

TRILL ECN Support

draft-eastlake-trill-ecn-support-00.txt

Donald E. Eastlake, 3rd
d3e3e3@gmail.com

Bob Briscoe
ietf@bobbriscoe.net

ECN Background

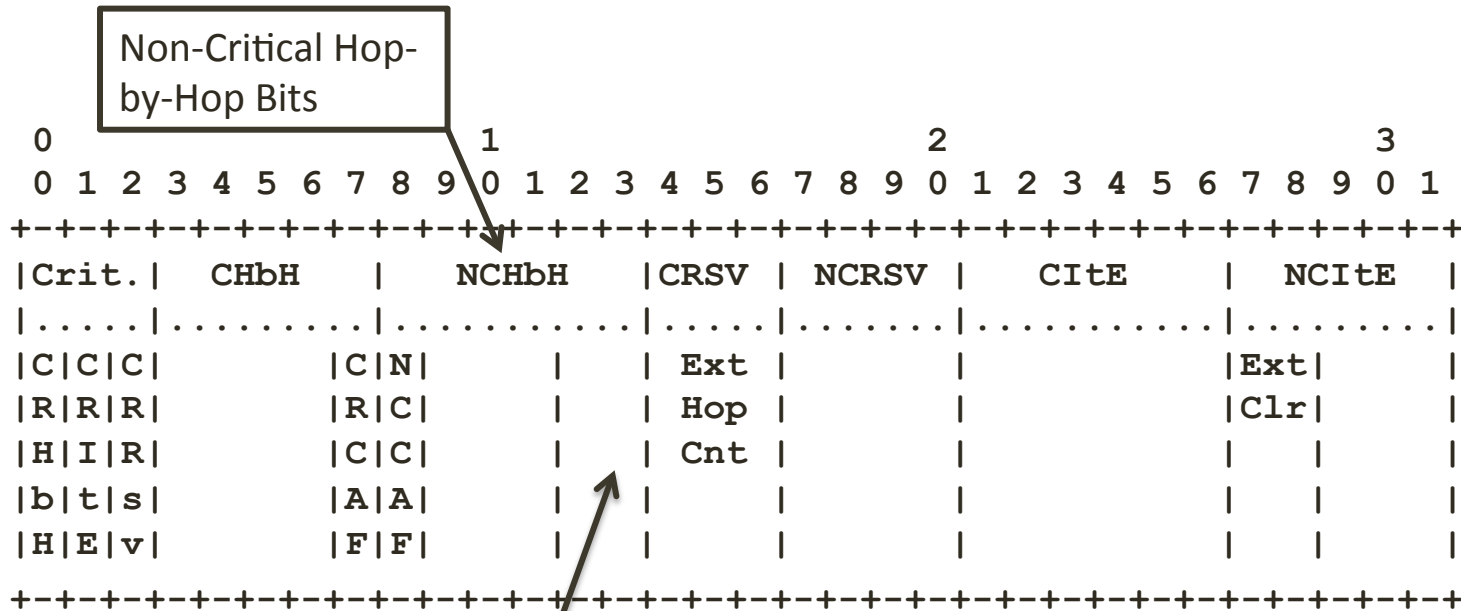
- Explicit Congestion Notification (ECN) in IP [RFC3168]
 - Mark, not drop, packet at congestion onset
 - used extensively in data centres
 - usually for low queue delay with Data Centre TCP (DCTCP)
 - L3 switches mark the IP header
- Enduring incremental deployment problem
 - if legacy receiver or sender would not understand marking
 - congested buffer MUST drop not mark
 - otherwise legacy hosts would black-hole congestion signals
- Solution for e2e transports over IP (eg. TCP, RTP)
 - e2e ECN capability negotiation at flow set-up
 - IP-ECN codepoints for “ECN-capable transport” (ECT)
 - then congested buffers:
 - mark ECT packets or
 - drop Not-ECT packets

IP-ECN codepoint	value	meaning
Not-ECT	00	Not ECN-capable transport
ECT(0)	10	ECN-Capable Transport
ECT(1)	01	
CE	11	Congestion Experienced ('marked')

