Congestion Signaling (CSIG)

Simple and Effective In-band Network Signals for Efficient Traffic Management in Datacenter Networks

Paul Bottorff (HPE)
Ramesh Sivakolundu (Cisco)
Jai Kumar (Broadcom)
Paul Congdon (Huawei)
Brad Karp (Google)

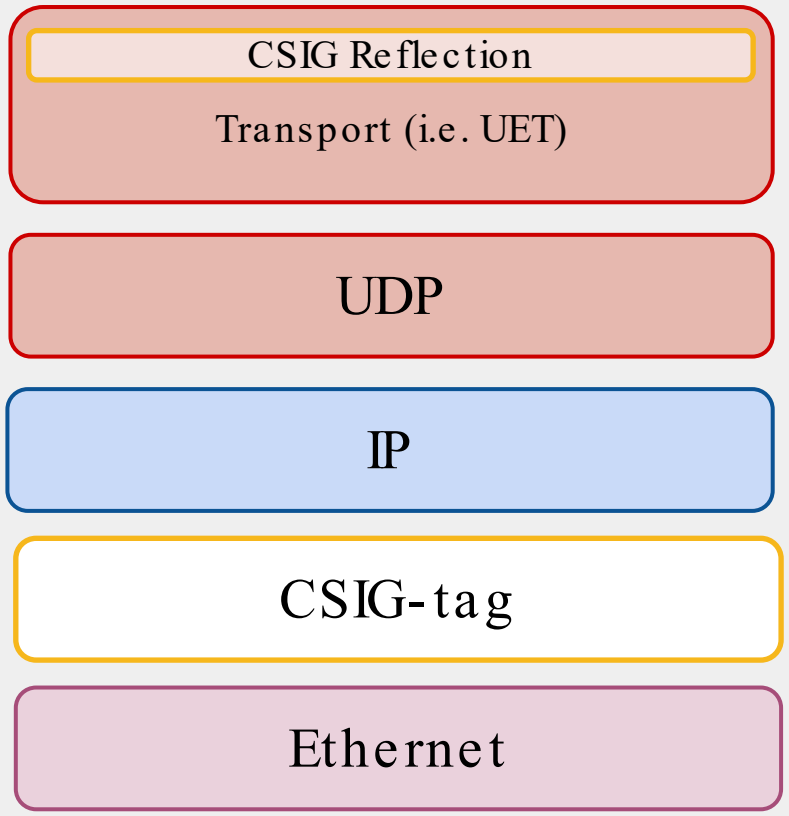
IEEE 802.1 Meeting
November 10, 2025

Workloads: Era of Extreme Network Demands

Accurate and fine-grained congestion signals are needed for network control

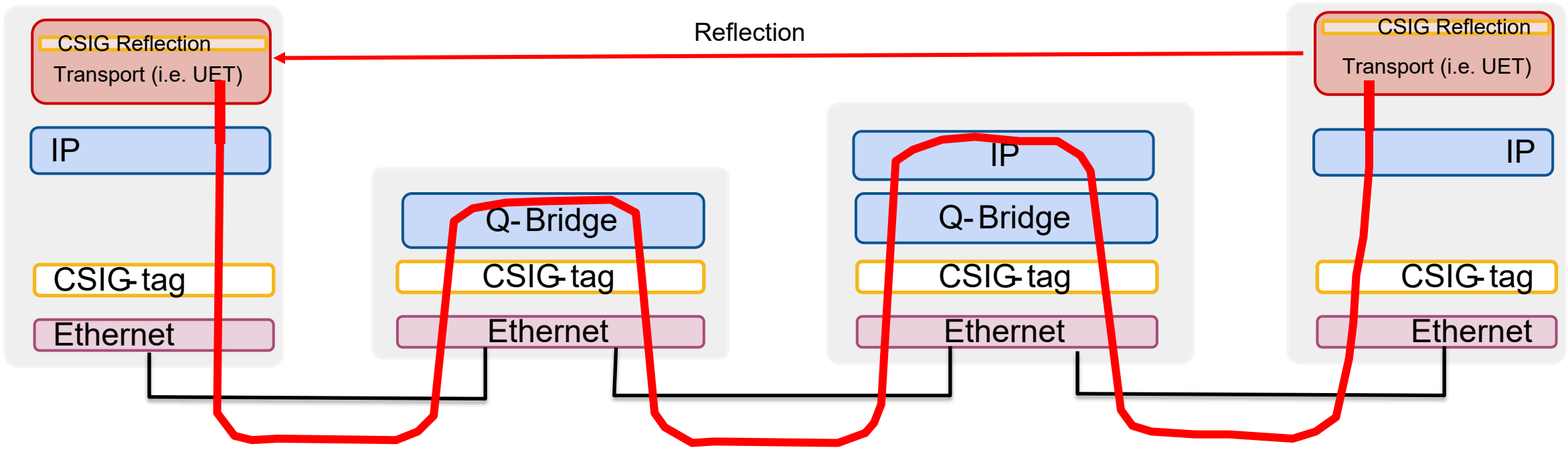
- Continuing trends in the AI era: Horizontal scaling is inevitable
  - Extreme *reliability*, *performance and efficiency* requirements for scale-up and scale-out networks serving AI workloads
  - AI workloads are extremely bandwidth-hungry and tail latency-intolerant
- New norms for network congestion in AI workloads
  - Massive, synchronized bursts that amplify as the network fabric scales
  - Congestion events that manifest at sub-millisecond timescales on network switches
  - Predictable and repeating patterns of short-lived congestion
- Many control loops operate at different timescales for congestion control, load balancing, multipathing, scheduling, traffic engineering, provisioning to
  - Efficiently utilize available network capacity *at fine-grained timescales*
  - Enable tight tail latency and throughput for collectives

