802.1EH IEEE 802 Modeling

IEEE 802.1 Meeting

Paul Bottorff (HPE)

September 14, 2026

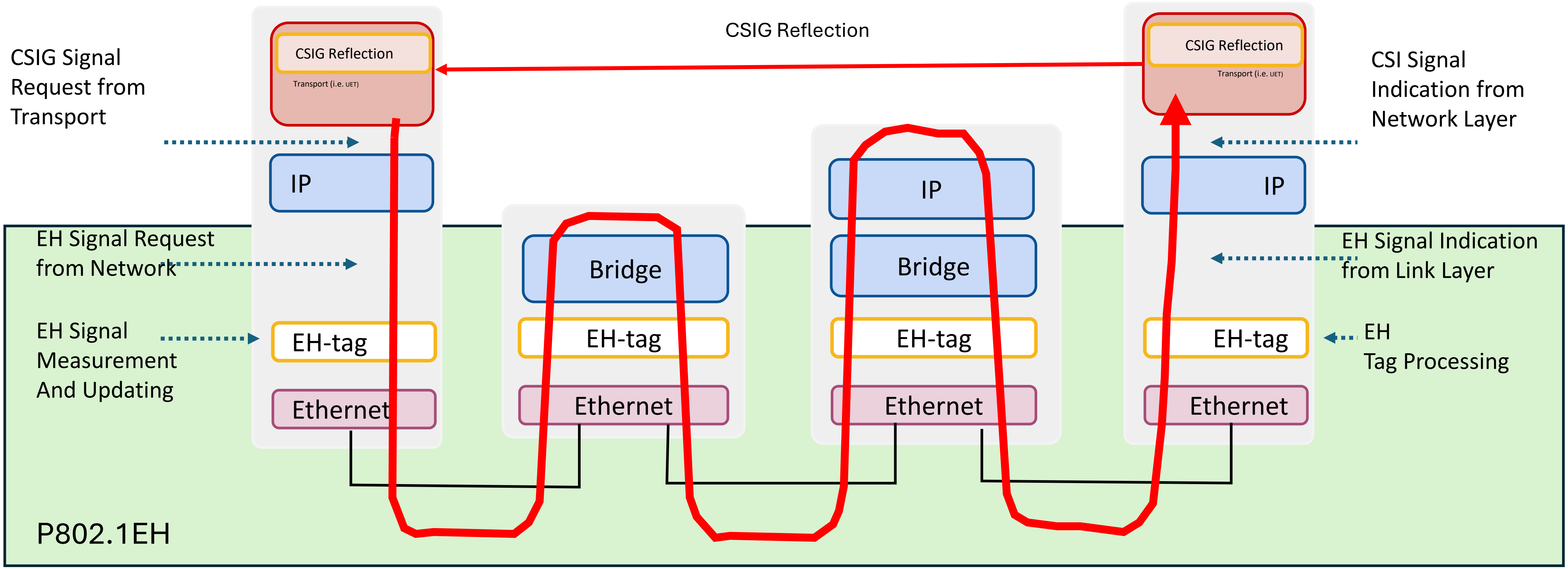
Agenda

- CSIG basic concepts (brief introduction)
- IEEE Modeling for Telemetry Tagging of Data Frames
  - Service Interface Extensions for the MAC Service Interface
  - Modeling for End Stations
  - Modeling for End Systems
  - Modeling for MAC and VLAN Bridges
  - Modeling for Provider Bridges
- Thoughts 802.1EH draft clause organization

CSIG Basic Concepts

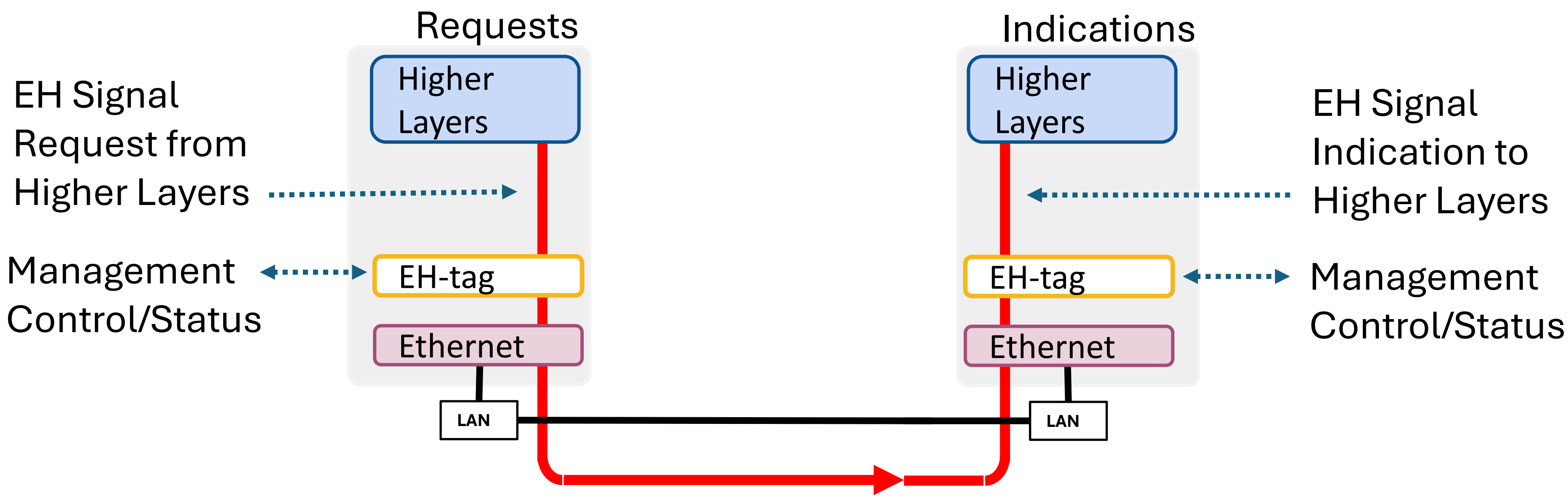
- Telemetry is transmitted in-band in fixed size, 4 or 8 byte, L2 tags
- Each tag specifies a single ***signal*** parameter indicating the type of measurement to be made.
  - Only a single measurement type is made per packet
- Four signals are currently defined which are:
  - ABW	Min available bandwidth per port
  - ABW/C	Min relative available bandwith per port
  - Delay	Max per-hop delay
  - nQD	Max queue depth normalized by port speed
- Each hop along the path compares the current ***signal value*** parameter carried in the CSIG tag with the value calculated for this hop
  - If this hop is more extreme (depending on min or max signal type) the ***signal value*** along with the ***locator*** carried in the tag is updated with the information from this hop.

