status

- Tunnelling of Explicit Congestion Notification
  - intended status: standards track
  - updates: 3168, 4301 (if approved)
  - RFC pub target: Dec ‘09
  - immediate intent: tsvwg review (again) of changes to error states then Security Directorate review
  - w-gs & r-gs affected: TSVWG, PCN, ICCRG, IPsecME, Int Area?

- relentless discussion since mid-Sep:
  - David Black, Gorry Fairhurst, Phil Eardley & I
  - reaching consensus since I-D deadline
  - minutiae of egress output for invalid combinations of inner & outer
  - but minutiae are important – these are changes to IP

- detailed re-review of -04 text by Gorry Fairhurst
egress behaviour in existing RFCs

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

<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>Not-ECT</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>Not-ECT</td>
<td>Not-ECT</td>
</tr>
<tr>
<td>Not-ECT</td>
<td>Not-ECT</td>
</tr>
</tbody>
</table>

Outgoing header (RFC4301 \ RFC3168)
egress rules in -04 (same as -03)

- 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>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>

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

(!!!) = currently unused combination, egress MAY raise an alarm
( ! ) = ditto, but alarm will need to be turned off (e.g. if PCN used)
a change in ECT(1) propagates from outer
**Egress rules proposed for -05**

- **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 unused combinations **unless inconsistent and not unsafe**
- **drop potentially unsafe unused combination**
  - where **high severity** 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: Egress Rules

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

**Outgoing header (proposed update)**

**(bold = proposed change for all IP in IP)**

(!!!) = currently unused combination, egress MAY raise an alarm

PCN objected to one alarm;
other removed for consistency;
OK – not a safety alarm
### egress behaviour in existing RFCs

- **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: Incoming vs. Outgoing Header

<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>

- **Outgoing header (RFC4301 \ RFC3168)**
main text changes draft-03→ 04

• no functional changes
• added appendix on ‘Open Issues’
• minor textual clarifications
next steps

• Nov 09: request tsvwg re-review
  • 2 reviews volunteered (Jason Livingood & David Black)

• Nov/Dec 09: socialise in Security Directorate
  • reviewers already lined up

• Once resolved: WG last call?
Tunnelling of Explicit Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-04.txt

Q&A
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)
## Backward & Forward Compatibility

<table>
<thead>
<tr>
<th>Ingress</th>
<th>I-D.ecn-tunnel</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>egress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'Comprehensive'</td>
<td>normal</td>
<td>'copy'</td>
<td>C</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>'compat'</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>'3g IPsec'</td>
<td>RFC4301</td>
<td>-</td>
<td>'copy'</td>
<td>C</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>
</tr>
<tr>
<td></td>
<td>limited</td>
<td>'zero'</td>
<td>C</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>'2g IPsec' IP in IP</td>
<td>RFC2401 RFC2003</td>
<td>-</td>
<td>'copy'</td>
<td>C</td>
<td>n/a</td>
</tr>
</tbody>
</table>

- **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

Notes:
- RFC 2401/2003
- RFC 2481 IPsec
- Broken: loses CE

**Explanation**
- **Calculation C**: Calculation C is used when more severe multi-level markings are necessary.
- **Calculation B**: Calculation B preserves the CE from the outer header.
- **Calculation A**: Calculation A is for when the ECN field was divided into separate bits.
- **Inner**: Forwards the inner header while discarding the outer.
- **Not Allowed**: Not allowed by configuration or negotiation due to RFC standards.
ingress recap

encapsulation at tunnel ingress

decapsulation at tunnel egress

<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>

proposal
unchanged compatibility mode for legacy
'reset' CE no longer used
'copy' CE becomes normal mode for all IP in IP