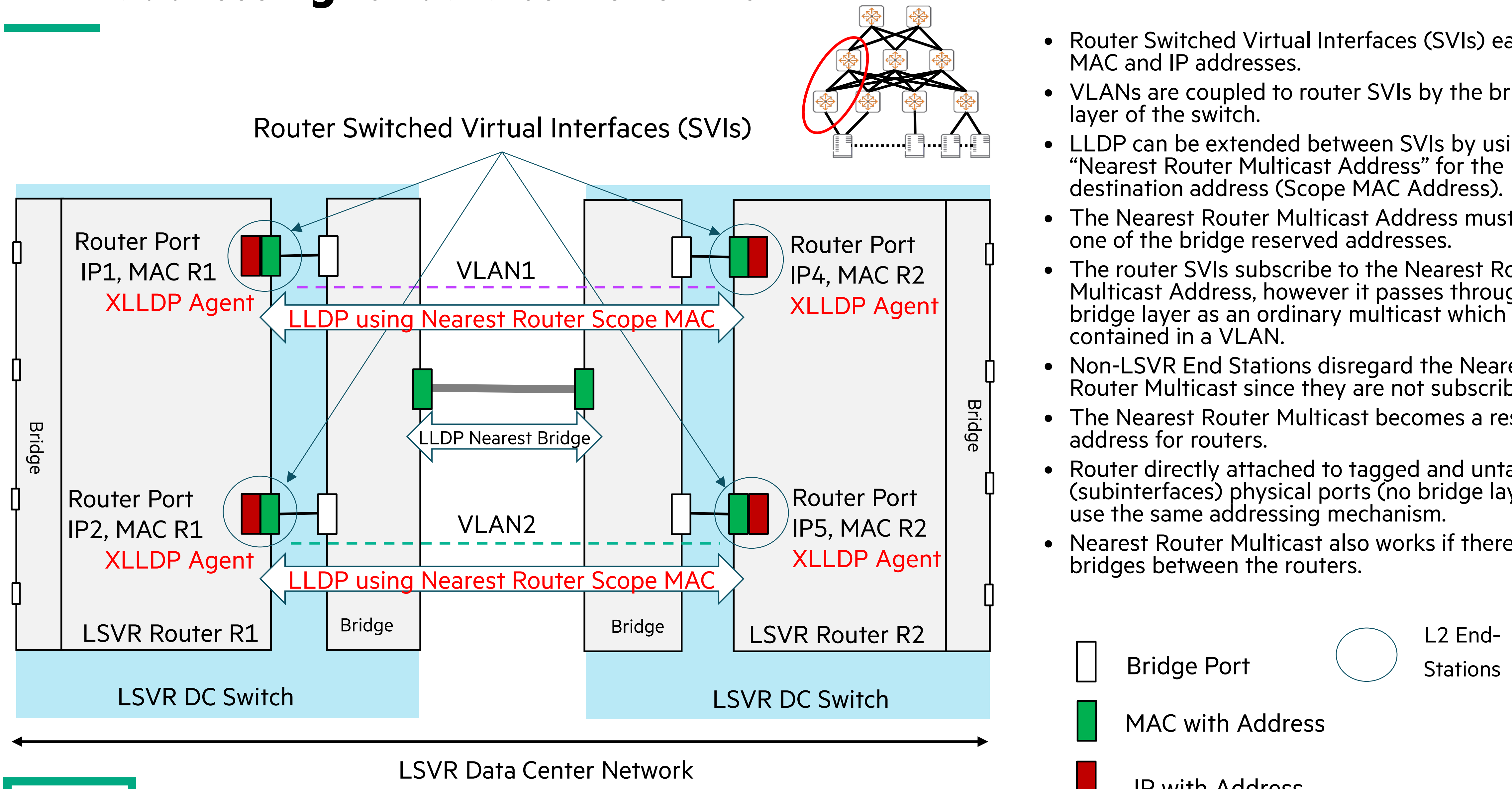
- Most datacenters are configured as 2-3 layer Clos networks using ECMP for distribution over the mesh and IEEE Std 802.1AX link aggregation for server attachment.
- Typically these networks provide an IPv4/IPv6 topology organized with ToR and Spine switches within Pods (around 8-128 racks)
- Servers at the network edge manage virtual and tenant networks which are encapsulated into the IP packets for transmission over the data center
- The orchestrator controls the creation of the virtual and tenant networks along with coupling to services