CSIG is a multi-layer telemetry protocol



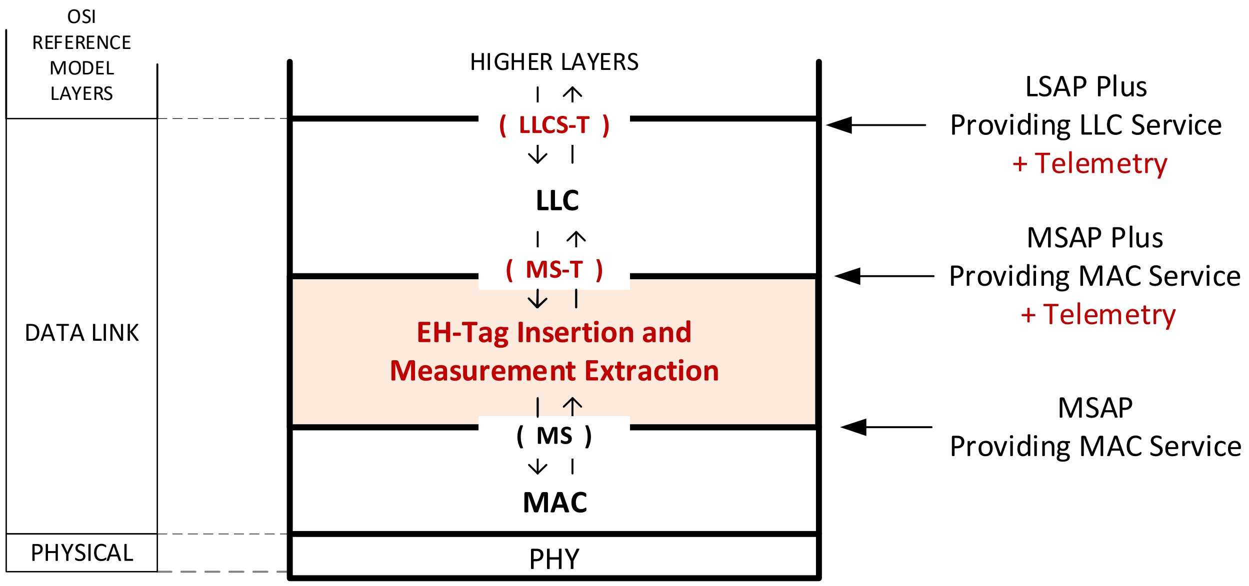
- The P802.1EH project covers the link layer operations used by CSIG and operations for telemetry tagging where no higher layer protocols are involved.
  - Higher layer protocols including reflection and congestion control are not specified by other standards bodies

End Station Tag Processing



- Higher layers making use of 802.1EH signals make measurement requests in-band via their service interface to the lower layers. These requests are encoded at L2 within a EH-tag.
- Requests could also be initiated directly from L2 under control from the management module rather than the service interface.
- On receipt of the EH-tag at an L2 destination measurements are extracted from the tag and delivered with the service indication to the Higher Layers.
- Measurements could also be collected at L2 for delivery to the management interface or a collector application.

Simple Open System Interconnect (OSI) End Station Stack



- Here the SAPs of an OSI stack have been enhanced to propagate the telemetry requests and indications from the Higher Layers to/from the Data Link.
- At the Link layer the Telemetry entity sits between the LLC and MAC entities.
- Frames are transmitted from and received by the telemetry entity using the MAC Service, however the service provided by the telemetry entity is an enhanced MAC Service including the telemetry signals.

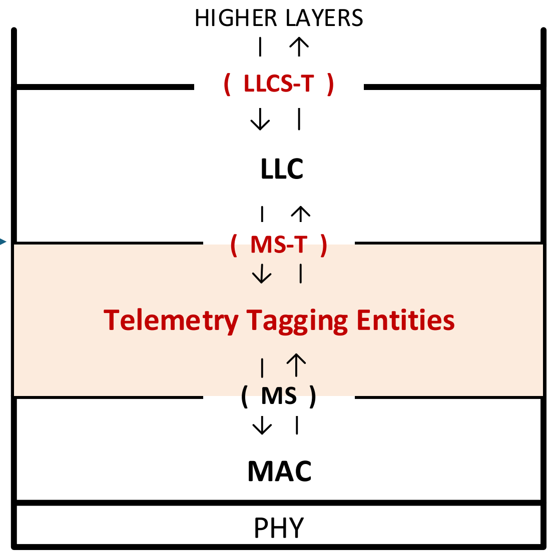
Note: The telemetry entity is not a shim since the MS-T service interface is extended to support the telemetry signals.

MAC Service Telemetry Extension (MS-T) For End Stations

MAT_UNITDATA.indication (
destination_address,
source_address,
mac_service_data_unit,
priority,
signal_type,
signal_value,
location_metadata,
flags
)

MAT_UNITDATA.request (
destination_address,
source_address,
mac_service_data_unit,
priority,
signal_type,
signal_value,
location_metadata,
flags
)

MSAP Plus

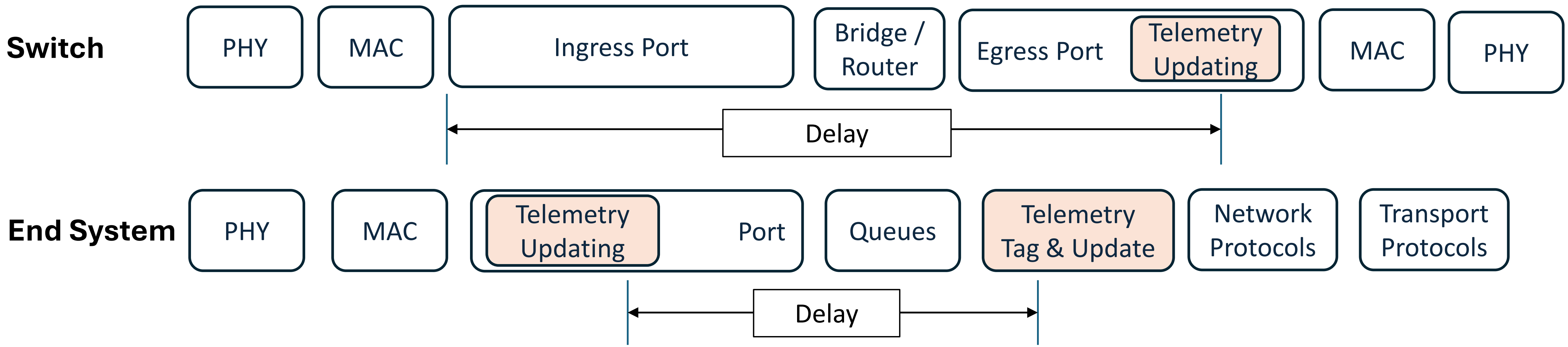


- Similar telemetry extensions are needed at LLCS

Timestamp needed for hop delay calculation

- We need a way to pass a timestamp along with each packet over the MS as well as over the ISS and EISS.
- The hop delay is measured in nanoseconds, however the precision of the measurement needs specification.
  - My strawman for hop delay time precision is 1 max sized packet time. For 100 Gbit Ethernet this would be ~121 nsec.
- The UEC specification is silent on end system hop delay, however ideally these would be included in the max hop delay signal.
- For router implementations the router ports, which are end stations, need to have a timestamp along with telemetry for each packet.
  - The timestamp needs to be forwarded with the packet over the router relay to the egress router port where it will be used to calculate the hop delay for the router.

Timestamp for hop delay calculation



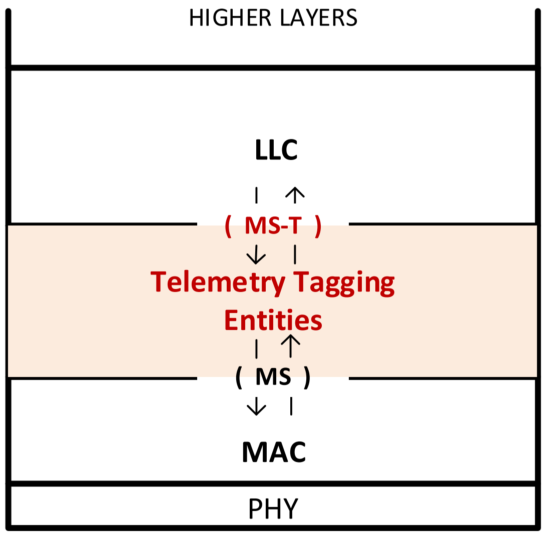
- For transit nodes a timestamp of the ingress time is needed
  - UEC liaison shows the timestamp within the ingress port before queuing
- For end systems both receive and transmit timestamps are needed
  - UEC has not provided any recommendations on this
  - Receive timestamp can follow the switch recommendation
  - Transmit timestamp needs to be before the packet enters the transmit queues

MAC Service Telemetry Extension (MS-T) with timestamp

MAT_UNITDATA.indication (
destination_address,
source_address,
mac_service_data_unit,
priority,
signal_type,
signal_value,
location_metadata,
flags,
time_stamp
)