CSIG: Practical & Effective In-band Signaling protocol



- Provides fixed-length simple summaries from the path bottlenecks
- Designed for Congestion Control, Traffic Management and Network debuggability use-cases
- Designed for brownfield deployment with backward compatibility / interoperability
- Link to UEC Draft 0.50 from UEC liaison is in public domain- https://github.com/opencomputeproject/OCP-NET-UEC-CSIG

CSIG is a multi-layer telemetry protocol



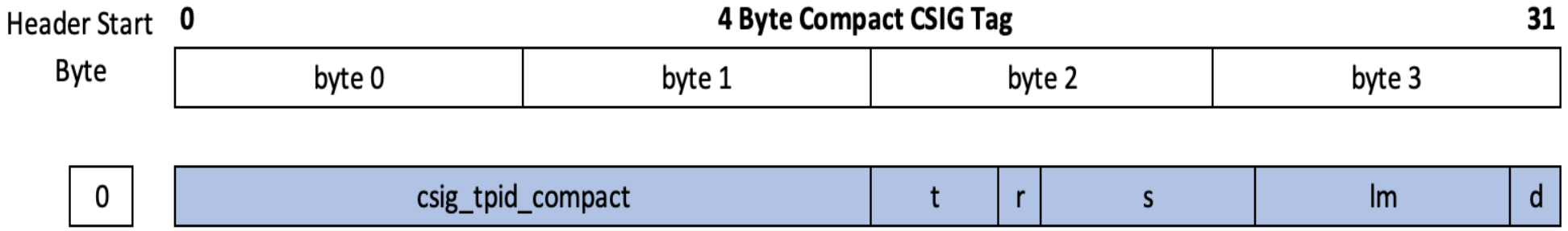
- The CSIG telemetry originates and terminates at the transport layer
  - Current UEC specification, however future work might allow postcards from inside the network
- The CSIG telemetry passes through both L2 and L3 switch layers
- The CSIG telemetry is encoded in an L2 tag on the wire
- The CSIG telemetry tag can be placed on every packet

CSIG tag is carried under the VLAN tag

Untagged	dstmac / srcmac / **csig-tag** / ethertype / payload
802.1q:	dstmac / srcmac / c-tag / **csig-tag** / ethertype / payload
802.1ad	dstmac / srcmac / s-tag / **csig-tag** / c-tag / ethertype / payload
802.1ae	dstmac / srcmac / sec-tag / c-tag / **csig-tag** / ethertype / payload

- For bridges the csig-tag will travel through the bridge with the frame
- For IP hops the content of the tag with travel through the switch along with the packet
- The CSIG telemetry will terminate at the transport layer where the signal will be delivered and reflected within the transport layer protocol

