More Accurate ECN Feedback in TCP

draft-ietf-tcpm-accurate-ecn-06

Bob Briscoe <ietf@bobbriscoe.net>
Mirja Kühlewind <mirja.kuehlewind@tik.ee.ethz.ch>
Richard Scheffenegger <rscheff@gmx.at>
Background & Problem

• **Explicit Congestion Notification (ECN):** Routers make packets as Congestion Experienced (CE) instead of dropping them in case of incipient congestion

• **ECN Feedback in RFC6831:** Receiver only provides feedback once per RTT to the sender

• **Accurate ECN (AccECN):** Receiver feeds back the accurate number of seen markings (within each RTT)
Overview AccECN

- **Capability Negotiation**: Repurposing the former NS (ECN Nonce Sum) TCP header flag
  - fully backward compatible

- **Essential Feedback**: Overloading the ECN TCP header flags (NS/ECE/CWR) as Accurate ECN (ACE) field
  - feed back the number of received CE marks (including control packets without payload)
  - no overhead compared to classic ECN but limited resilience to loss

- **Supplementary Feedback**: Using a new AccECN TCP Option
  - provide additional feedback on the number of marked bytes

- Both essential and supplementary parts: receiver maintains ECN-IP-codepoint counters and AccECN repeats LSBs of counters for resilience
The ACE field

The (post-ECN Nonce) definition of the TCP header flags (bytes 13 & 14):

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               |           |
| A | C | E | U | A | P | R | S | F |
| Header Length | Reserved |
| E | W | C | R | C | S | S | Y | I |
|               |           |   | R | E | G | K | H | T | N | N |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Definition of the ACE field (when AccECN has been negotiated and SYN=0):

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               |           |           |
| U | A | P | R | S | F |
| Header Length | Reserved | ACE |
| R | C | S | S | Y | I |
|               |           | G | K | H | T | N | N |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```
The AccECN Option

| Kind = TBD1 | Length = 11 | EE0B field |
| EE0B (cont’d) | ECEB field |
| EE1B field |

**EE0B**  number of bytes received with ECT(0) marked
**ECEB**  number of bytes received with CE marked
**EE1B**  number of bytes received with ECT(1) marked

*optional*
Usage of the AccECN TCP Option

• **Change-Triggered ACKs**
  MUST send immediate ACK if an arriving packet increments a different byte counter

• **Continual Repetition**
  SHOULD include if CE-bytes-counter has incremented (MUST give precedence to SACK if space is limited)

• **Full-Length Options Preferred**
  SHOULD always use full-length AccECN Options; MAY use shorter AccECN Options if space is limited, but it MUST include the counter(s) that have incremented since the previous AccECN Option

• **Beaconing Full-Length Options**
  MUST include a full-length AccECN TCP Option on at least three ACKs per RTT
Implementation Status

- Linux patch available: https://github.com/mirjak/linux-accecn/
  - Use of net.ipv4.tcp_ecn=4 to enable AccECN
  - Fallback detection mechanisms incl. recently added IP codepoint feedback in handshake not implemented yet
  - No counter wrap detection implemented yet

- TCP Experimental Option Experiment Identifier (TCP ExID) registered with IANA:
  - 0xACCE
  - TCP Option is requested with publication (IESG approval)
Re-assignment of the „NS“ flag

• RFC8311 "Relaxing Restrictions on Explicit Congestion Notification (ECN) Experimentation“ declares RFC 3540 (ECN Nonce) as historic and de-assigns the NS bit; now marked as „reserved“

• IANA TCP Header Flags registration policy is „Standards Action“
  
  • AccECN is an experimental TCP extension that uses the former NS bit for negotiation and as part of the ACE field
  
  • Hum at last tcpm meeting to assign to AccECN with IESG approval