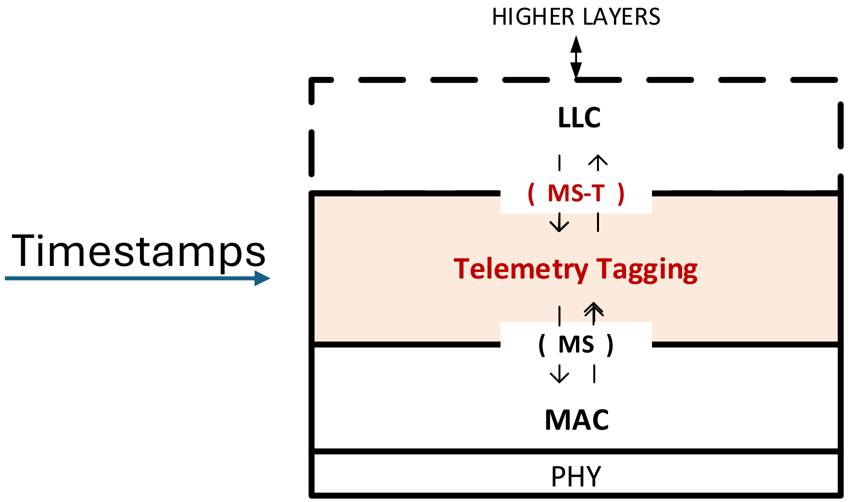
MAT_UNITDATA.request (
destination_address,
source_address,
mac_service_data_unit,
priority,
signal_type,
signal_value,
location_metadata,
flags,
time_stamp
)

MSAP Plus



- A timestamp could be added to the extended MAC service interface (MS-T).
- A timestamp will also be needed over the ISS and EISS.
    - Could be loaded as a sub-parameter of the ISS and EISS connection_identifier

Simple End Station

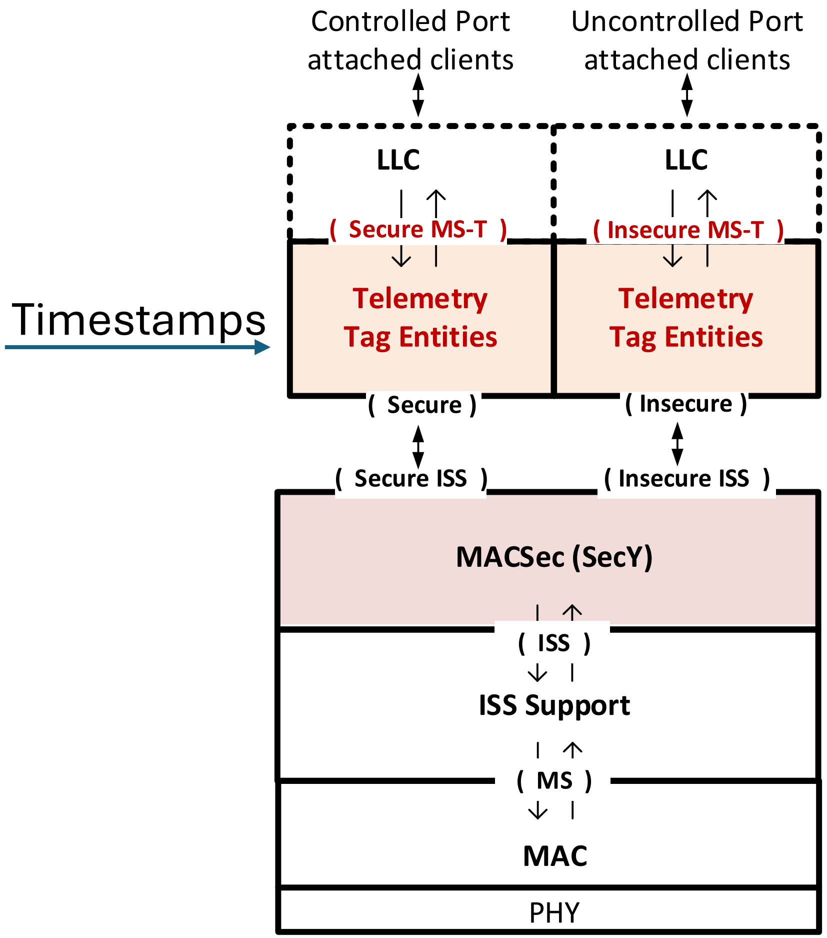


- In a simple end station the telemetry tag is inserted and telemetry extracted by a layer above the MAC Service interface.
- The telemetry layer uses the telemetry in the extended MAC Service (MS-T) requests to generate telemetry tags.
- The telemetry layer strips inbound telemetry tags and passes their telemetry data in the extended MAC Service (MS-T) indications.

DA	SA	TelTAG	Data	FCS

Frame on wire with simple End Station

End Station with MACSec

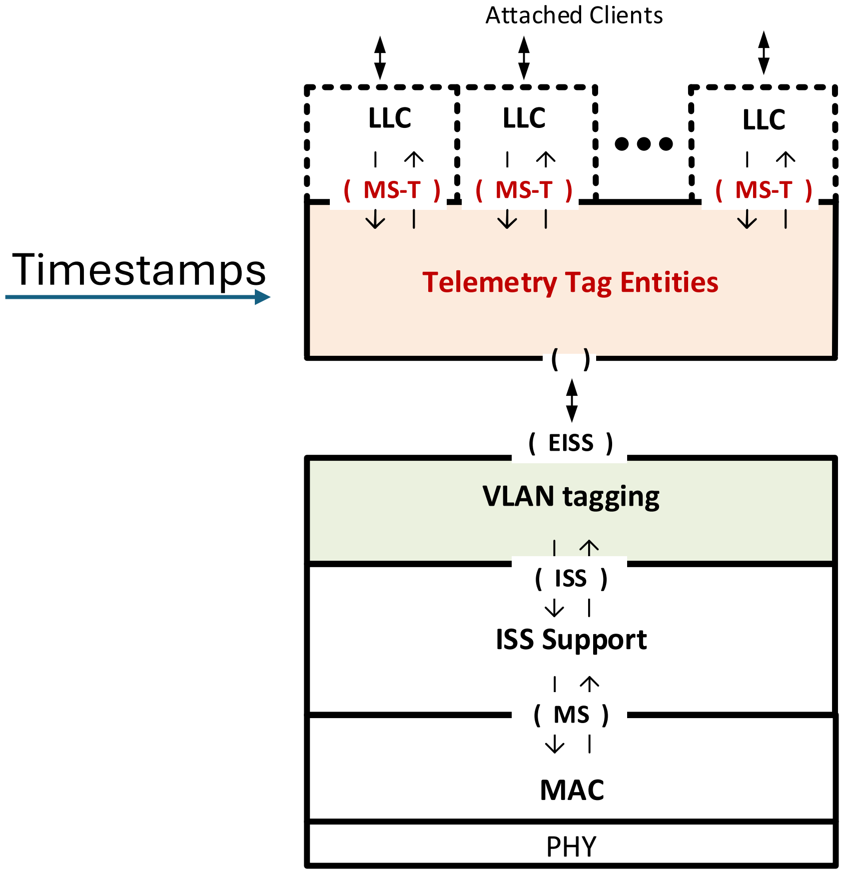


- The telemetry tag is inserted, updated, and extracted above MACSec since the telemetry operations need to be done in the clear.
- There may be other layers between MACSec and telemetry tagging.
- The telemetry tag entities interface to LLC using the extended Mac Service interface (MS-T) to pass the telemetry tag information.
- Here the timestamp is taken above MACSec so MACSec operations are not included in the hop delay.

