Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP

draft-briscoe-tsvwg-ecn-encap-guidelines-02

Bob Briscoe, BT
John Kaippallimalil, Huawei
Pat Thaler, Broadcom
IETF-86 Mar 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-02, ready for WG adoption
- new co-authors
  - John Kaippallimalil, using ECN for GTP in 3GPP
  - Pat Thaler, IEEE 802 1st vice-chair, Data Centre Bridging taskgroup chair

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]
explicit congestion notification (ECN)

• growing interest again
  • in recognition of the importance of low delay
  • particularly in L2 networks (backhaul, data centres) & mobile

• drop: both congestion signal and impairment
  • compromise: deliberately delay the signals (bufferbloat)

• ECN: a signal without impairment
  • can signal as early as needed
problem

• 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)
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
new in draft-02

Technical

• §4.1 IP-in-IP Tunnels with Tightly Coupled Shim Headers
  • L2TP, GRE, PPTP, GTP, VXLAN, ...
  • General advice: RFC6040 applies (ECN/IP-in-IP)
• §4.5 Sequences of Similar Tunnels or Subnets
  • Optimisation: skip decap & re-encap of ECN
• Within §3.1, included a 3GPP example
  • see spare slide #12 for full motivating example

Document

• Added authors: JK & PT
• Roadmap at the start of §4, given the no. of subsections now
• §9 "Conclusions"
changes in draft-02

- Clarified why transports are starting to be able to saturate interior links
- Under § 1.1, addressed the question of alternative signal semantics and included multicast & anycast.
- § 4.2. "Wire Protocol Design":
  - guideline 2: clarified that check egress capability check only applies to the immediate subnet egress, not later ones
  - Added a reminder that it is only necessary to check that ECN propagates at the egress, not whether interior nodes mark ECN
  - Added example of how QCN uses 802.1p to indicate support for QCN.
  - Added references to Appendix C of RFC6040, about monitoring the amount of congestion signals introduced within a tunnel
- Appendix A: Added more issues to be addressed, including plan to produce a standards track update to IP-in-IP tunnel protocols.
- Updated acks and references
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)

• reviewers pls
Outstanding Document Issues

- [GF] Concern that certain guidelines warrant a MUST (NOT) rather than a SHOULD (NOT). Esp:
  - If inner is a Not-ECN-PDU and Outer is CE (or highest severity congestion level), MUST (not SHOULD) drop?
  - Approach: Given the guidelines say that if any SHOULD (NOT)s are not followed, a strong justification will be needed, they have been left as SHOULD (NOT) pending further list discussion.
- [GF] Impact of Diffserv on alternate marking schemes (referring to RFC3168, RFC4774 & RFC2983)
- 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.
- Guidelines referring to subnet technologies should also refer to tunnels and vice versa.
- Check that each guideline allows for multicast as well as unicast.
- Security Considerations
Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP

draft-briscoe-tsvwg-ecn-encap-guidelines-02

Q&A
& spare slides
status of congestion notification in protocols that encapsulate IP

• IETF
  - 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

\[\text{[eNB]} \ll \text{S1-U} \gg \text{[SGW]} \ll \text{S5/S8} \gg \text{[PGW]}\].
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 \( \approx \) marked bytes out
  - timeliness – don’t hold back any remainder
the main problem: incremental deployment

• IP-ECN designed for incremental deployment

<table>
<thead>
<tr>
<th>congested queue supports ECN?</th>
<th>IP header</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>transport supports ECN?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Not-ECT</td>
<td>drop</td>
<td>drop</td>
</tr>
<tr>
<td>Y</td>
<td>ECT</td>
<td>drop</td>
<td>CE</td>
</tr>
</tbody>
</table>

• 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

- often designed for where the subnet is the whole network

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

Incoming load unchanged

Backs up into L3

Slows down L2

Congestion f/b

Not a good fit