Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP

draft-briscoe-tsvwg-ecn-encap-guidelines-03

Bob Briscoe, BT
John Kaippallimalil, Huawei
Pat Thaler, Broadcom

IETF-88 Nov 2013

aim of this draft

- guidelines for writing specs to propagate ECN up to IP from:
  - L2 protocols (e.g. IEEE802, TRILL)
  - tunnelling protocols (L2TP, GRE, PPTP, GTP,…)
- for authors who may not be ECN experts

draft status

- intended status: best current practice
- individual draft-03, **ready for WG adoption**

ECN = explicit congestion notification
L2TP = layer 2 tunnelling protocol [RFC2661]
PPTP = Point-to-point Tunnelling Protocol [RFC2637]
GRE = generic routing encapsulation [RFC1701, RFC2784]
QCN = quantised congestion notification [IEEE 802.1Qau]
GTP = GPRS tunnelling protocol [3GPP TS 29.060]

context / problem

- growing interest in ECN again
  - in recognition of the importance of low delay
  - particularly in L2 networks (backhaul, data centres) & mobile
- AQM & ECN are for queues at any layer
  - not just IP
- ECN has to be explicitly propagated
  - up the layers
- in contrast drop is easy
  - it naturally propagates up the layers

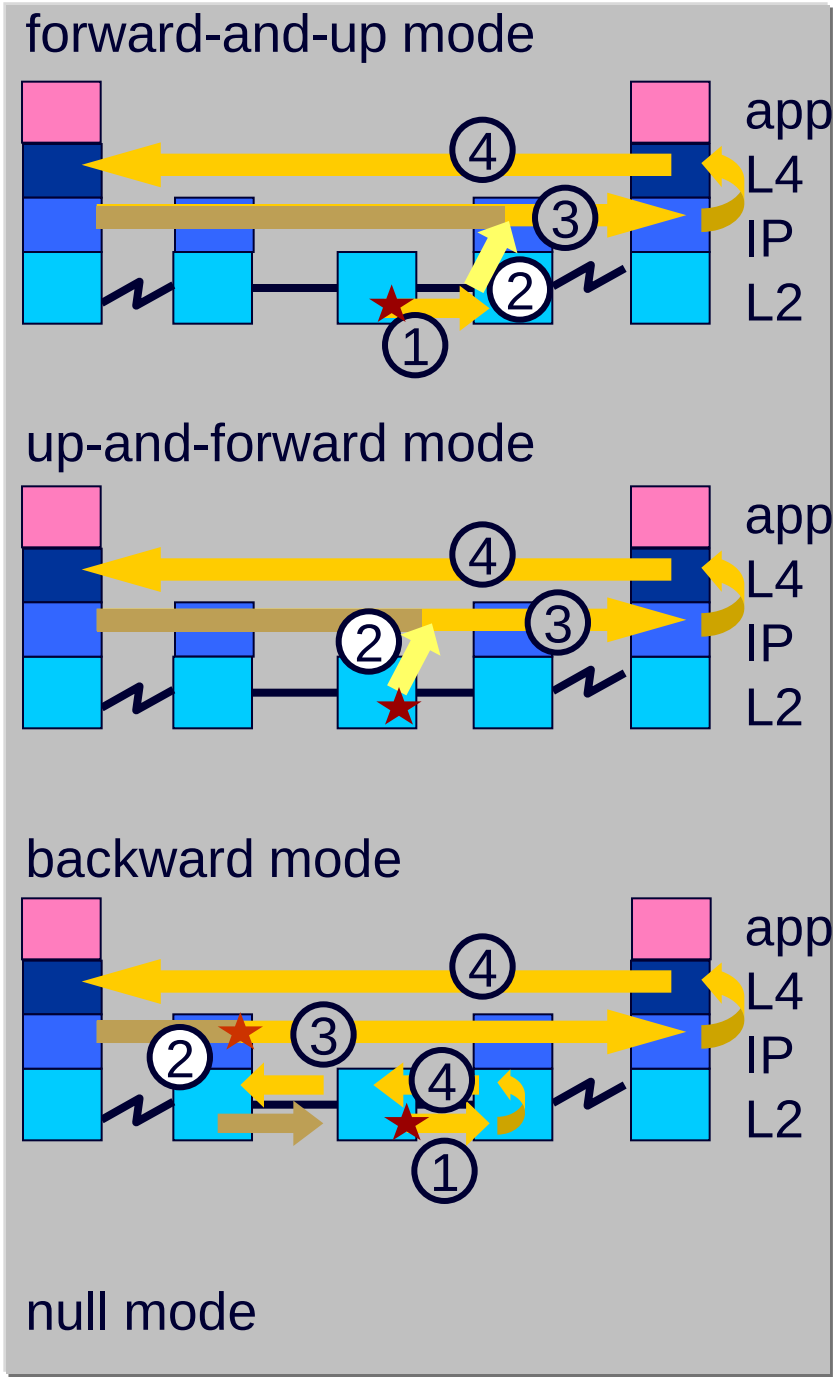
AQM = active queue management (e.g. RED, CoDel, PIE, DCTCP threshold, PCN)