DA	SA	SecTAG	TelTAG	Data	ICV	FCS

Frame on wire with MACSec End Station

End Station with VLAN tagging

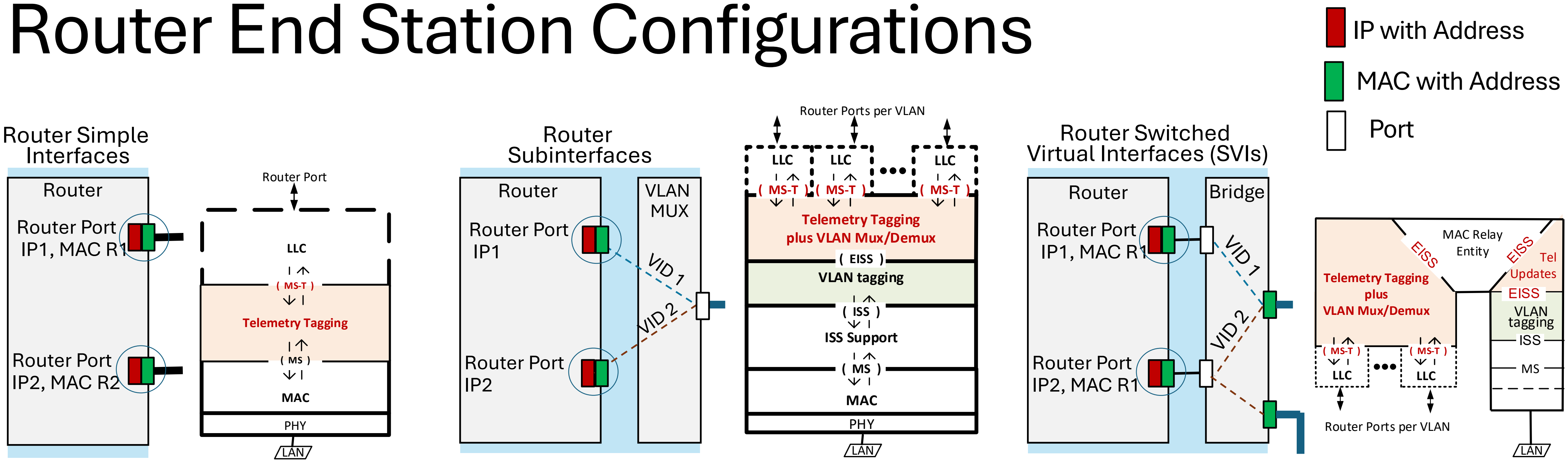


- The telemetry tag is inserted, updated, and extracted above VLAN tagging to allow it to pass transparently through downstream Bridge relays.
- There may be other layers between VLAN tagging and the telemetry tagging.
- The telemetry tag entities interface to LLC using the extended Mac Service interface (MS-T) so the telemetry tag information is passed up the stack.
- A reference point for the timestamping is needed.
  - Ideally this would be at the top of the port stack just before queuing to eliminating the port pipeline from the hop delay.

DA	SA	VLAN	TelTAG	Data	FCS

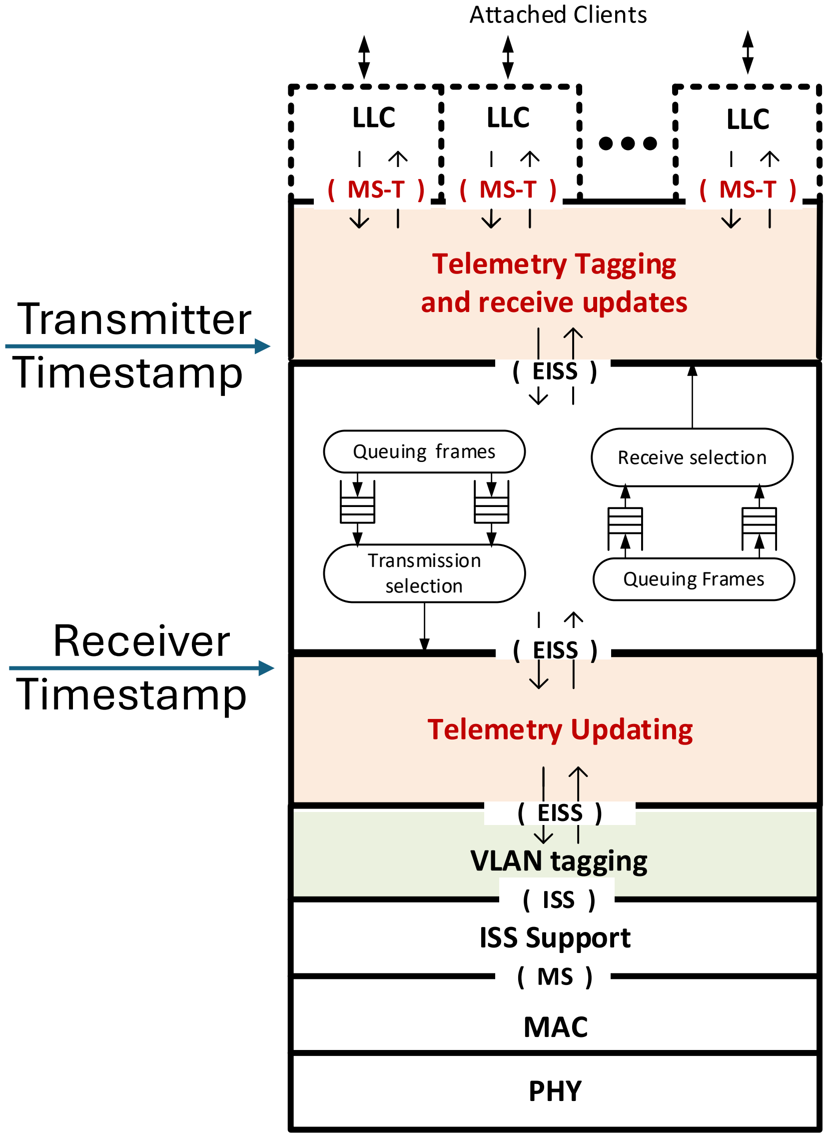
Frame on wire with VLAN aware End Station

Router End Station 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
- In each router interface before ingress to the router port the telemetry tag is stripped and the telemetry is forwarded to the router port as meta-data.
- In each router interface after egress from the router port a telemetry tag is inserted based on the telemetry meta-data forwarded from the router