Router Interface Configurations



- Three distinct types of router interfaces are possible. These are:
  - Simple interface were the router has a direct attachment to a physical MAC port
  - Subinterfaces were multiple router interfaces share a MAC through a VLAN multiplexer
  - Switched Virtual Interfaces (SVIs) where multiple router interfaces have a virtual MAC connected to an internal bridge
- Any of the three router interface types can also be attached to other routers through an external bridge

LLDP addressing for data center switch



- Router Switched Virtual Interfaces (SVIs) each have MAC and IP addresses.
- VLANs are coupled to router SVIs by the bridging layer of the switch.
- LLDP can be extended between SVIs by using a “Nearest Router Multicast Address” for the LLDP destination address (Scope MAC Address).
- The Nearest Router Multicast Address must not be one of the bridge reserved addresses.
- The router SVIs subscribe to the Nearest Router Multicast Address, however it passes through the bridge layer as an ordinary multicast which is contained in a VLAN.
- Non-LSVR End Stations disregard the Nearest Router Multicast since they are not subscribed.
- The Nearest Router Multicast becomes a reserved address for routers.
- Router directly attached to tagged and untagged (subinterfaces) physical ports (no bridge layer) can use the same addressing mechanism.
- Nearest Router Multicast also works if there are bridges between the routers.

LLDP ChassisID + PortID = L3DL Logical Link Endpoint Identifier (LLEI)

Chassis ID TLV Format

TLV type = 1 (7 bits)	TLV information String length (9 bits)	Chassis ID subtype (8 bits)	Chassis ID (1-255 octets)

Port ID TLV Format

TLV type = 2 (7 bits)	TLV information String length (9 bits)	Port ID subtype (8 bits)	Port ID (1-255 octets)

- The Chassis ID and Port ID are the first two TLVs in all LLDP PDUs (frames). They identify the data base element this LLDP PDU is associated with.
- The ChassisID provides a unique system identifier which can be a MAC address, any IANA network address (such as an NSAP, IP, etc), a management identifier, or any locally defined string identifier.
- The PortID provides a unique identifier for the interface within the system which can be an IfIndex, a port number, or any locally defined string identifier.

- Together the ChassisID+PortID are the same function as the l3dl Logical Link Endpoint Identifier (LLEI)

LLDP uses periodic LLDPDU transmissions to assure liveliness

Time To Live TLV format

TLV type = 3 (7 bits)	TLV information String length=2 (9 bits)	Time to live (TTL) (2 octets)

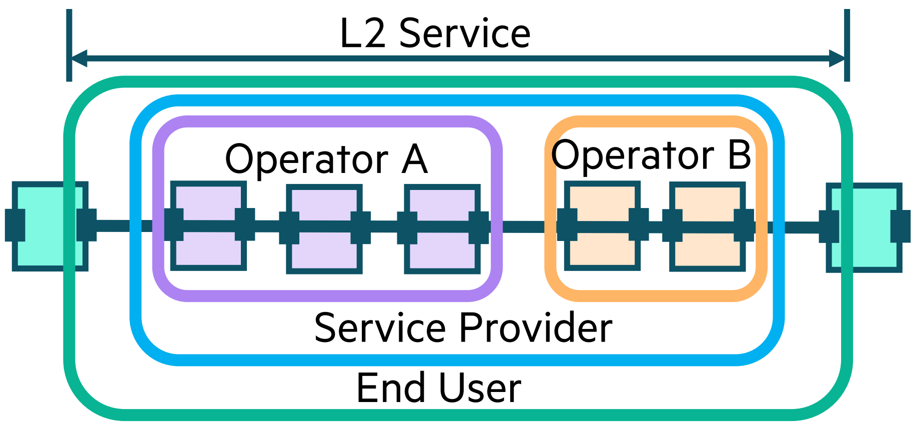
- Periodic transmissions can be set at 1-3600 seconds intervals.
  - Default is 30 seconds.
- The LLDP data base is invalid and discarded if a periodic transmission is not received within the time to live interval.
  - Default is 4 transmission times.
- A 0 time to live invalidates the database immediately.