a variety of arrangements

- avoid precluding L2 innovation
- must not be over-prescriptive

- guidelines for each mode
  - see draft (or spare slides)

- wide expertise needed for authoring & review



how would this draft BCP be used?

- authors of L2 & tunnel protocols often not L4 experts
- for IETF maintained protocols
  - e.g. trill, L2TP, GRE, PPTP
  - they can be referred to this draft BCP (e.g. by IESG)
- for protocols maintained by other SDOs
  - while considering this for BCP, and once issued as a BCP IAB would issue liaisons, e.g.
    - to IEEE for 802 protocols
    - to 3GPP for GTP
    - etc.

new in draft-03

Technical

- §1.1 Scope:
  - Added dependence on correct propagation of traffic class
  - For the feed backward mode only: deemed multicast and anycast out of scope
- §4 Feed-Forward-and-Up Mode
  - Wherever it only talked of subnet technologies widened it to tunnels
  - and the converse
- §8 Security Considerations added (all sections now complete)
  - congestion signal fields should be classed as immutable
  - congestion signal integrity best done end-to-end rather than hop-by-hop

Editorial

- none (document is fairly mature now)

next steps

- process
  - request adoption onto wg agenda
  - if adopted, need liaison with other WGs & SDOs
    - notify IETF TRILL, IEEE 802, 3GPP, at least
    - setting requirements for interfacing IP with their protocols
- outstanding document issues
  - listed in Appendix A (next slide)
- thanks to those volunteering to review, so far:
  - Andrew McGregor
  - Wei Xinpeng
  - Richard Scheffenegger
  - Dirk Kutscher
  - (and Gorry Fairhurst already reviewed draft-01)
- and thanks for 12+ expressions of support for adoption on list

Outstanding Document Issues

- Recent review comments
  - [PO'H] clarify that ECN 'as is' gives only incremental benefit
  - [JT] INT area not even motivated by wider recommendations
- Outstanding from previous reviews (recorded in Appendix A):
  - [GF] Certain guidelines warrant MUST (NOT) rather than SHOULD (NOT). Esp:
    - If inner is a Not-ECN-PDU and Outer is CE (or highest severity congestion level), MUST (not SHOULD) drop?
    - *Proposed approach: Express overall intent, not just decap, as MUST (NOT)*
  - Consider whether an IETF Standard Track doc will be needed to Update the IP-in-IP protocols listed in Section 4.1 – at least those that the IETF controls – and which Area it should sit under.
    - *Proposed approach: we think a proposed standard RFC will be needed (probably INT Area, or in TSV for INT), but too early to call*
- Double check: should intended status be BCP or INF?
  - *Proposed approach: Contains normative statements and extrapolates approach in IP-in-IP and MPLS proposed standards, so BCP not just INF seems correct?*

Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP

draft-briscoe-tsvwg-ecn-encap-guidelines-03

Q&A

& spare slides

status of congestion notification
in protocols that encapsulate IP

- IETF

  done: MPLS-in-MPLS, IP-in-MPLS [RFC5129], IP-in-IP [RFC6040]

  to do: trill-rbridge-options (in progress),
  & pass ECN thru tunnel protocols, eg. L2TP, GRE

- Other standards bodies:

