Tunnelling of Explicit Congestion Notification

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

Bob Briscoe, BT

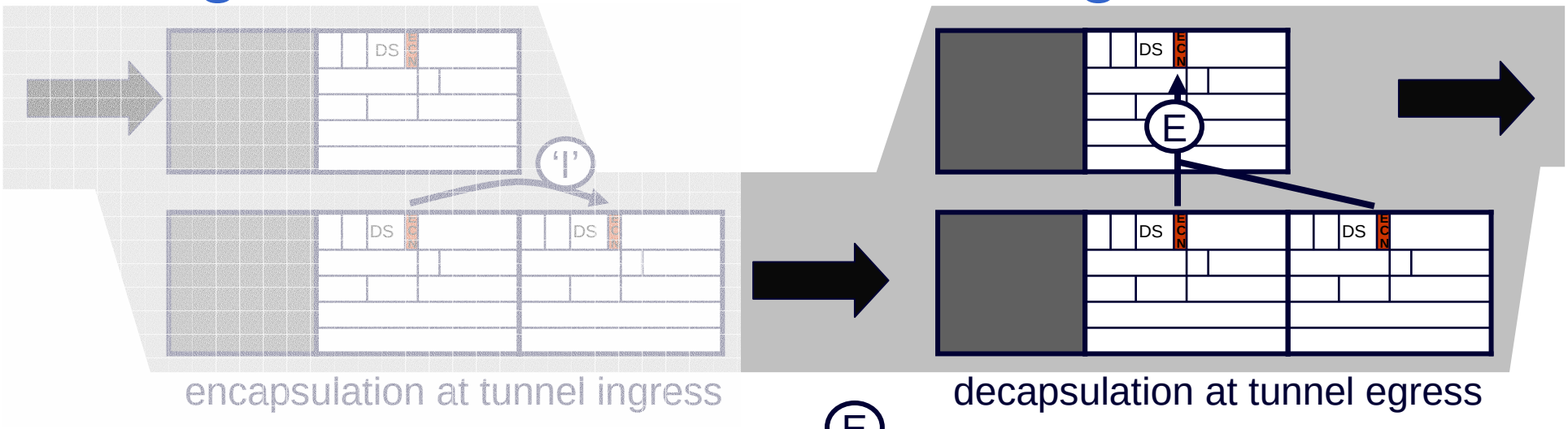
IETF-76 tsvwg Nov 2009

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
  - **revised WG draft:** draft-ietf-tsvwg-ecn-tunnel-04.txt 24 Oct '09
  - **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

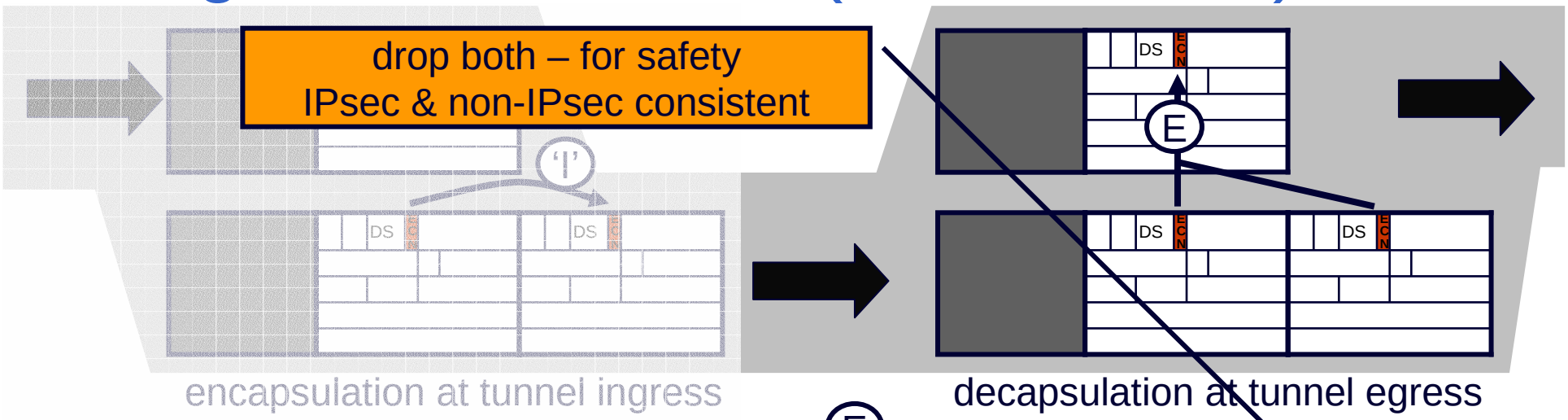


- OK for current ECN
  - 1 severity level of congestion
- any outer changes into ECT(0/1) lost
  - reason: to restrict covert channel (but 2-bit now considered manageable)
  - effectively wastes ½ bit in IP header