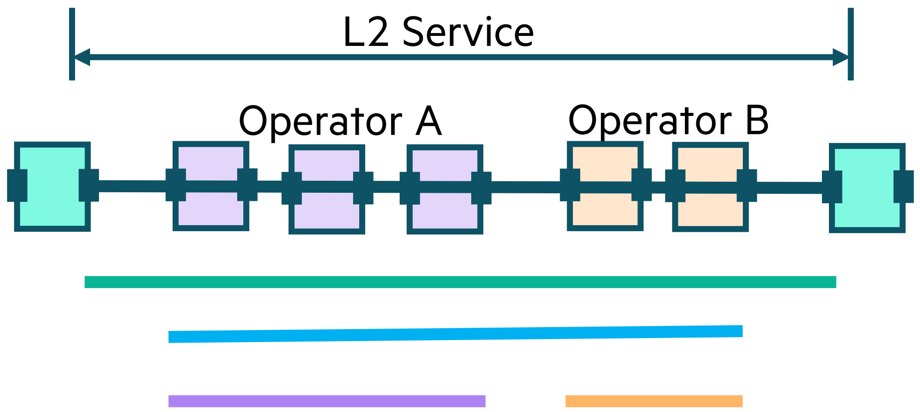
- To duplicate with l3dl operation the periodic transmissions can be set to 10 seconds (same as l3dl) and the TTL can be set to 3 transmission times.

Connectivity Fault Management (CFM)

- Currently standardized in IEEE Std 802.1Q-2022 Clauses 18-22
- Originally published as amendment IEEE Std 802.1Qag-2007
- Family of protocols that provides capabilities to detect, verify, isolate, and report end-to-end Ethernet connectivity faults
- Employs regular Ethernet frames that travel in-band with the customer traffic
- Devices that cannot interpret CFM Messages forward them as normal data frames
- CFM frames are distinguished by Ether-Type (0x8902) and DA address (for multicast messages)
- Features
  - Nested Maintenance Domains (MDs) that break up the responsibilities for network administration of a given end-to-end service
  - Maintenance Associations (MAs) that monitor service instances under a given MD
  - Maintenance Points (MPs) that generate and respond to CFM Protocol Data Units (PDUs)
  - Protocols (Continuity Check, Loopback, and Linktrace) used for Fault Management activities

Connectivity Fault Management (CFM) Basic Concepts (802.1Q cl 18-21)





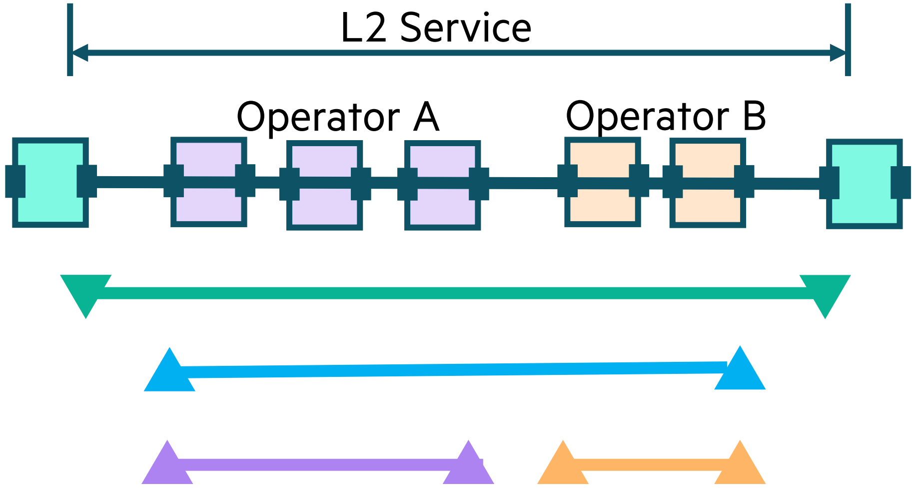
Maintenance Domains (MDs)

