More Accurate ECN Feedback in TCP

Bob Briscoe, Independent  
Mirja Kuhlewind, ETH Zürich  
Richard Scheffenegger, NetApp

TCPM WG, IETF-106, Nov 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

<table>
<thead>
<tr>
<th>IP-ECN</th>
<th>Codepoint</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>not-ECT</td>
<td>No ECN</td>
</tr>
<tr>
<td>10</td>
<td>ECT(0)</td>
<td>ECN-Capable Transport</td>
</tr>
<tr>
<td>01</td>
<td>ECT(1)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CE</td>
<td>Congestion Experienced</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>kind</th>
<th>length</th>
<th>EE0B [init=1]</th>
<th>ECEB [init=0]</th>
<th>EE1B [init=0]</th>
</tr>
</thead>
<tbody>
<tr>
<td>kind</td>
<td>length</td>
<td>EE1B [init=0x800001]</td>
<td>ECEB [init=0]</td>
<td>EE0B [init=0]</td>
</tr>
</tbody>
</table>

- **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