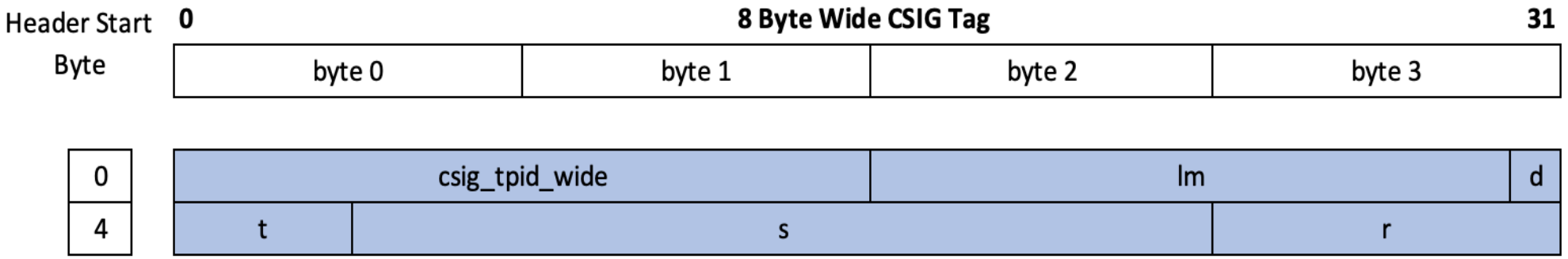
Two L2 Tags: the 4 byte tag is aligned to VLAN tag fields



Bit Offset	Width (in bits)	Field Name	Comments
0-15	16	csig-tpid-cpct	New Ethertype allocated by IEEE
16-18	3	t	Type of Signal
19	1	r	Reserved
20-24	5	s	Quantized Signal Value
25-30	6	lm	Locator Metadata
31	1	d	Do not update (Packet Trimmed) (“(D)ropped“)

- The alignment with VLAN tags is critical to enable retrofitting some existing switches (by changes in firmware and microcode) to support CSIG.
- This has enabled the current largescale deployments at Google.

Two L2 Tags: the 8 byte tag provides fine grained measures



Bit Offset	Width (in bits)	Field Name	Comments
0-15	16	csig-tpid-wide	New Ethertype allocated by IEEE
16-30	15	lm	Locator Metadata
31	1	d	Do not update (Packet Trimmed) (“(D)ropped”)
32-35	4	t	Signal Type
36-55	20	s	Quantized Signal Value
56-63	8	r	Reserved

- With silicon developments currently in progress it will be possible to implement 8 byte CSIG tagging to provide fine grained measurements.

One signal is carried in each tag

t	Signal	Profile	Aggregation Function	Comments
0	ABW	base	min	Available bandwidth per port
1	ABW/C	base	min	Relative available bandwidth per port
2	Delay	base	max	Per-hop delay
3	nQD	extended	max	Queue depth normalized by port speed

- The signal is measured for each hop and updated only if the measure is either greater or less than, depending on signal type, the current CSIG signal value.
- At the destination transport the CSIG signal value is the value of the hop with either the minimum or maximum value.
- Transport is responsible for generating a collection of signal measurement types it needs for it to control congestion and manage multiple network paths.

IEEE Role in CSIG development

- Standardize the CSIG Tag formats.
- Standardize forwarding of CSIG tags within the Bridge architecture.
- Develop an L2 service interface for exposing the CSIG telemetry to Bridge and Router relays and to end stations.
- Provide an IEEE 802.1 management module for CSIG.
- Clarify how forwarding of CSIG operates in provider and provider backbone bridged networks.

Proposed CSIG Project Title

Standard for Local and Metropolitan Area Networks –
Congestion Signaling Telemetry

This project would be a standalone standard rather than a Q amendment.

Proposed Project Scope

- This standard specifies two layer 2 tags, one 4 byte long tag and one 8 byte long tag, which are used to carry light weight telemetry containing congestion signaling information between hosts and switches. It specifies a service interface for exposing the CSIG telemetry at the L2 layer, an architecture model for how CSIG is passed through Bridges, and an IEEE management module for configuring CSIG tag processing.
- The telemetry information carried in these tags is specified by the UE CSIG standard and would be cross referenced by the IEEE standard.
- Management of CSIG tags is specified in UE standards, however would need to be adaptation to the IEEE 802.1 management framework.

Thank You