Adding ECN to TCP control packets (and retransmits)

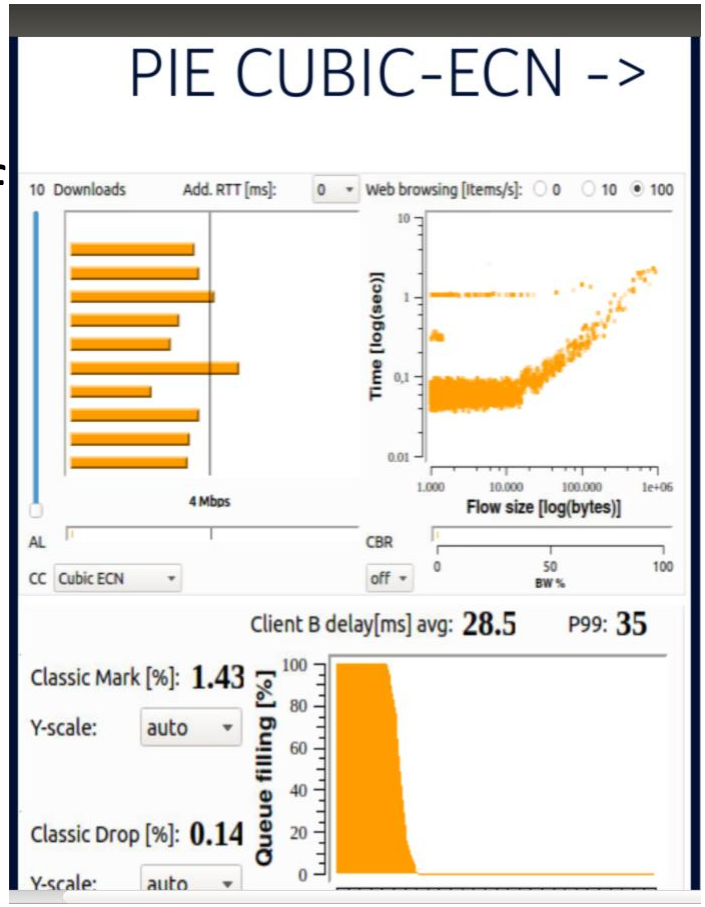
draft-bagnulo-tsvwg-generalized-ecn
tcpm – IETF96
Marcelo Bagnulo
Bob Briscoe

Current situation

- RFC3168 stats that ECN must not be used in SYNs, SYN/ACKS, Pure ACKs, Window Probes, retransmitted packets
  - It is silent about RSTs and FINs
  - RFC5562 enables it for SYN/ACKs
- Provides a number of arguments for doing this.

Performance penalty

- In congestion situations, non-ECT packets are likely dropped instead of marked
- SYN: 0.14% drop leads to flow completion time = 1s for 0.14% of short flows (top stripe in figure)
- In DCs and L4S, dropping TCP control packets results in severe performance penalties
  - See Judd, G., "Attaining the promise and avoiding the pitfalls of TCP in the Datacenter", NSDI 2015, 2015.



Reliability argument

- Overarching principle in RFC3168
  - *To ensure the reliable delivery of the congestion indication of the CE codepoint, an ECT codepoint MUST NOT be set in a packet* ***unless the loss of that packet in the network would be detected by the end nodes and interpreted as an indication of congestion****.*
- Overly conservative
- Suggest to use the do not harm principle
  - The CE signal must be as reliable as the congestion signal resulting by the loss of the marked packet.
  - This implies that by definition any packet can be marked

SYNs

- Argument #1: Discard ECT SYN by the responder
- Non-issue in DC environments
- Possible behaviours in public Internet
  - Reply with some form of SYN/ACK
    - 99,18% of Alexa top 1M as per Trammell et al. study
  - Reply with RST
    - Retransmit a non ECT SYN (adds 1 RTT, caching may help)
  - Silent discard
    - 0,82% of Alexa top 1M as per Trammell et al. study
    - Replaces performance penalty due to congestion per policy-based discard
    - Caching can help
    - Send one ECT SYN and a non ECT SYN with small delay

SYNs

- Argument #2: Loss of congestion information when non ECT SYN/ACK is returned
- Neither TCP nor DCTCP provides means to feedback ECE information in SYN/ACKs
- AccECN provides a solution
- But what to do when a SYN/ACK is received?
  - The initiator doesn't know if the CE was set in the SYN
  - Reduce initial CWND? Too much penalty?
  - Caching may help?

SYNs

- Argument #3: DoS attacks
  - Second, the ECN-Capable codepoint in TCP SYN packets could be misused by malicious clients to "improve" **the well-known TCP SYN attack**. By setting an ECN-Capable codepoint in TCP SYN packets, a malicious host might be able to inject a large number of TCP SYN packets through a potentially congested ECN-enabled router, **congesting it** even further.
- Attack to the endpoint (SYN flood)
  - Attackers are likely to set the ECT when launching attacks anyway (rfc3168 does not reccomends dropping these packets)
  - DoS attacks can be cause with all kinds of packets (and this is not an argument for not marking them)
- Attack to the router (congesting if further)
  - RFC3168 already mandates that “AQM MUST turn off ECN support if under persistent overload” which addresses this issue

Pure ACKs

- Argument #1: Reliability, commented before
- Argument #2: lack of means to react
  - The receiver of the congestion signal may only be sending ACKs, so no means to reduce the load.
  - This would be no worse than the current situation (i.e. if an ACK is lost, the sender of the ACK does not react to the congestion signal) and potentially better (if it is sending data, it can reduce the CWND)

Retransmitted packets

- Argument #1: reliability (commented before)
- Argument #2: DoS attacks
  - Using CE marked packet to reduce the CWND
  - First, an attacker would set the CE anyway
  - Second, the protection comes not from not allowing the set of CE but to not react to out of window packets, as recommended by 3168

Window probes

- Argument #1: reliability, commented before

Argument against ECT	Rebuttal / Solution
SYN	
Responder may discard ECT SYN	Not required, but 0.6%-0.8% do.
	Cache the failure & rexmt Not-ECT SYN
Responder may have no means to f/b CE on SYN	If AccECN confirmed, assures CE f/b support
	If no AccECN, reduce IW conservatively
DoS attacks	See common rebuttal
Pure ACK	
Unreliable CE delivery	See common rebuttal
No means to f/b CE (no ACKs of ACKs)	No worse than drop of Pure ACK, and better performance
Re-xmt	
Unreliable CE delivery	See common rebuttal
DoS attacks	See common rebuttal
Over-reacting to congestion	Correct to react twice to congestion in different RTTs
Window Probe	
Unreliable CE delivery	See common rebuttal

Next steps

- read and comment, pls – thx for useful comments so far
  - more arguments against ECT control pkts?
- add discussion of FINs & RSTs
- more experiments
- turn draft into an experimental track spec before seek adoption
- write brief separate draft enabling ECT control pkt experiments:
  - network “MUST NOT” drop ECT control packets
  - updates RFC3168 (PS)