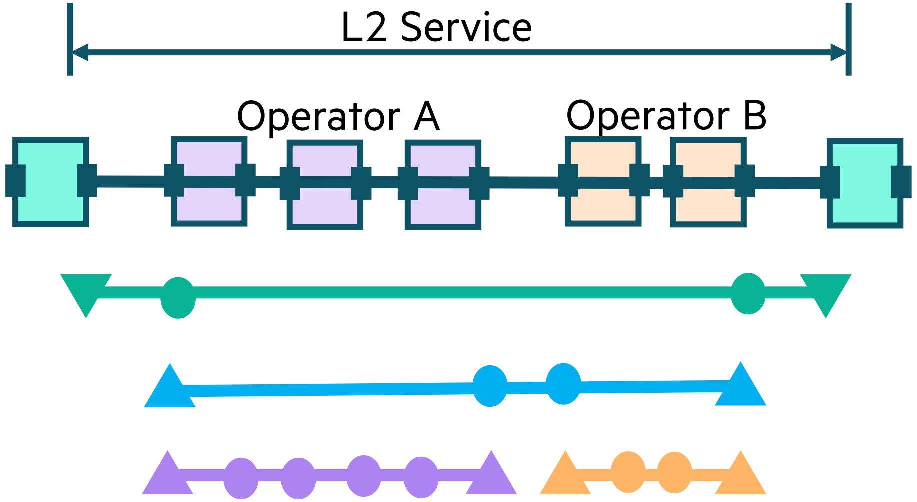
- Defined by Operational/Contractual Boundaries, such as Customer/Service Provider/Operator
- MD may nest and touch, but never intersect
- Up to eight levels of "nesting": MD Level (0..7) - the higher the level, the broader it's reach
- MD Name Format: null, MAC address, DNS or string-based

Maintenance Associations (MAs)

- Monitors connectivity of a particular service instance in a given MD, such as one service that traverses four MDs = four MAs
- Defined by a set of Maintenance End Points (MEPs) at the edge of a domain
- Identified by MAID - "Short MA" Name + MD Name
- Short MA Name Format - Vlan-ID, VPN-ID, integer or string-based

Connectivity Fault Management (CFM) Basic Concepts





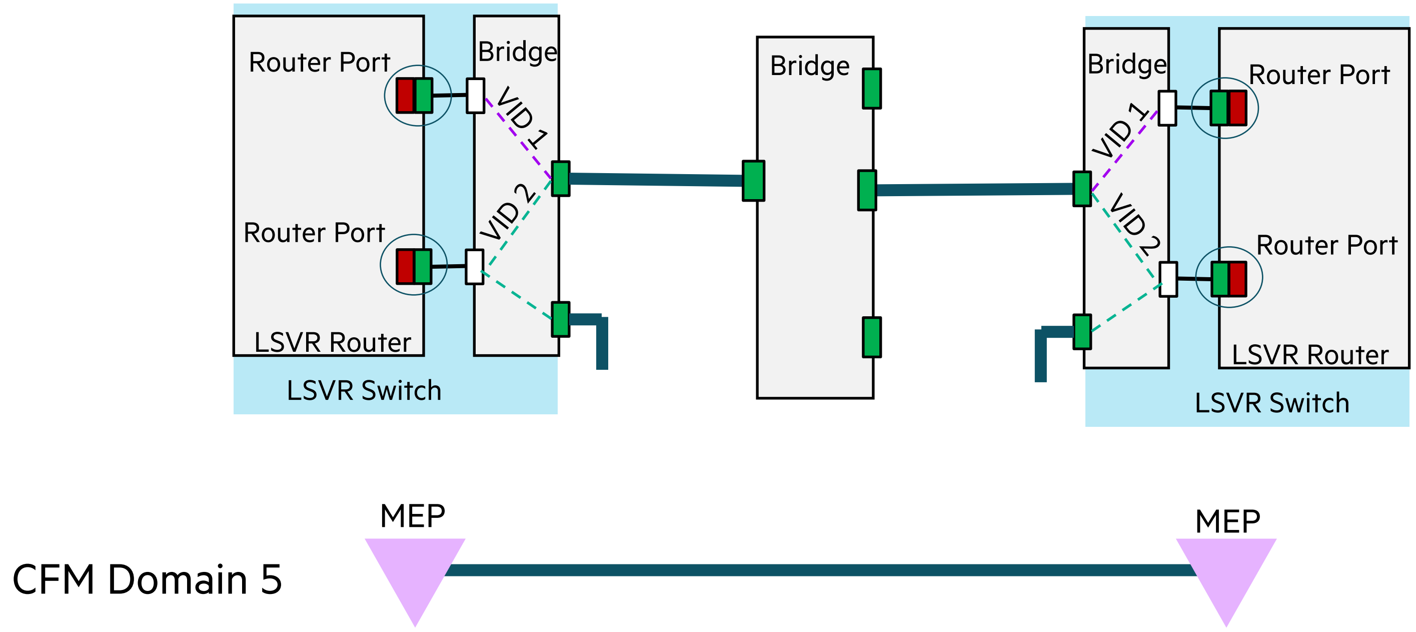
Maintenance End Points (MEPs)