Adding ECN to TRILL: the problem

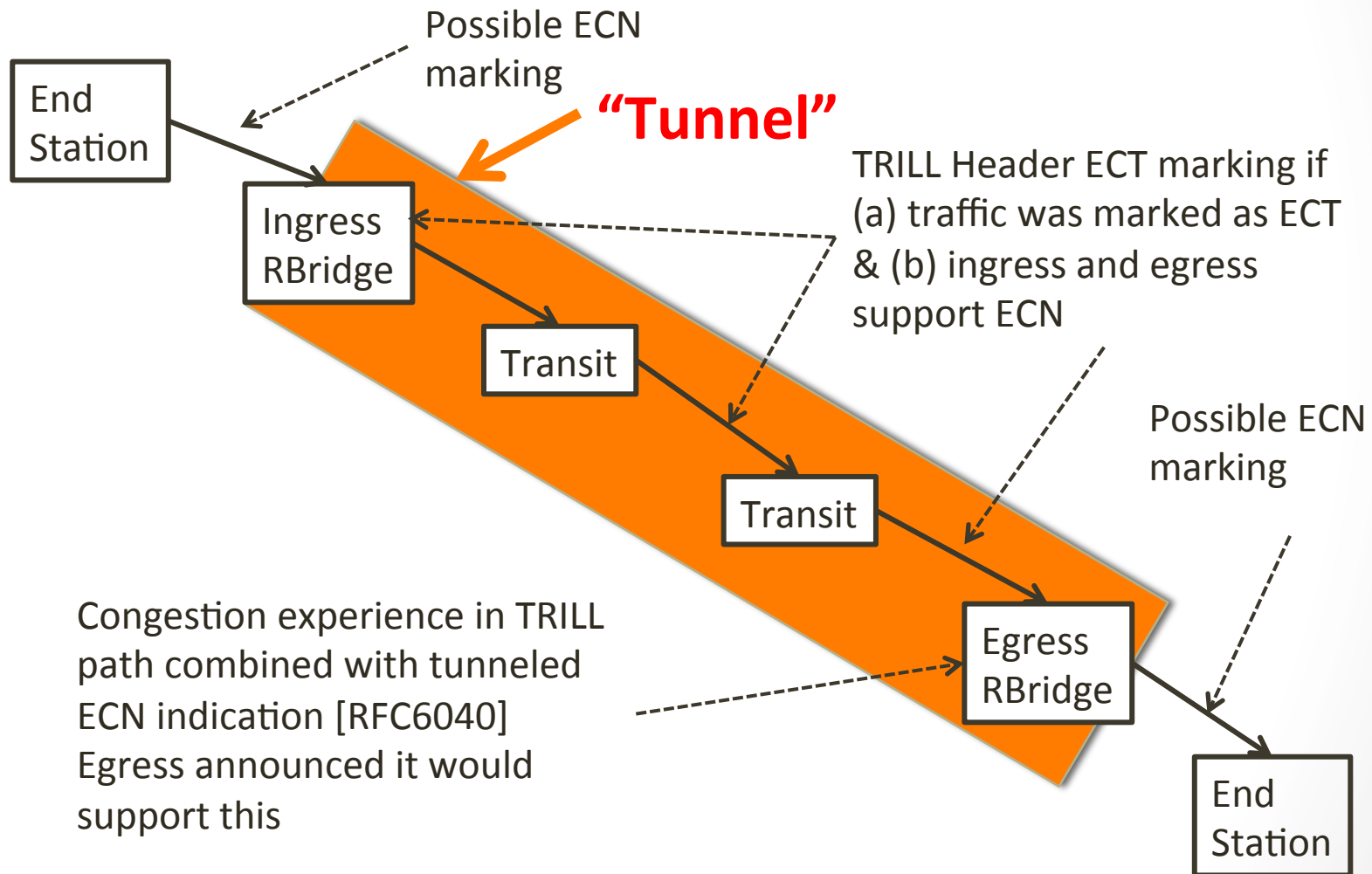
- Similar incremental deployment problem
 - if legacy egress does not understand ECN
 - will not propagate upward to forwarded IP inner header
 - would black-hole congestion signals
- Three potential types of solution:
 1. require TRILL RBridges to act as “L3 switches” and mark IP-ECN field
 2. require ingress-egress ECN capability advertisement (via routing)
 3. use a critical ingress-to-egress flag
- rest of talk
 - #1: set aside as unreasonable
 - #2: described first and in [draft-eastlake-trill-ecn-support-00]
 - #3: described second but not in draft yet (developed after draft deadline)

