Tunnelling of Explicit Congestion Notification

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

Bob Briscoe, BT

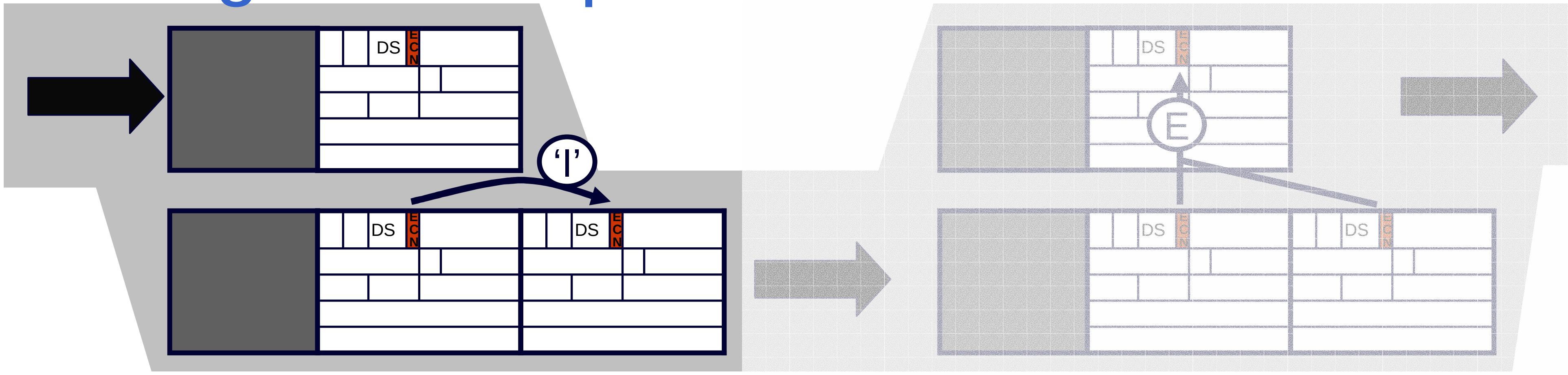
IETF-75 tsvwg Jul 2009

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



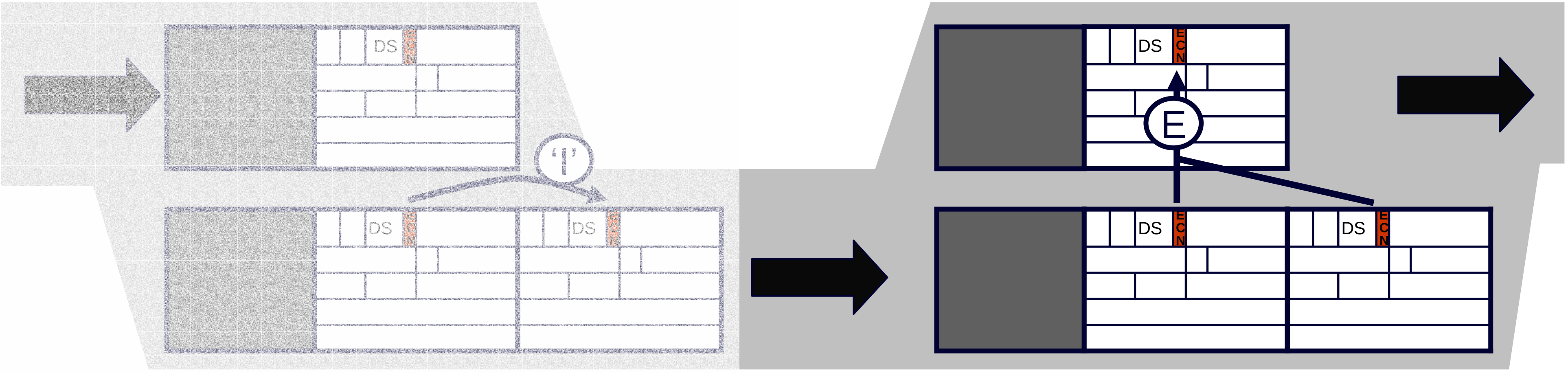
encapsulation at tunnel ingress

decapsulation at tunnel egress

‘I’

incoming header (also = outgoing inner)	outgoing outer		
	RFC3168 ECN limited functionality	~~RFC3168 ECN full functionality~~	RFC4301 IPsec
Not-ECT	Not-ECT	~~Not-ECT~~	Not-ECT
ECT(0)	Not-ECT	~~ECT(0)~~	ECT(0)
ECT(1)	Not-ECT	~~ECT(1)~~	ECT(1)
CE	Not-ECT	~~**ECT(0)**~~	**CE**
proposal	unchanged **compatibility state** for legacy	**'reset' CE no longer used**	'copy' CE becomes **normal state** for all IP in IP