- Maintenance Association End Point
- Define the boundaries of an MD
- Support the detection of connectivity failures between any pair of MEPs in an MA
- Associated per MA and identified by a MEPID (1-8191)
- Can initiate and respond to CFM PDUs

Maintenance Intermediate Points (MIPs)

- Maintenance Domain Intermediate Point (MIP)
- Supports the discovery of paths among MEPs and location of faults along those paths
- Can be associated per MD and VLAN/EVC (manually or automatically created)
- Can add, check, and respond to received CFM PDUs

Connectivity Fault Management (CFM, IEEE 802.1Q clauses 18-21)



- CFM provides OAM for paths extending through L2 networks.
- Maintenance Domains (MDs) provide a hierarchical management space with 8 possible levels (0-7)
- For the LSVR application MD level 5 seems most appropriate.
- For data center LSVR link discovery it is possible for router interfaces to be connected to a provider L2 service and therefore the MD level should be in the customer domain.
- It is also possible data center LSVR routers provide L2 services over the data center such as EVPNs. In this case the LSVR discover should be a service below network extended by an EVPN.
- Within a MD Maintenance Associations (MAs) are established between Maintenance association End Points (MEPs)
- MD internal points are identified by Maintenance domain Intermediate Points (MIPs)

LSVR Encapsulation Types and Addressing TLVs

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=127      |    Length           |          OUI = 00-00-5E       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |  Subtype = xx | LsvrType = e |  Encapsulation ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~   List …                                                        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Each encapsulation includes flags indicating primary, underlay/overlay, loopback
- The encapsulation subtypes are:
  - LsvrType = 1: IPv4 encapsulation list each with flags, IPv4 address, prefix length
  - LsvrType = 2: IPv6 encapsulation list each with flags, IPv6 address, prefix length
  - LsvrType = 3: MPLS IPv4 encapsulation list each with flags, label list, IPv4 address + prefix length list
  - LsvrType = 4: MPLS IPv6 encapsulation list each with flags, label list, IPv6 address + prefix length list