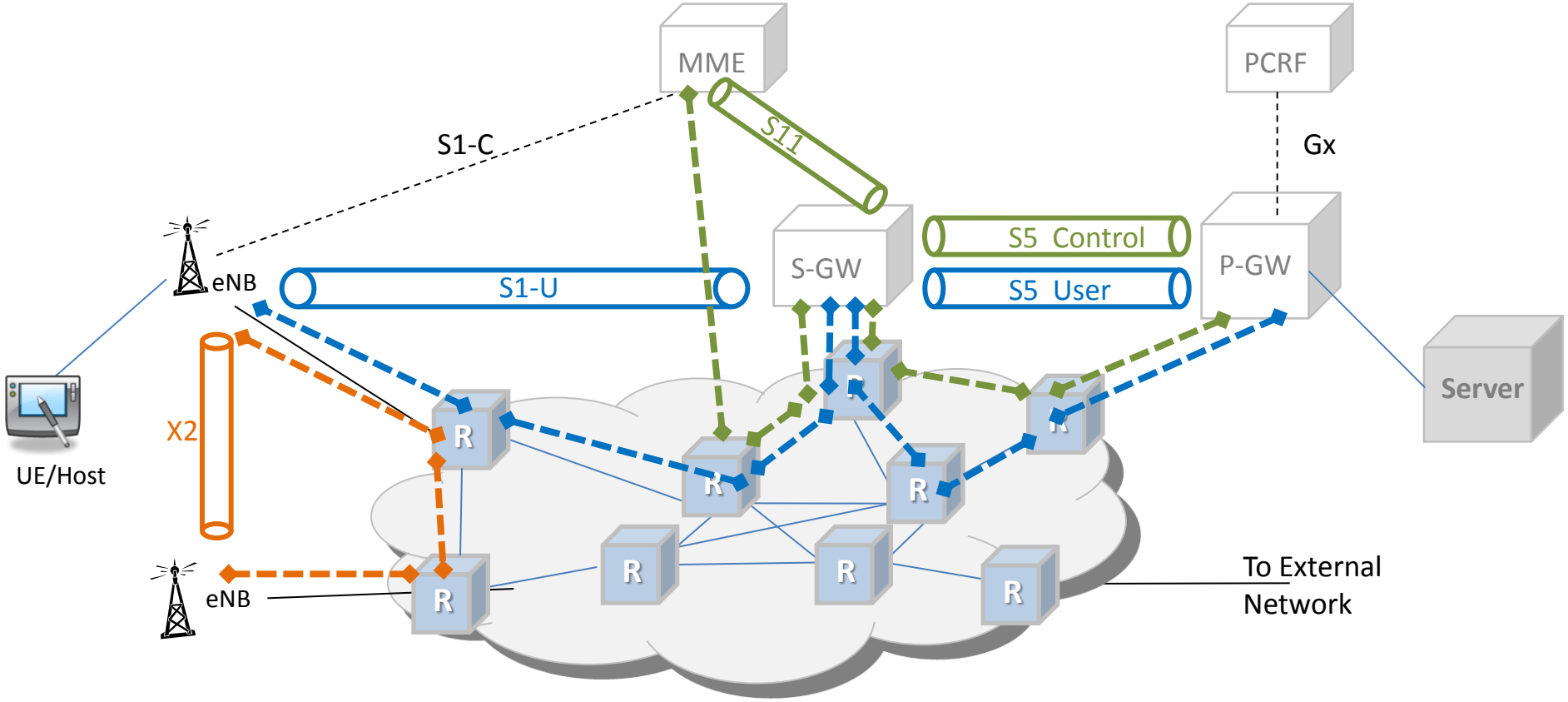
  done: QCN [802.1Qau], Frame Relay, ATM [I.371]
  (all subnet-local)

  todo: IEEE 802.1, (802.3, 802.11), …?
  & pass ECN thru tunnel protocols, eg. 3GPP GTP

L2TP = layer 2 tunnelling protocol [RFC2661]
GRE = generic routing encapsulation [RFC1701, RFC2784]
QCN = quantised congestion notification
GTP = GPRS tunnelling protocol - user plane [3GPP TS 29.281]

motivating example

3GPP LTE/SAE – sequence of tunnels

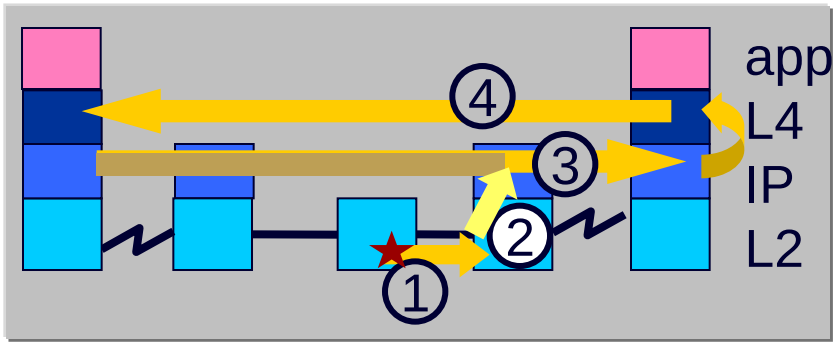


More than 1 tunnel between policy enforcement points.