End System stack detail

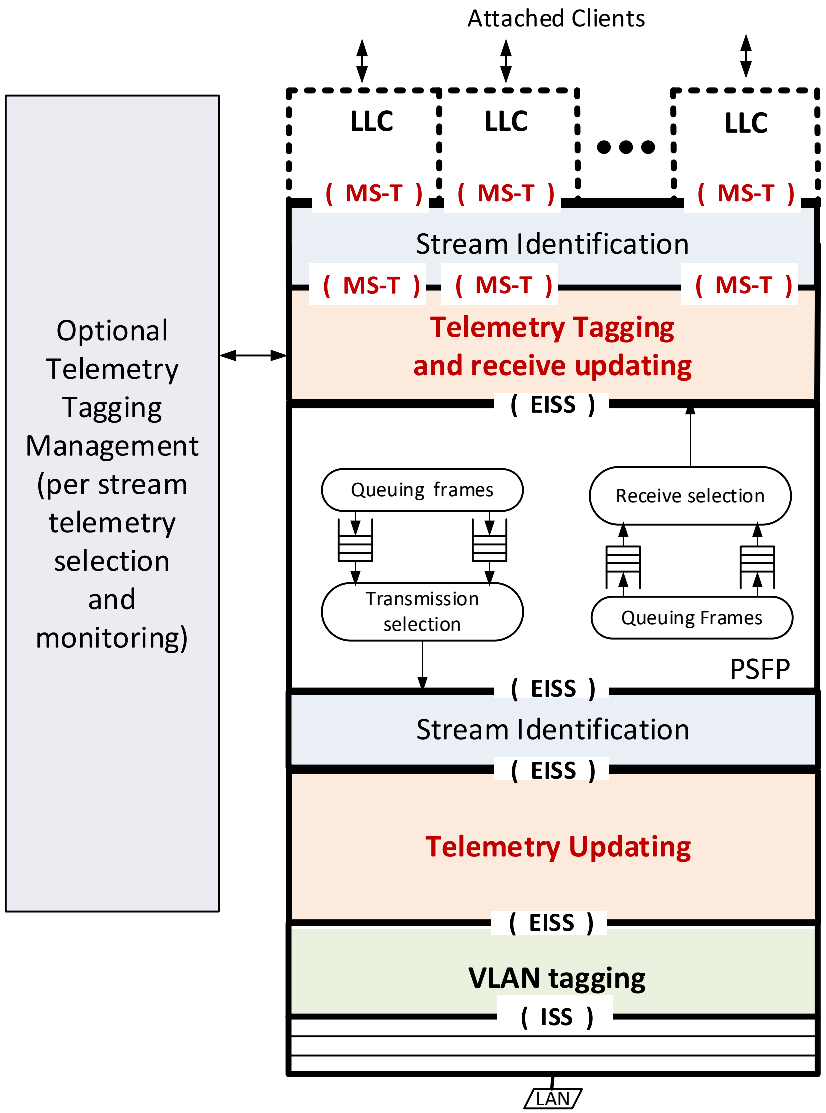


- End Systems could insert the telemetry tag before transmit queuing and extract telemetry after receiver dequeuing.
- Since the receive hop delay can't be calculated until dequeuing from the receive queues the calculation of hop delay must be after receive dequeue.
- Since the transmit hop delay can't be calculated until dequeuing from the transmit queues the calculation of hop delay must be after transmit dequeue.

DA	SA	VLAN	TelTAG	Data	FCS

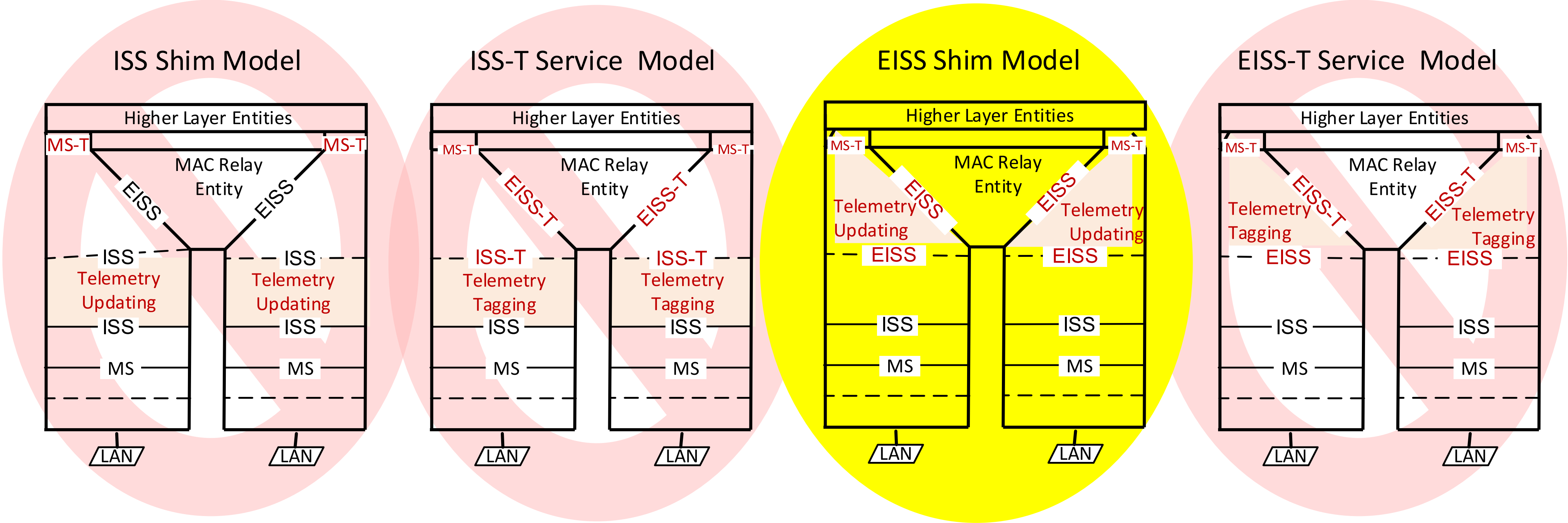
Frame on wire with VLAN aware End Station

Optional End System Telemetry Initiation/Collection



- Current CSIG telemetry requests are initiated at the transport layer.
  - The extended MAC Service interface (MS-T) allows higher layers to control telemetry generation.
- Alternatively, telemetry requests could be initiated at the link layer under management control.
  - A manageable option could be used to enable higher layer telemetry or link layer telemetry.
  - Management could specify a set of triples composed of Signal, Sample Period,and Stream Identifier indicating the telemetry requested.
  - For telemetry collection the link layer could capture all telemetry data along with the stream handle and local time.

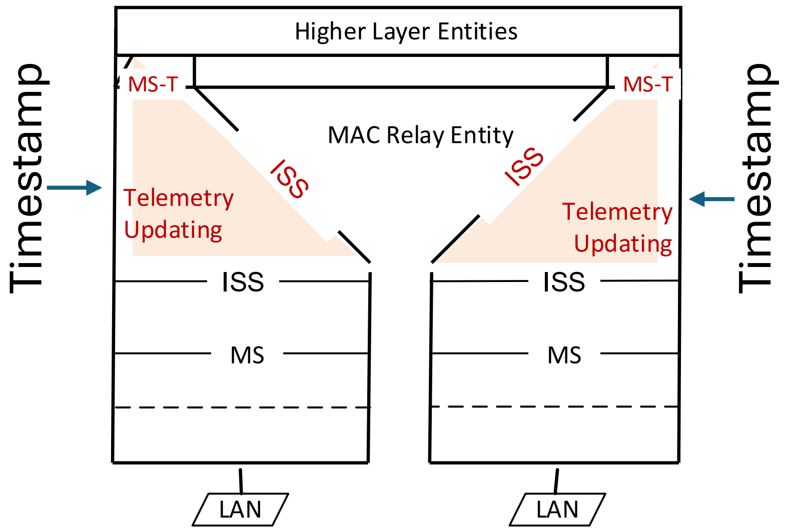
Alternative VLAN Bridge Models



- The two Service Interface models (ISS-T and EISS-T) force major changes to bridge implementation since they imply reducing the telemetry to meta-data which needs to be carried over the bridge relay with each frame.
- All current CSIG implementations use the shim model which allows the telemetry to pass over the bridge relay within the CSIG tag.
  - Passing the CSIG tag over the bridge relay has allowed retrofitting CSIG support in many existing switches
- Since the telemetry tag is located under the VLAN tag the ISS shim model implies parsing past the VLAN.
- In the EISS shim model the VLAN tag has been removed allowing direct access to the telemetry tag.
- In a MAC Bridge there will be no VLAN tag so a telemetry tag updating needs to be an ISS shim.

