Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-02.txt

Bob Briscoe, BT
IETF-74 tsvwg Mar 2009
exec summary

Tech changes:

- **ingress (no change from -01 draft):**
  - brings into line with RFC4301 IPsec

- **egress:**
  - save two wasted codepoint combinations
  - one proposed at IETF-73: generally agreed to go for it
    - needed by PCN but more general
  - one proposed by Anil Agarwal on list
  - both have no backward compatibility issues
    - because they use previously unused codepoint combinations

- **Baked: ready for review**
  - apologies for late posting
  - complete re-write
  - solely standards action text (17pp)
  - shifted motivation, impact analysis etc to appendices or trash

- **Plan**
  - list of 6 volunteer reviewers
    - question: all 3 changes ok?
  - socialise in PCN now
  - socialise with IPsec w-g once rough consensus in tsvwg (Jul)
Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-02.txt

Bob Briscoe, BT
IETF-74 PCN Mar 2009
status

- Layered Encapsulation of Congestion Notification
  - new WG draft:  [draft-ietf-.tsvwg-ecn-tunnel-02.txt](draft-ietf-.tsvwg-ecn-tunnel-02.txt) 24 Mar '09
  - intended status: standards track
  - RFC pub target:  ? TBA
  - immediate intent: review specifically: fix to decap as well as encap?
  - w-gs & r-gs affected: TSVWG, PCN, ICCRG, IPsec, Internet Area?
recap (exec summary)

• scope
  • all IP in IP (v4, v6) tunnels, all DSCPs
  • solely wire protocol processing of tunnelled ECN, not marking or response algorithms

• sequence of standards actions led to perverse position
  • non-IPsec ECN tunnels [RFC3168] have vestige of stronger security than even IPsec [RFC4301] decided was necessary!
  • limits usefulness of 3168 tunnels
    – ingress: PCN stds track "excess rate marking" works with 4301 but not 3168
    – egress: PCN 2-level marking lost requires complex work-rounds or reduced function

• ingress: bring ECN tunnelling [RFC3168] into line with IPsec [RFC4301]
• egress: use two wasted combinations of inner & outer codepoints
  • absolutely no backwards compatibility issues
### Ingress Recap

**Encapsulation at Tunnel Ingress**

<table>
<thead>
<tr>
<th>Incoming Header (also = outgoing inner)</th>
<th>Outgoing Outer</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFC3168 ECN limited functionality</td>
<td>RFC3168 ECN full functionality</td>
</tr>
<tr>
<td>Not-ECT</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>ECT(0)</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>ECT(1)</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>CE</td>
<td>Not-ECT</td>
</tr>
</tbody>
</table>

- **Proposal**: unchanged compatibility state for legacy
- **'Reset' CE**: no longer used
- **'Copy' CE**: becomes normal state for all IP in IP
current egress behaviour

- OK for current ECN
- but any changes to ECT lost
  - effectively wastes ½ bit in IP header
  - again, for safety against marginal threat that IPsec decided was manageable
- PCN tried to use ECT(0/1)
  - but having to waste DSCPs instead
  - or other complex work-rounds
  - or hobbled function

<table>
<thead>
<tr>
<th>incoming inner</th>
<th>incoming outer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not-ECT</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>ECT(0)</td>
<td>ECT(0)</td>
</tr>
<tr>
<td>ECT(1)</td>
<td>ECT(1)</td>
</tr>
<tr>
<td>CE</td>
<td>CE</td>
</tr>
</tbody>
</table>

(!!!) = illegal combination, egress MAY raise an alarm

Outgoing header (RFC3168 & RFC4301)
new egress rules (appendix in -01, normative in -02)

- no effect on any legacy
  - adds new capability using previously illegal combinations of inner & outer
  - only tunnels that need the new capability need to comply
  - an update, not a fork

Dropping unnecessarily prevented future use

![Diagram showing encapsulation at tunnel ingress and decapsulation at tunnel egress]

- Outgoing header (proposed update)
  (bold = proposed change for all IP in IP)

- (!!!) = illegal combination, egress MAY raise an alarm

- propagates changed outer
text changes draft-01→ 02

• scope reduced solely to ECN in IP in IP tunnels
  • removed ECN design guidelines for any layered encapsulation (e.g. ethernet)

• changes to egress made normative
  • one was tentative in appendix (proposed last IETF)
  • other suggested by Anil Agarwal on list

• completely restructured and largely rewritten
  • solely standards action text
  • bloat (justification, analysis) removed or shifted to appendices
next steps

• ready for full review now
  • list of 6 volunteers
  • main question: all three changes ok?
  • remember, these are nuances to the behaviour of the neck of the hour-glass

• socialise in PCN

• once rough concensus in tsvwg, socialise in IPsec (Jul)
  • will need to assure IPsec folks that they don't have to change (again)
### Backward & Forward Compatibility

<table>
<thead>
<tr>
<th>ingress</th>
<th>egress</th>
<th>I-D ecntunnel</th>
<th>RFC 4301</th>
<th>RFC 3168</th>
<th>RFC 2481</th>
<th>RFC 2401/2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>mode</td>
<td>comprehensive</td>
<td>4301</td>
<td>full</td>
<td>lim</td>
<td>2481</td>
<td>lim?</td>
</tr>
<tr>
<td>action</td>
<td>calc C</td>
<td>calc B</td>
<td>calc B</td>
<td>inner</td>
<td>calc A</td>
<td>inner</td>
</tr>
<tr>
<td>compre-hensive</td>
<td>I-D.ecntunnel</td>
<td>normal</td>
<td>'copy'</td>
<td>C</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>compat</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
<td>inner</td>
<td>inner</td>
</tr>
<tr>
<td>'3g IPsec'</td>
<td>RFC4301</td>
<td>4301</td>
<td>'copy'</td>
<td>C</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>ECN</td>
<td>RFC3168</td>
<td>full</td>
<td>'reset CE'</td>
<td>C</td>
<td>n/a</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limited</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ECN expt</td>
<td>RFC2481</td>
<td>2481</td>
<td>'copy'?</td>
<td>C</td>
<td>n/a</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limited?</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>'2g IPsec'</td>
<td>RFC2401</td>
<td>-</td>
<td>'copy'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**Calculation Notes:**

- **C**: calculation C (more severe multi-level markings prevail)
- **B**: calculation B (preserves CE from outer)
- **A**: calculation A (for when ECN field was 2 separate bits)
- **inner**: forwards inner header, discarding outer
- **n/a**: not allowed by configuration
Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-02.txt

Q&A