Approach #2: Requires ingress-egress ECN negotiation

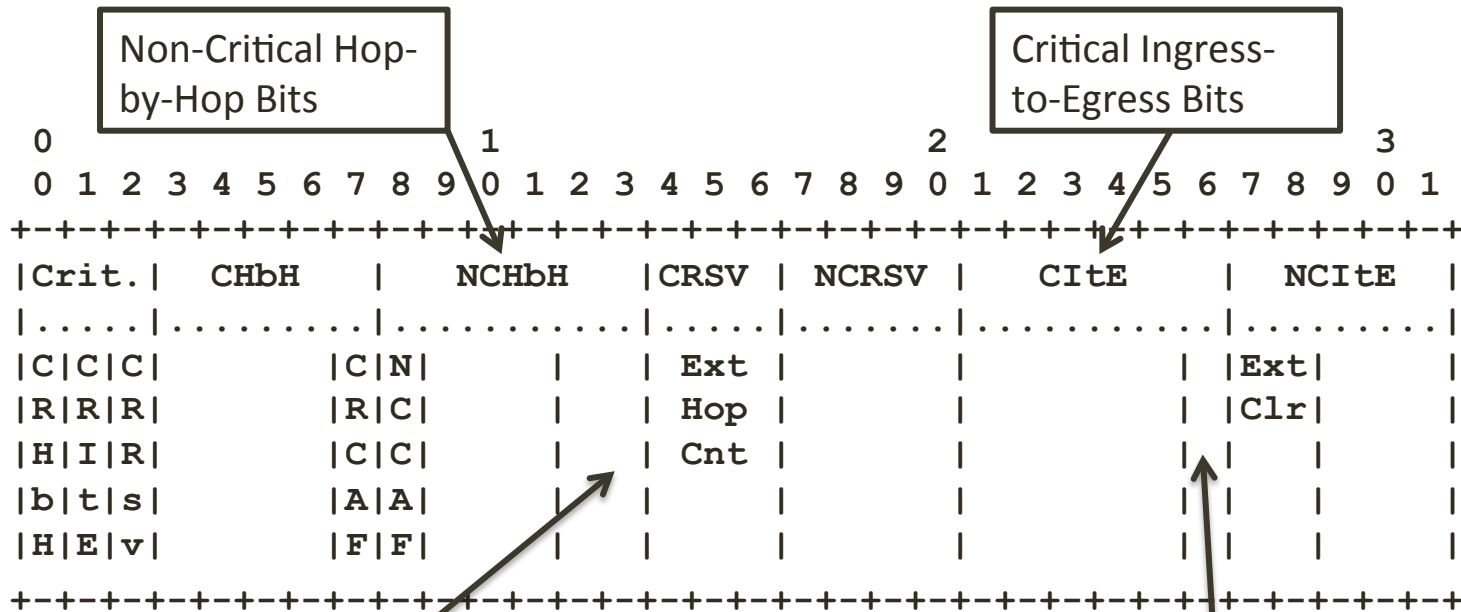


ECN codepoint	value	meaning
Not-ECT	00	Not ECN-capable transport
ECT(0)	10	ECN-Capable Transport
ECT(1)	01	ECN-Capable Transport
CE	11	Congestion Experienced