Telemetry Updating in MAC Bridge

MAC Bridge
ISS Telemetry Shim

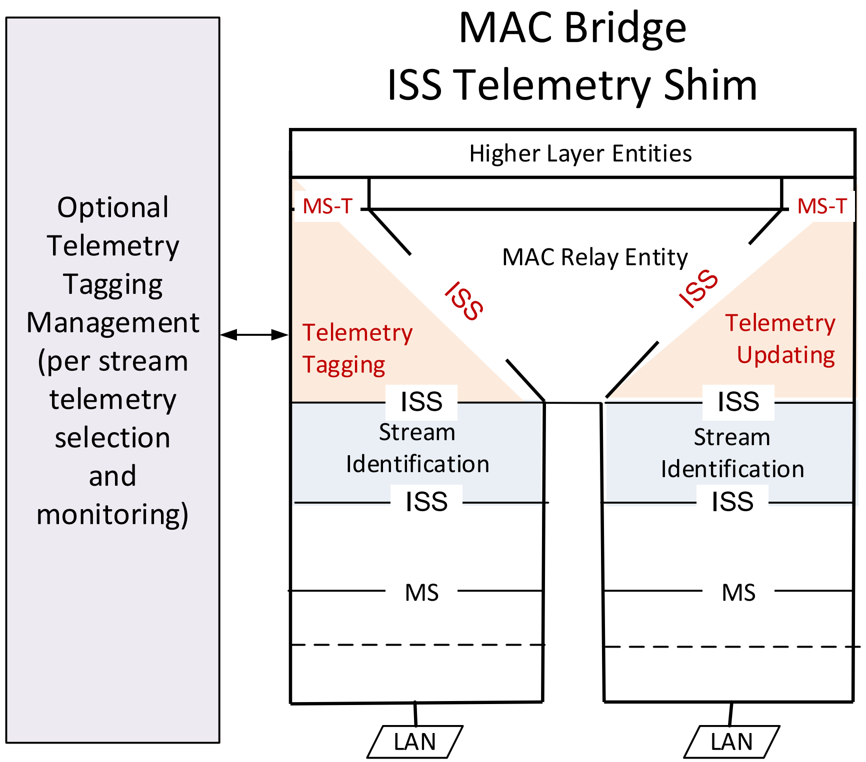


- A MAC Bridge will only update the telemetry tag if it is the top tag in the frame
- On ingress the telemetry shim time stamps the frame before queuing it
  - The timestamp is carried as a parameter of the ISS connection_identifier
- At egress the telemetry shim:
  - Calculates the signal value for the specified signal type
  - Compares the calculated signal value to the signal value carried in the Telemetry TAG
  - If the compare is for max and the calculated signal value is greater than the tag's signal value the tag is updated
  - If the compare is min and the calculated signal value is less than the tag's signal value the tag is updated

DA	SA	TelTAG	Data	FCS

Frame on wire for MAC Bridge telemetry update

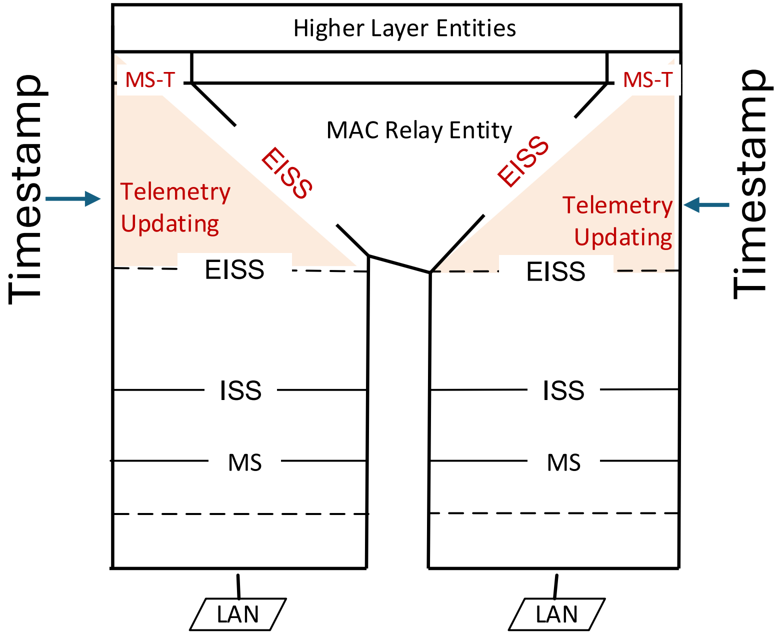
Optional L2 Telemetry Initiation and Collection



- Telemetry collection and initiation could be initiated from a Bridge Port
  - Could be used within provider or enterprise networks.
- Collection and initiation of telemetry at the link would be controlled by a management interface.
  - Needs some configurable options to enable end station telemetry or in network telemetry.
  - For initiation, management could specify a set of triples composed of Signal, Sample Period,and Stream Identifier indicating the telemetry requested.
  - For collection, the link layer could capture all telemetry data along with the signal, signal value, locator, stream handles, local time, initial packet fields, etc.
  - Alternatively, collection could be configured to only record data from packets with specific stream handles.

Telemetry Updating in VLAN Bridges

VLAN Bridge
EISS Shim Model



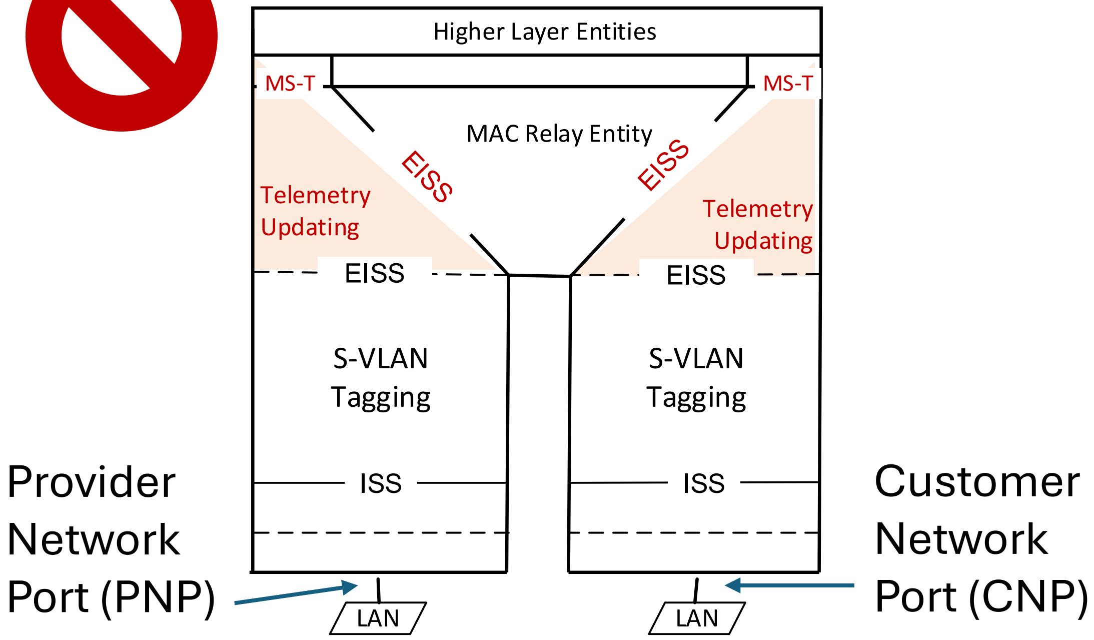
- A VLAN Bridge will only update the telemetry tag if it is immediately after the VLAN tag
- On ingress the telemetry shim timestamps the frame before queuing it
  - The timestamp is carried as a parameter of the EISS connection_identifier
- At egress the telemetry shim:
  - Calculates the signal value for the specified signal type
  - Compares the calculated signal value to the signal value carried in the Telemetry TAG
  - If the compare is for max and the calculated signal value is greater than the tag's signal value the tag is updated
  - If the compare is min and the calculated signal value is less than the tag's signal value the tag is updated

DA	SA	VLAN	TelTAG	Data	FCS

Frame on wire for VLAN Bridge telemetry update

Provider Bridge Port-based service issue

Provider Bridge
With EISS telemetry shim



- This configuration does not provide consistent results
  - For untagged or S-VLAN tagged frames from the customer the telemetry tag is updated normally by the Provider Bridge
  - However, for C-VLAN tagged frames from the customer the telemetry tag is hidden under the C-VLAN tag and therefore not updated by the Provider Bridge