Example: UE PDN connection traverses

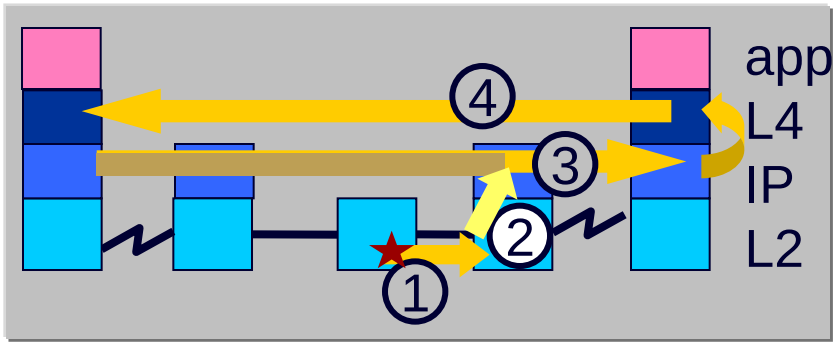
[eNB] << S1-U >> [SGW] << S5/S8 >> [PGW].

forward and upward mode: requirements



- identifying whether transport will understand ECN
- identifying whether egress will understand ECN
- propagating ECN on encapsulation
- propagating ECN on decapsulation
- reframing issues

forward and upward mode: guidelines



- identifying whether transport will understand ECN
  - ‘ECN-capable transport’ codepoint or other approaches
- identifying whether egress will understand ECN
  - new problem
- propagating ECN on encapsulation
  - copying ECN down for monitoring purposes
- propagating ECN on decapsulation
  - combining inner & outer
- reframing issues
  - marked bytes in ≈ marked bytes out
  - timeliness – don’t hold back any remainder

the main problem: incremental deployment

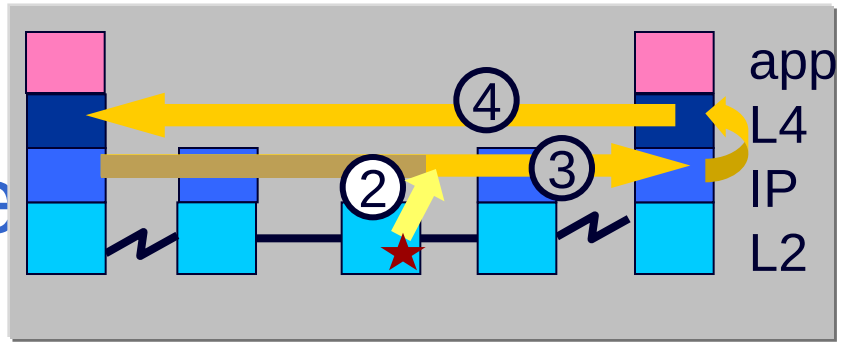
- IP-ECN designed for incremental deployment

		congested queue supports ECN?	
transport supports ECN?	IP header	N	Y
N	Not-ECT	drop	drop
Y	ECT	drop	CE

- if transport only understands drop
  - lower layer must not send it congestion indications
- need not mimic IP mechanism (grey)
  - but needs to achieve same outcome (white)
  - also, must check egress understands ECN too

ECT = ECN-capable transport
CE = Congestion Experienced

up and forward mode guidelines

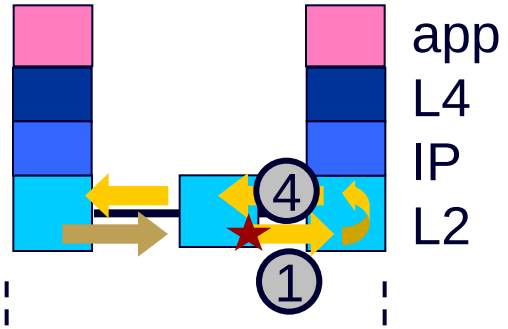


- identifying whether transport will understand ECN
  - use IP mechanism
- ~~identifying whether egress will understand ECN~~
- ~~propagating ECN on encapsulation~~
- ~~propagating ECN on decapsulation~~
- ~~reframing issues~~
- a layering violation
  - but safe if guidelines apply

backward mode

IEEE 802.1Qau (QCN)
ATM ITU-T-I.371
Frame Relay

- often designed for where the subnet is the whole network



- doesn't interwork efficiently with IP's forwards-only mode