Upper-Layer Protocol Configuration TLV

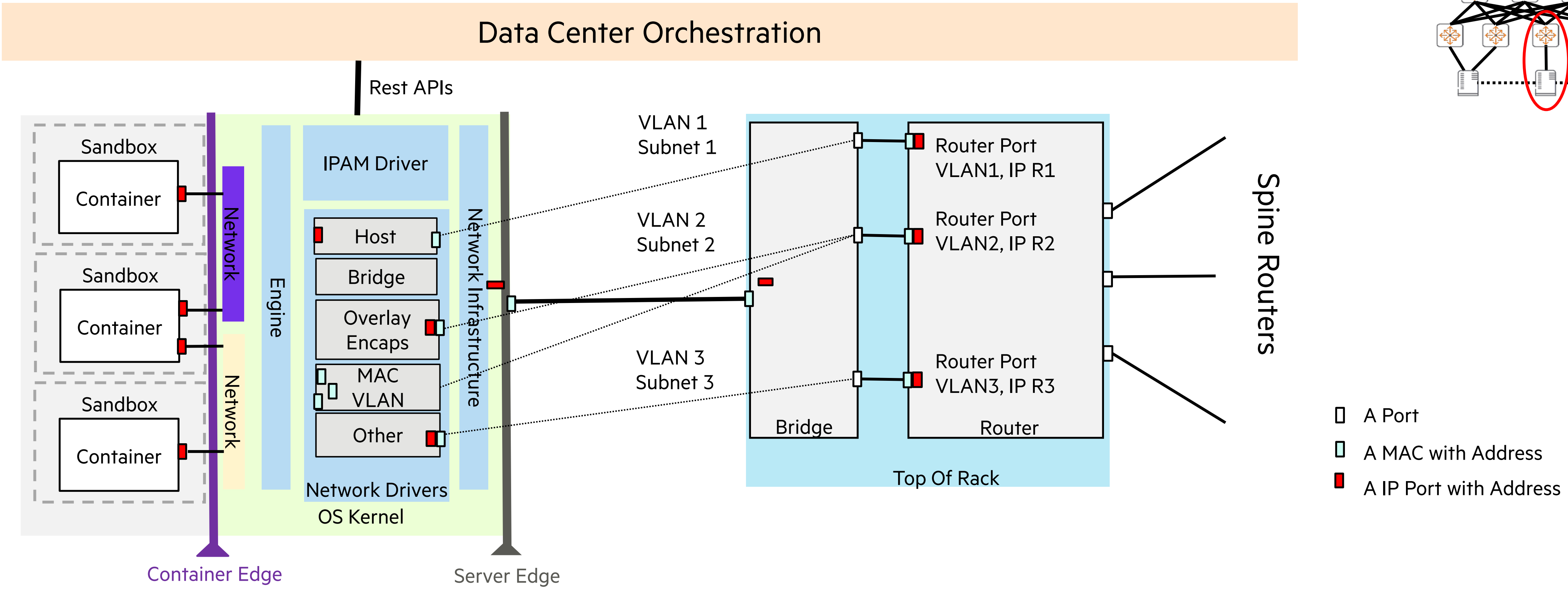
```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=127     |    Length          |        OUI = 00-00-5E      ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~              |  Subtype = xx |  LsvrType = a |  AttrType = a |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Attribute                                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| AttrType = a |            Attribute …                         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

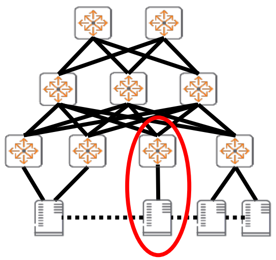
- Upper layer protocol configuration information LsvrType = 5 and Attribute types:
  - AttrType = 1: BGP ASN
  - AttrType = 2 : BGP IPv4 Peering Address
  - AttrType = 3 : BGP IPv6 Peering Address
  - AttrType = 4 : BGP Authentication Data
  - AttrType = 5: Misc Flags, Bit 0: GTSM, Bit 1: BFD, Bit 2-15 reserved (must be 0)

Chassis ID:

subtype	Basis	reference
0	Reserved	-
1	Chassis component	EntPhysicalAlias (RFC 6933)
2	Interface alias	IfAlias (RFC 2863)
3	Port component	EntPhysicalAlias (RFC 6933)
4	MAC address	MAC address (IEEE Std 802)