Untagged or S-VLAN tagged Customer Frame

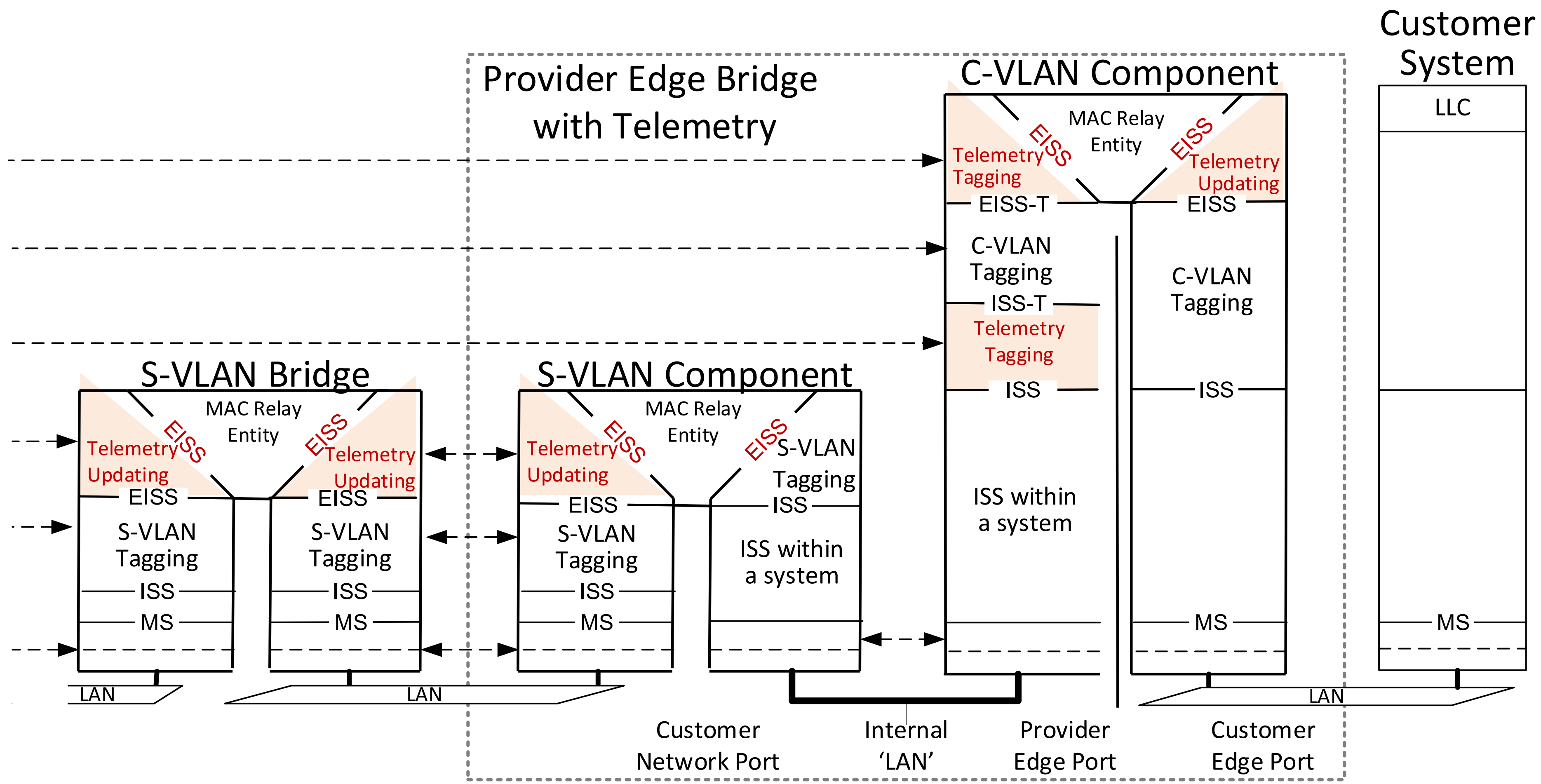
DA	SA	S-VLAN	TelTAG	Data	FCS

C-VLAN tagged Customer Frame

DA	SA	S-VLAN	C-VLAN	TelTAG	Data	FCS

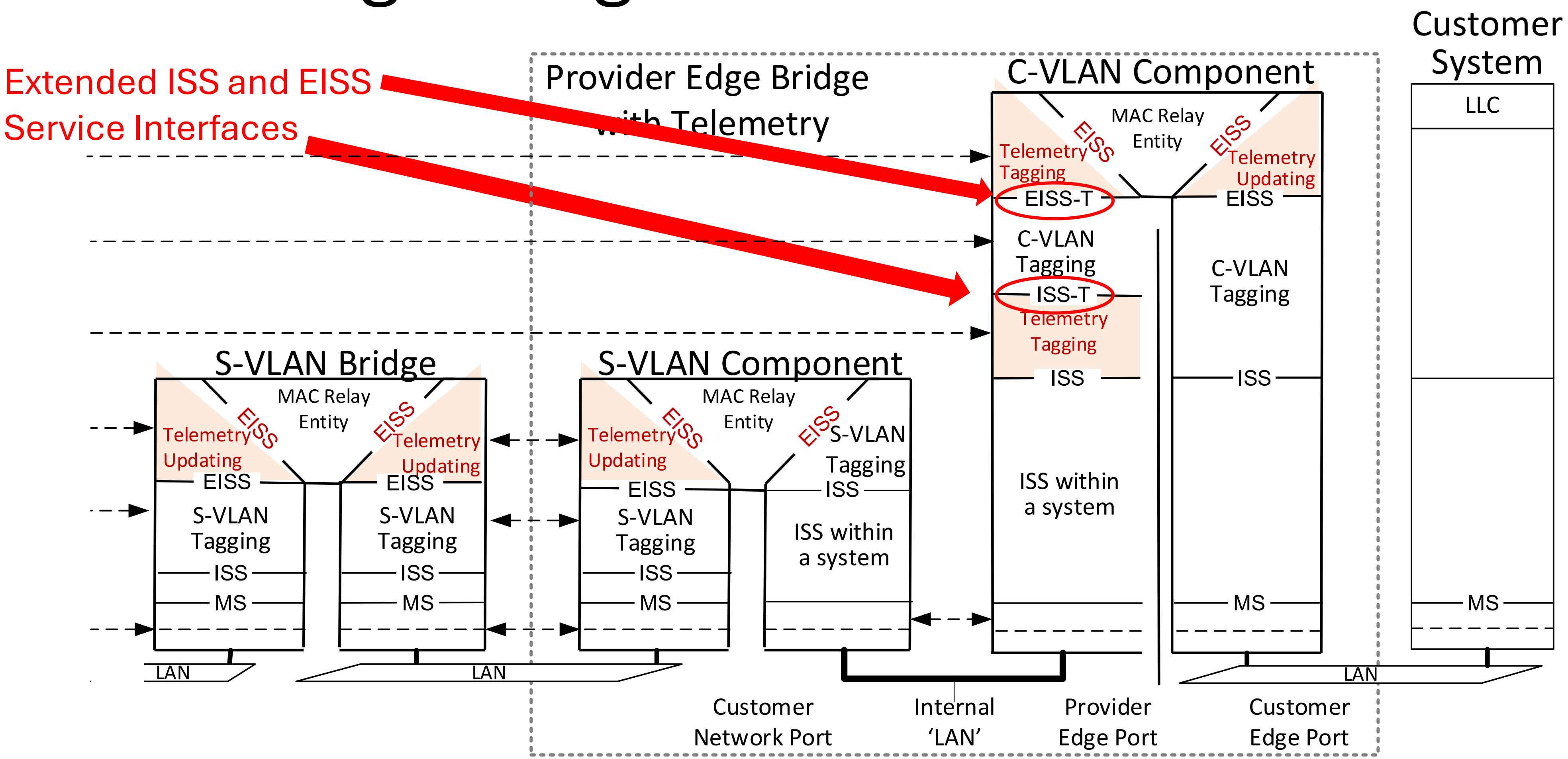
Since the C-VLAN tag is unrecognized by a Provider Bridge it hides the telemetry tag sitting under it

Provider Edge Bridge with telemetry support



- Here the C-VLAN component is used to move the telemetry tag from under the C-VLAN tag to under the S-VLAN tag allowing telemetry updating in the PBN

Provider Edge Bridge service interface extensions



- The ISS-T and EISS-T service interfaces are extensions to the ISS and EISS used to pass the telemetry between the telemetry tagging layers within the C-VLAN Component port.
- These two Telemetry tagging layers are used to move the telemetry tag from under the C-VLAN tag to the first tag on the Provider Edge Port.

