Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-03.txt

Bob Briscoe, BT
IETF-75 tsvwg Jul 2009

This work is partly funded by Trilogy, a research project supported by the European Community www.trilogy-project.org
status

- Tunnelling of Explicit Congestion Notification
  - new WG draft: draft-ietf-tsvwg-ecn-tunnel-03.txt 21 Jul '09
  - intended status: standards track
  - updates: 3168, 4301
  - RFC pub target: Dec ‘09
  - immediate intent: reviews req’d from Sec Area & tsvwg (again)
  - w-gs & r-gs affected: TSVWG, PCN, ICCRG, IPsecME, Int Area?

- 5 reviews, 4 very extensive
  - resulted in major re-write (again), apologies for late posting
  - one tech change (optional alarm)
  - shifted all non stds stuff to end or deleted.
### Ingress Recap

#### Encapsulation at Tunnel Ingress

<table>
<thead>
<tr>
<th>Incoming Header (also = outgoing inner)</th>
<th>RFC3168 ECN Limited Functionality</th>
<th>RFC3168 ECN Full Functionality</th>
<th>RFC4301 IPsec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not-ECT</td>
<td>Not-ECT</td>
<td>Not-ECT</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>ECT(0)</td>
<td>Not-ECT</td>
<td>ECT(0)</td>
<td>ECT(0)</td>
</tr>
<tr>
<td>ECT(1)</td>
<td>Not-ECT</td>
<td>ECT(1)</td>
<td>ECT(1)</td>
</tr>
<tr>
<td>CE</td>
<td>Not-ECT</td>
<td>ECT(0)</td>
<td>CE</td>
</tr>
</tbody>
</table>

#### Decapsulation at Tunnel Egress

- 'copy' CE becomes normal state for all IP in IP
- 'reset' CE no longer used
- Proposal unchanged compatibility state for legacy

- RFC3168 ECN limited functionality
- RFC3168 ECN full functionality
- RFC4301 IPsec
current egress behaviour

- OK for current ECN
  - 1 severity level of congestion
- any outer changes into ECT(0/1) lost
  - reason: to restrict covert channel (but 2-bit now considered manageable)
  - effectively wastes ½ bit in IP header

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

Outgoing header (RFC4301 \ RFC3168)

encapsulation at tunnel ingress
decapsulation at tunnel egress

got these wrong in -02 whoops!
new egress rules

- cater for ECT(1) meaning either more severe or same severity as ECT(0)
  - for PCN or similar schemes that signal 2 severity levels
- only changing currently unused combinations
  - optional alarms added to all unused combinations
- drop potentially unsafe unused combinations
  - where congestion marked in outer but inner says transport won’t understand
- only tunnels that need the new capability need to comply
  - an update, not a fork
  - no changes to combinations used by existing protocols (backward compatible)

<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>

<table>
<thead>
<tr>
<th>outgoing header (proposed update)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(bold = proposed change for all IP in IP)</td>
</tr>
</tbody>
</table>

- (!!!) = currently unused combination, egress MAY raise an alarm
- ( ! ) = ditto, but alarm will need to be turned off (e.g. if PCN used)

a change into ECT(1) propagates from outer

drop both – for safety
IPsec & non-IPsec now consistent

encapsulation at tunnel ingress

decapsulation at tunnel egress

ECN
DS
encapsulation at tunnel ingress

decapsulation at tunnel egress

'I'
E
E
E
E

Not-ECT
ECT(0)
ECT(1)
CE

Not-ECT
Not-ECT
( !!! )
drop (!!!)
drop (!!!)

ECT(0)
ECT(0)
ECT(1) ( ! )
CE

ECT(1)
ECT(1) ( !!! )
ECT(1)
CE

CE
CE
CE ( !!! )
CE

Outgoing header (proposed update)
draft-ietf-tsvwg-ecn-tunnel-03.txt

Tech changes to RFC3168 or 4301

(red = changed since -02)

• ingress:
  – brings RFC3168 into line with 4301 IPsec

• egress:
  – only changes to previously unused combinations (guarantees backward compatible)
  – propagates 2 severity levels of congestion
    • uses previously unused codepoint combination
    • no change for packets using 1 severity level
  – optional alarms on all currently unused combinations (PCN considered unused – turn off alarm when deployed)
  – two unused combinations dropped for safety (originally one in RFC3168, none in RFC4301)
  – future standards actions SHOULD NOT use ECT(0) outer + Not-ECT inner as indication of congestion, without giving strong reasons
main text clarifications draft-02→ 03

- shifted all non stds stuff nearer to end or deleted
- “Changes from Earlier RFCs” & “Backward Compatibility”
  - organised by RFC, not by ingress / egress
- added appendix on ECN tunnelling in earlier RFCs
  - 2003 (original IP in IP), 2401 (obsolete IPsec), 2481 (ECN expt)
- distinguished static & discovered tunnels more clearly
  - out of scope to specify (proprietary) legacy mode negotiation
  - instead lays down constraints on legacy mode negotiation
next steps

• Jul 09: socialise in Security Area
• Aug 09: request tsvwg re-review
  • 2/7 volunteered reviews still to come
• Nov 09: ask for WG last call
Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-03.txt

Q&A
### backward & forward compatibility

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mode</td>
<td>full</td>
<td>lim</td>
<td>2481 IPsec</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</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>
<td>inner</td>
<td>inner</td>
<td>inner</td>
</tr>
<tr>
<td>'comprehensive'</td>
<td>I-D.ecntunnel</td>
<td>normal</td>
<td>'copy'</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</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>
<td>inner</td>
<td>inner</td>
<td>inner</td>
</tr>
<tr>
<td>'3g IPsec'</td>
<td>RFC4301</td>
<td>-</td>
<td>'copy'</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</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>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>limited</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
<td>inner</td>
<td>inner</td>
<td>inner</td>
<td>inner</td>
</tr>
<tr>
<td>RFC2481</td>
<td>2481</td>
<td>'copy'</td>
<td>C</td>
<td>n/a</td>
<td>B</td>
<td>n/a</td>
<td>A</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>2481 IPsec</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
<td>inner</td>
<td>n/a</td>
<td>inner</td>
<td>inner</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>
<td>inner</td>
<td>A</td>
<td>inner</td>
</tr>
</tbody>
</table>

| RFC2003 |                |
|---------|                |
| 10      |                |

- **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 or negotiation
path support for 2 severity levels of congestion

- do all decapsulators on path propagate 2 levels?
  - PCN: controlled domain: configured by operator
  - future e2e scheme: hosts can’t tell (open issue)