Approach #2: ECN negotiated between ingress and egress



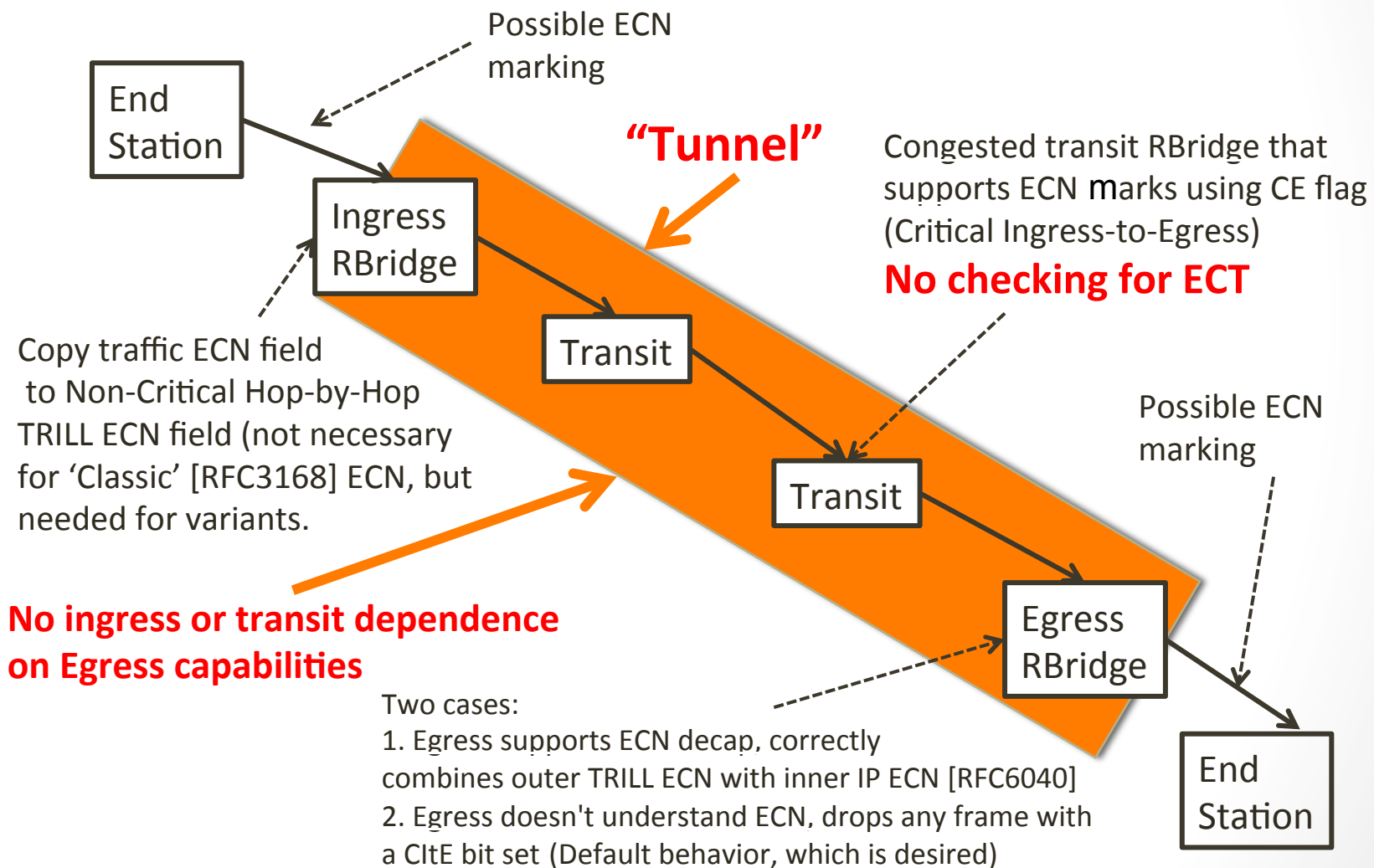
Approach #3: Defer mark / drop decision to egress



ECN codepoint	value	meaning
Not-ECT	00	Not ECN-capable transport
ECT(0)	10	ECN-Capable Transport
ECT(1)	01	
CS	11	Congestion Signalled (Reserved for future use)