5	Network address	IP address string, first octet IANA family
6	Interface name	ifName (RFC 2863)
7	Locally assigned	Alpha-numeric string, local assigned
8-255	Reserved	-

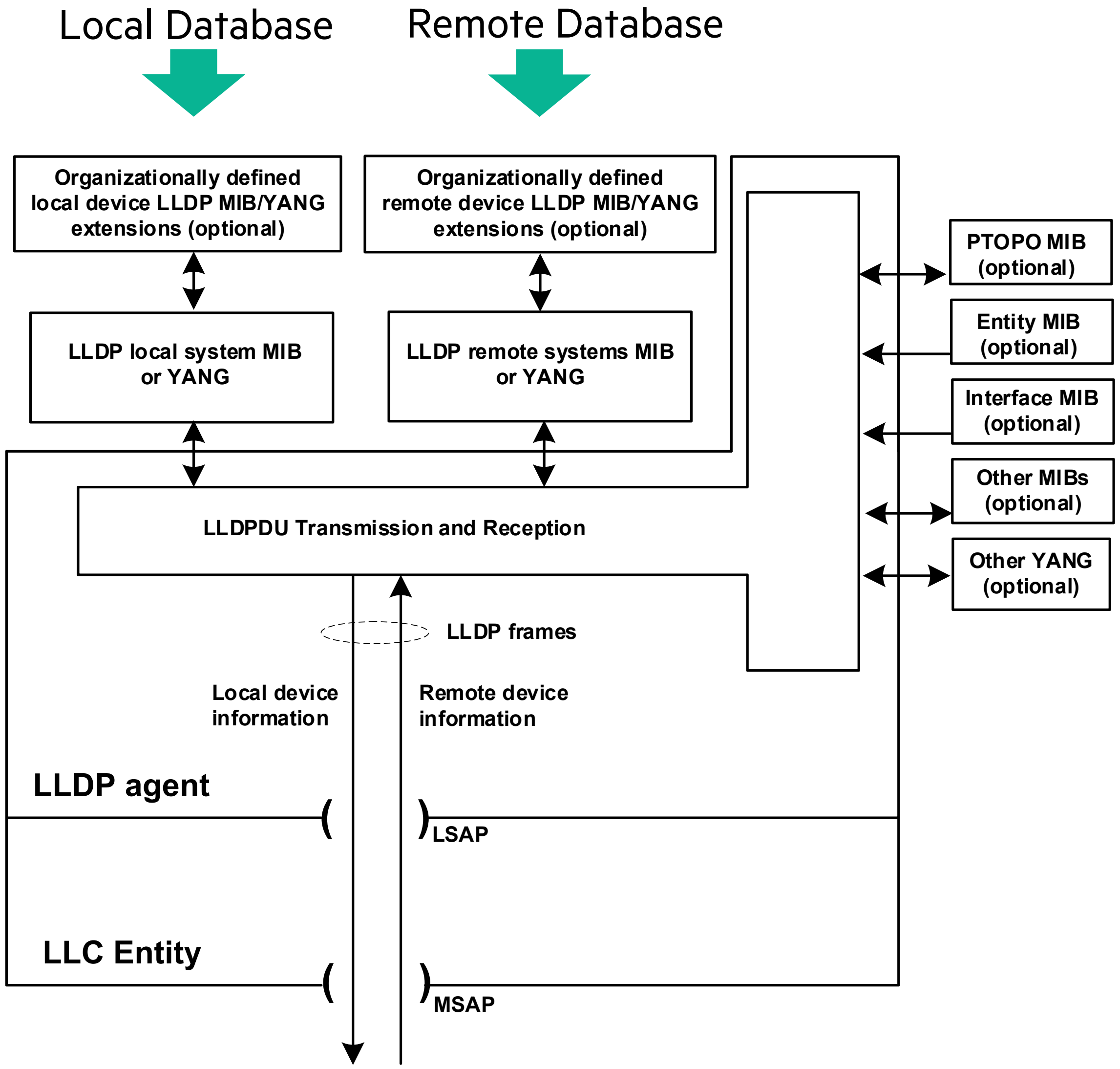
Server Network Interfaces – Containers (i.e. Docker)





- Container Solutions use Linux Namespaces and Groups to isolate containers
- These solutions provide a variety of network connections, overlays for large scale datacenters
- DC network is a simple IP network. Scaling L3 encapsulations with Network Virtualization Edge like procedures

LLDP is an extremely simple database exchange protocol



- LLDP starts with a local database which it advertises in a single frame LLDPDU using a Scope MAC Address as the L2 destination.
- The Scope MAC Address determines the extent of the LLDPDU propagation.
- LLDP advertisements are received and stored at each LLDP port listening to the corresponding Scope MAC Address.
- The LLDP receivers build a remote database containing the advertised information from each LLDP transmitter.
- LLDP transmitters are identified by a Chassis ID and Port ID encoded in each LLDPDU
- LLDP transmitters advertise their local database periodically to maintain liveliness