

More Accurate ECN Feedback in TCP

Bob Briscoe, Independent

<ietf@bobbriscoe.net>



Mirja Kuhlewind, **ETH** zürich

<mirja.kuehlewind@tik.ee.ethz.ch>



Richard Scheffenegger,  **NetApp**[®]

<rs.ietf@gmx.at>



TCPM WG, IETF-106, Jul 2019

Problem (Recap): Congestion Existence, not Extent

- Explicit Congestion Notification (ECN)
 - routers/switches mark more packets as load grows
 - RFC3168 added ECN to IP and TCP

IP-ECN	Codepoint	Meaning
00	not-ECT	No ECN
10	ECT(0)	ECN-Capable Transport
01	ECT(1)	
11	CE	Congestion Experienced

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

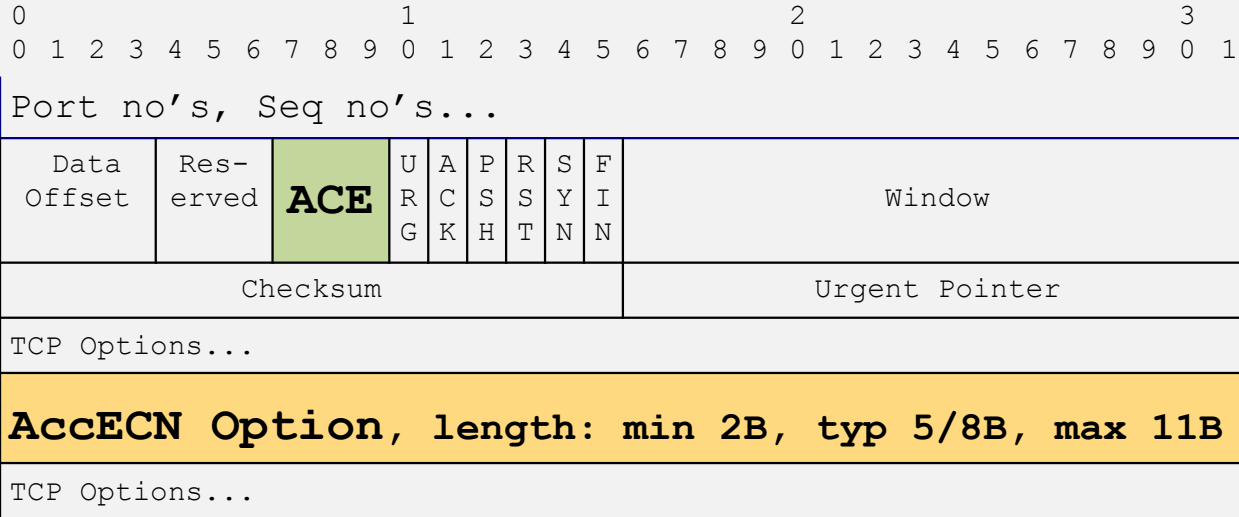
Port no's, Seq no's...

Data Offset	Reserved	N	C	E	U	A	P	R	S	F	Window					
		S	W	C	R	C	S	S	Y	I						
		R	E		G	K	H	T	N	N						
Checksum										Urgent Pointer						
TCP Options...																

- Problem with RFC3168 ECN feedback:
 - only one TCP feedback per RTT
 - rcvr repeats **ECE** flag for reliability, until sender's **CWR** flag acks it
 - suited TCP at the time – one congestion response per RTT

Solution (recap): Congestion Extent, not just Existence

- AccECN: Change to TCP wire protocol
 - Repeated count of CE packets (**ACE**) - essential
 - and CE bytes (**AccECN Option**) – supplementary



- Key to congestion control for low queuing delay
 - 0.5 ms (vs. 5-15 ms) over public Internet

AccECN TCP Option Field Order

kind	length	EE0B [init=1]	ECEB [init=0]	EE1B [init=0]
------	--------	---------------	---------------	---------------

kind	length	EE1B [init=0x800001]	ECEB [init=0]	EE0B [init=0]
------	--------	----------------------	---------------	---------------

- AccECN TCP Option
 - can omit fields that have not changed since the last ACK from the right hand end
- NEW?: Two types of AccECN TCP Options with the same kind
 - switch order of fields for whole connection
 - dependent on initial value of first field on first option received

Fixed inconsistencies

- Previously, WG made AccECN TCP Option optional
- Michael Scharf noticed that change-triggered ACKs text was inconsistent
 - said “...MUST send a change-triggered AccECN TCP Option...”
 - (a change-triggered ACK is an ACK that is triggered by change of a field)
 - made us realize we needed to describe change-triggered ACKs separately for the two cases:
 - ACE field only
 - ACE and AccECN TCP Option

Next Steps

-
- WGLC