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)			

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)

E

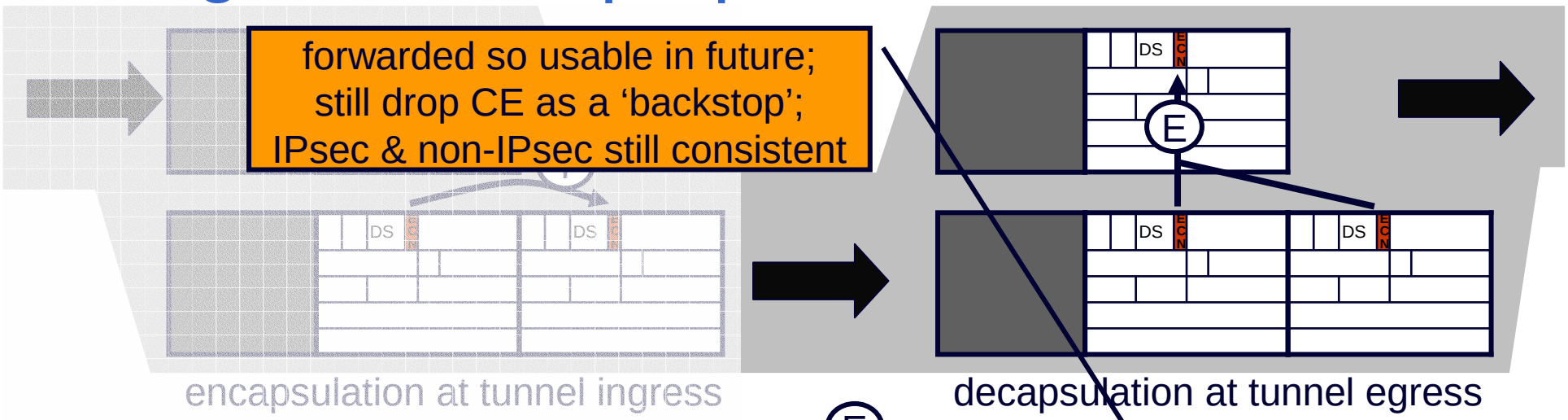
incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT (!!!)	**drop (!!!)**	**drop (!!!)**
ECT(0)	ECT(0)	ECT(0)	**ECT(1) (!)**	CE
ECT(1)	ECT(1)	ECT(1) **(!!!)**	ECT(1)	CE
CE	CE	CE	CE **(!!!)**	CE

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 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)

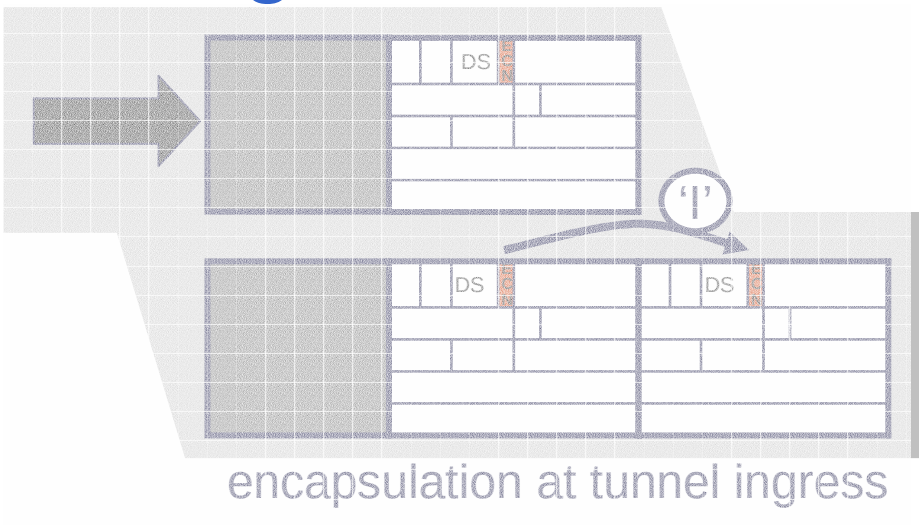
E

incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT (!!!)	Not-ECT (!!!)	**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)**			