Congestion Experienced CE flag

Approach #3: Defer mark / drop decision to egress



Recap: ECN tunneling rules at egress [RFC6040]

incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT	Not-ECT	drop
ECT(0)	ECT(0)	ECT(0)	ECT(1)	CE
ECT(1)	ECT(1)	ECT(1)	ECT(1)	CE
CE	CE	CE	CE	CE
	Outgoing header			

- TRILL egress same as [RFC6040] but 3 ECN bits in proposal #3 so map 3 bits to the 4 codepoints as shown in table:

NCHbH TRILL ECN	CitE TRILL ECN	Incoming Outer
00	0	Not-ECT
10	0	ECT(0)
01	0	ECT(1)
11	0	CE
00	1	CE
10	1	CE
01	1	CE
11	1	CE

Next Step

- Ask for comments on the mailing list
- Call for WG Adoption

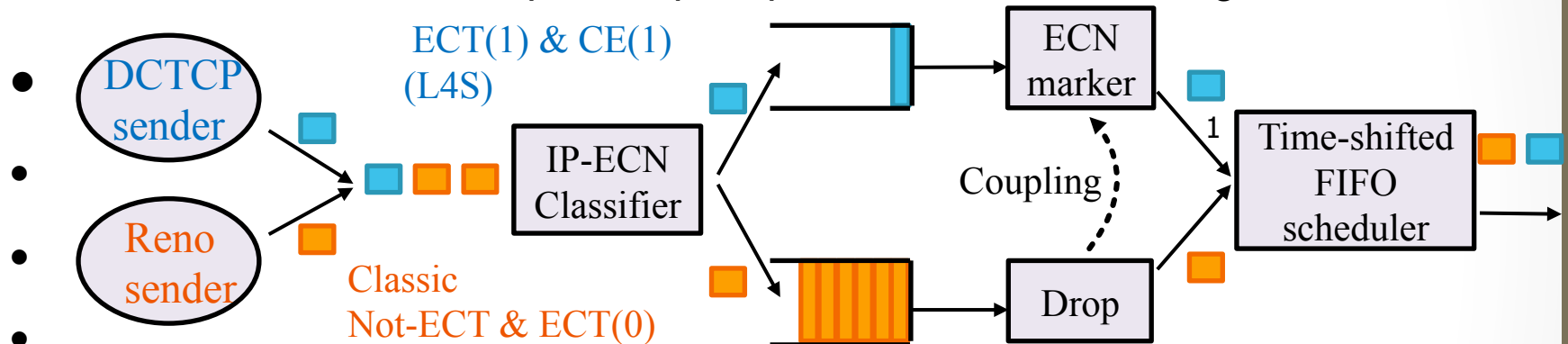
SUPPLEMENTAL SLIDES

Recap: ECN tunneling rules at ingress [RFC6040]

incoming header (also = outgoing inner)	outgoing outer	
	Normal mode (default)	Compatibility mode
Not-ECT	Not-ECT	Not-ECT
ECT(0)	ECT(0)	Not-ECT
ECT(1)	ECT(1)	Not-ECT
CE	CE	Not-ECT

Adding support to TRILL for Low Latency Low Loss Scalable throughput (L4S)

- for background on L4S see:
 - draft-briscoe-aqm-duaq-coupled, draft-briscoe-tsvwg-ecn-l4s-id



- On TRILL RBridge, use solution #3 “defer mark/drop decision”
 - plus: classify on NCHbH TRILL ECN field, as shown above

```
Classic queue:
if (p' > max(random(), random() ) {
    mark(frame, CE(0))
}
```

```
L4S queue:
if (p' > random() ) {
    if (p' > random() ) mark(frame, CE(0) )
    else mark(frame, CE(1) )
}
```

- then deferring mark/drop decision to egress gives desired outcome

End

Donald E. Eastlake, 3rd
d3e3e3@gmail.com

Bob Briscoe
ietf@bobbriscoe.net