current egress behaviour



encapsulation at tunnel ingress

decapsulation at tunnel egress

- 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

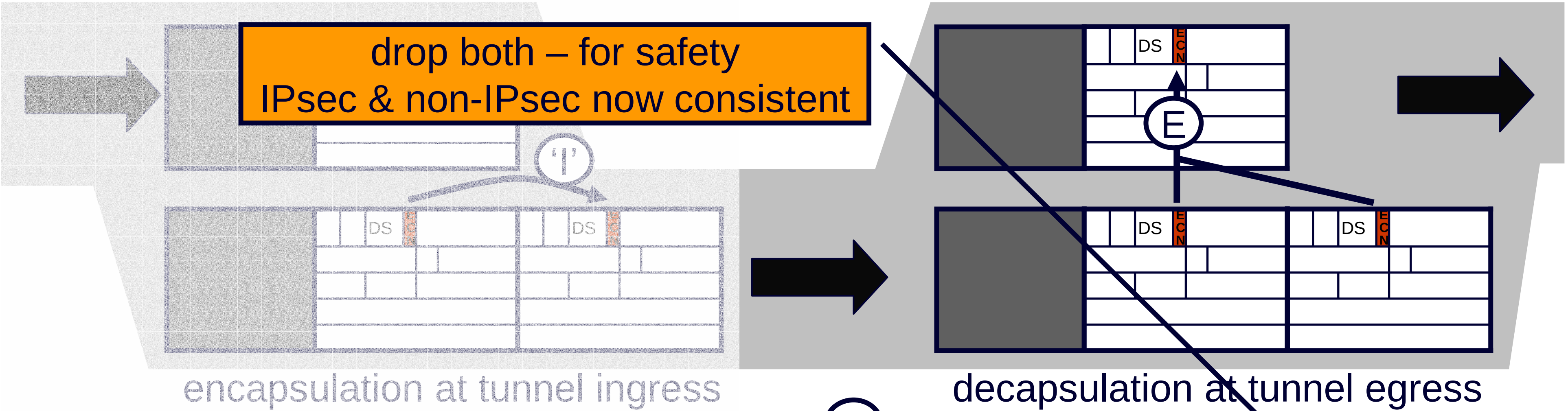
E

incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT	Not-ECT	drop Not-ECT
ECT(0)	ECT(0)	ECT(0)	ECT(0)	CE
ECT(1)	ECT(1)	ECT(1)	ECT(1)	CE
CE	CE	CE	CE	CE

Outgoing header (RFC4301 \ RFC3168)

got these wrong in -02
whoops!

new egress rules


‘I’
DS
ECN
DS
ECN
encapsulation at tunnel ingress
DS
ECN
E
DS
ECN
DS
ECN
decapsulation at tunnel egress

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

- 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)

E

incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT (!!!)	**drop (!!!)**	**drop (!!!)**
ECT(0)	ECT(0)	ECT(0)	**ECT(1) (!)**	CE
ECT(1)	ECT(1)	ECT(1) **(!!!)**	ECT(1)	CE
CE	CE	CE	CE **(!!!)**	CE
	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 into ECT(1) propagates from outer

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

ingress			egress	**I-D ecn-tunnel**	RFC 4301	RFC 3168		RFC 2481		RFC 2401/ 2003
		mode		**-**	-	full	lim	2481	2481 IPsec	-
			action	**calc C**	calc B	calc B	inner	calc A	inner	inner
‘comprehensive’	**I-D.ecn-tunnel**	**normal**	**'copy'**	C	B	B	n/a	n/a	n/a	n/a
		compat	**'zero'**	C	n/a	n/a	inner	inner	inner	inner
'3g IPsec'	RFC4301	-	'copy'	C	B	B	n/a	n/a	n/a	n/a
ECN	RFC3168	full	'reset CE'	C	n/a	B	n/a	n/a	n/a	n/a
		limited	'zero'	C	n/a	n/a	inner	inner	inner	inner
ECN expt	RFC2481	2481	'copy'	C	n/a	B	n/a	A	n/a	n/a
		2481 IPsec	'zero'	C	n/a	n/a	inner	n/a	inner	inner
'2g IPsec' IP in IP	RFC2401 RFC2003	-	'copy'	C	n/a	n/a	inner	A	inner	broken: loses CE

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)