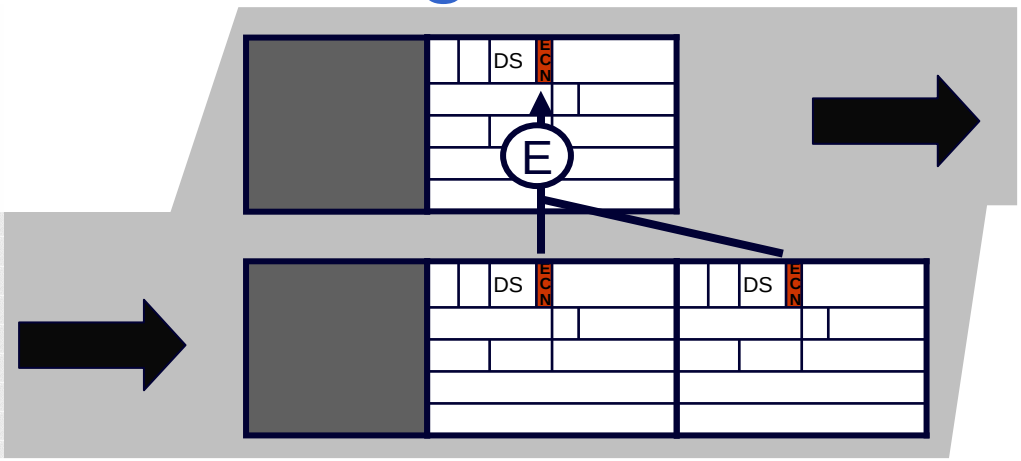
(**!!!**) = currently unused combination, egress MAY raise an alarm

PCN objected to one alarm;
other removed for consistency;
OK – not a safety alarm

egress behaviour in existing RFCs



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)			

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

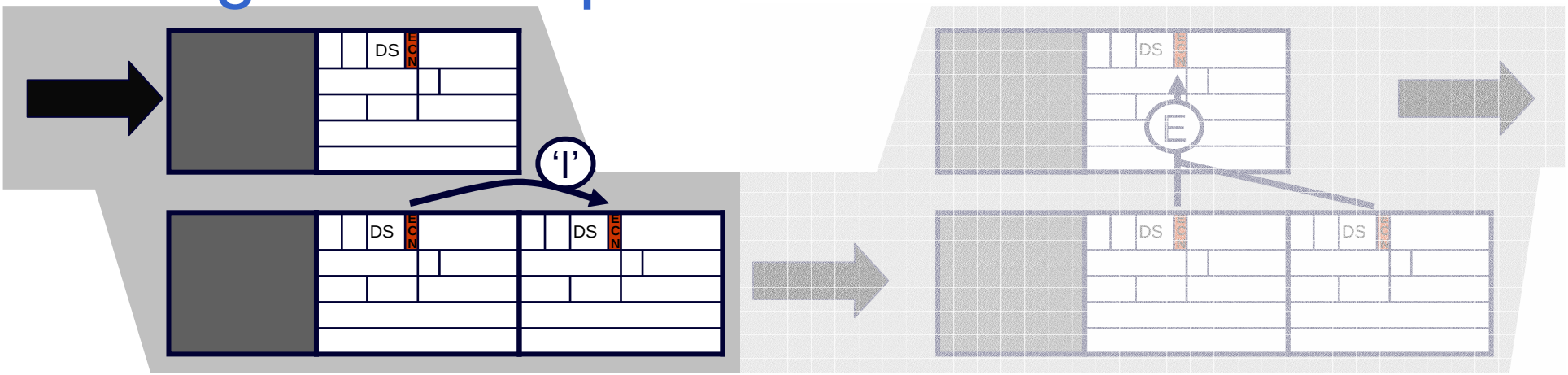
- do all decapsulators on path propagate 2 levels?
  - PCN: controlled domain: configured by operator
  - future e2e scheme: hosts can’t tell (open issue)

backward & forward compatibility

ingress		egress mode	egress action	**I-D ecn-tunnel**	RFC 4301	RFC 3168		RFC 2481		RFC 2401/ 2003
		-	-	-	-	full	lim	2481	2481 IPsec	-
			action	**calc C**	calc B	calc B	inner	calc A	inner	inner
'compre-hensive'	**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

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 mode** for legacy	**'reset' CE no longer used**	'copy' CE becomes **normal mode** for all IP in IP