Provider Edge Bridge (PEB) encodings choices

Untagged customer frame

DA	SA	S-VLAN	TelTAG	Data	FCS

C-VLAN tagged customer frame

DA	SA	S-VLAN	TelTAG	C-VLAN	Data	FCS

- On ingress to the Provider Bridge Network (PBN) the telemetry tag is moved under the S-VLAN tag to allowing the S-VLAN Bridges to update the tag.
- On egress from the PBN the telemetry tag is moved back under the C-VLAN tag.
- The pro of moving the telemetry tag is that no frame size expansion is generated for transmit over the PBN.
- The con is the measurement of the Provider Bridge Network can't be separated from the customer network.

Untagged customer frame

DA	SA	S-VLAN	TelTAG	TelTAG	Data	FCS

C-VLAN tagged customer frame

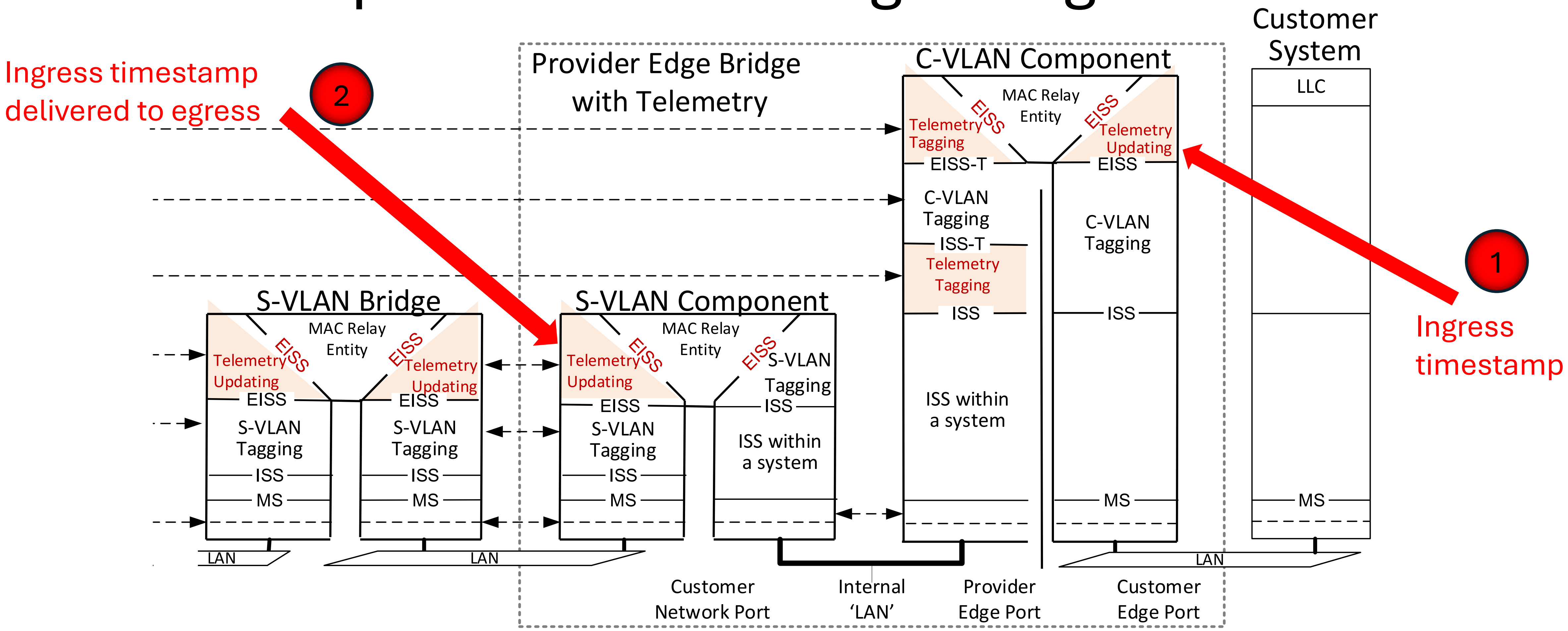
DA	SA	S-VLAN	TelTAG	C-VLAN	TelTAG	Data	FCS

- On ingress to the PBN a new telemetry tag is placed under the S-VLAN tag with the same signal as the customer telemetry tag. This new tag is then updated by the S-VLAN Bridges or the PBN.
- On egress from the PBN the telemetry tag under the S-VLAN tag is compared to the telemetry tag under the C-VLAN tag which is updated based on the signal.
- The pro of replicating the telemetry tag is this allows the PBN to perform a measure of the signal within the PBN separate from the customer network measure.
- The con is expansion of the frame size.

Provider Edge Bridge delay calculation

- Even thought a Provider Edge Bridge (PEB) is modeled as two components (an S-VLAN and C-VLAN components) typical switches have a single set of buffers and therefore the delay calculation should be from ingress to egress of the PEB.
- This requires the timestamp taken at the ingress at the C-VLAN component to be carried over the Internal LAN connecting the C-VLAN Component to the S-VLAN Component.
- Clause 6.14 or IEEE Std 802.1Q supports passing the ISS parameters over the internal LAN. In particular, if the timestamp is encoded as a sub-parameter within the connection_identifier it could be carried over the internal interface.

Timestamps in a Provider Edge Bridge



- Here the timestamp is delivered from the telemetry updating layer of the C-VLAN component (point 1) to the telemetry updating layer of the S-VLAN component (point 2).
- This operations requires passing the timestamp over the internal 'LAN' which is supported by IEEE Std 802.1Q clause 6.14.

Summary

- IEEE 802 Modeling requires supporting extensions to the MAC Service Interface to allow passing telemetry from(to) higher layer protocols
- All our service interfaces A(MS, ISS, and EISS) need a way to support passing timestamp metadata
  - The MAC Service Interface needs either a time_stamp parameter or a general metadata parameter such as the connection_identifier defined for the ISS and EISS.
  - The ISS and EISS could use a sub-parameter of the connection_identifier to support passing a time_stamp parameter
- To support Provider Bridge modeling we will need to be able to pass the telemetry parameters up and down the port stack.
  - The most natural way to do this is to provide ISS and EISS variants supporting the telemetry extension
  - We do not need to pass the telemetry as metadata over the Bridge relay

Thoughts on 802.1EH clauses

- 1.0-4.0 Overview - Abbreviations
- 5.0 Conformance
- 6.0 Support of MAC Service with telemetry
- 7.0 Principles of Telemetry Tagging of Data Frames
- 8.0 Signal type definitions
- 9.0 Principles of operation in End Station
- 10.0 Principles of operation in Bridges
- 11.0 Telemetry tag formats
- 12.0 Telemetry configuration
- 13.0 YANG data models for End Stations
- 14.0 YANG data models for Bridges
- Annex A: PICS proforma – Telemetry Tagging of Data Frames

Thank You

Backup

CSIG as General-Purpose Telemetry

- Many Possible Use Cases for CSIG:
  - End-to-End Congestion Control (requires integration with L4)
  - Path Selection and Load Balancing
  - Debugging and Fault Isolation
  - Traffic Engineering
  - SLA Compliance
- Currently, CSIG is used to provide precision telemetry at the transport layer to improve congestion control, path selection, and load balancing.
- However, since the telemetry is collected and transmitted at L2 it could be used for general purpose telemetry at any of the L2, L3 or L4 layers.
- Providing CSIG support at L2 enables the use of CSIG in environments with conventional transport protocols such as TCP.
  - Conventional transport protocols would require modification to enhance L4 congestion control, but other applications such as debugging and fault isolation can be supported without change to the transport protocol.