Layered Encapsulation of Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-01.txt

Bob Briscoe, BT

IETF-73 pcn Nov 2008

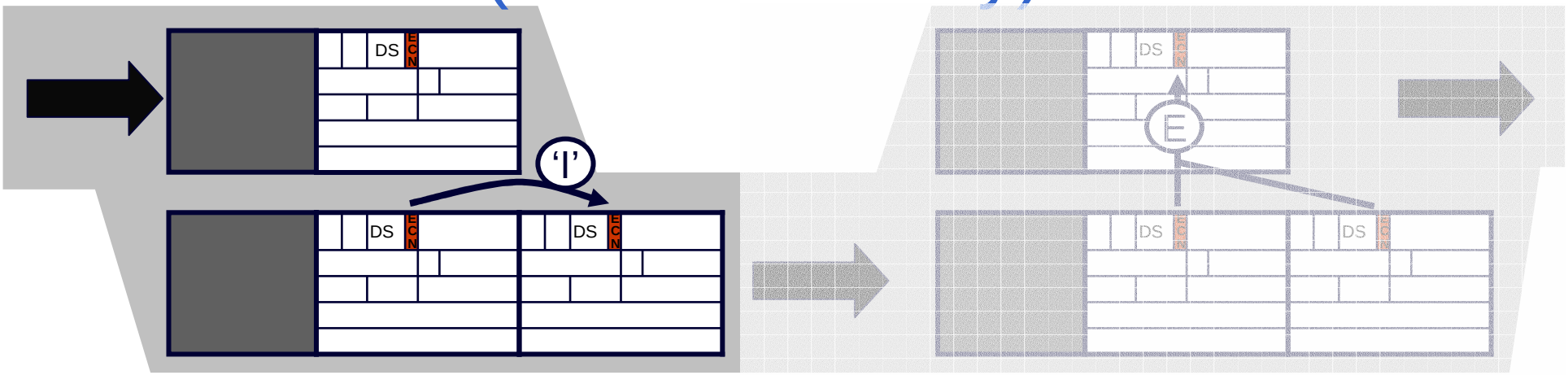
status

- Layered Encapsulation of Congestion Notification
  - **new WG draft:** draft-ietf-tsvwg-ecn-tunnel-01.txt as of late Oct'07
  - **previously:** draft-briscoe-tsvwg-ecn-tunnel-01.txt
  - **intended status:** standards track
  - **RFC pub target:** ? TBA
  - **immediate intent:** discuss including fix to decap as well as encap
    get people to sign up to review
  - **w-gs & r-gs affected:** TSVWG, PCN, ICCRG, IPsec, Internet Area?

reminder (exec summary)

- scope
  - 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
    - e.g. PCN "excess rate marking" works with 4301 but not 3168 tunnels
- bring ECN IP in IP tunnel ingress [RFC3168] into line with IPsec [RFC4301]
  - all tunnels can behave the same, revealing full congestion info
  - anyway, copying of whole ECN field is simpler
- thorough analysis of implications:
  - security, control, & management
  - guidance on specifying ECN behaviour for new links, for alternate PHBs
- ideally fix egress too (currently only 'for discussion')

reminder (exec summary)



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

text updates since IETF-72

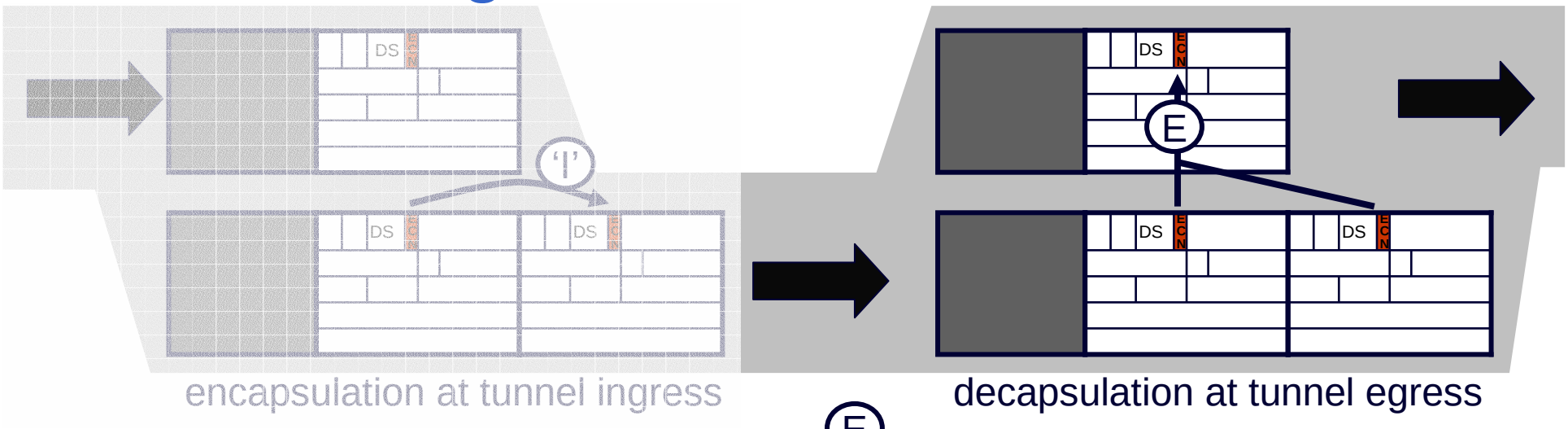
[draft-briscoe-tsvwg-ecn-tunnel-01.txt]

→ [draft-ietf-tsvwg-ecn-tunnel-00.txt]

→ [draft-ietf-tsvwg-ecn-tunnel-01.txt]

- much simpler method to monitor tunnel's contribution to congestion
  - see spare slide or Appendix B
- all significant edits concern decap – encap has stayed stable
- documented full set of illegal combinations of inner & outer at egress
  - on which egress should (optionally) raise a management alarm
- generalise egress behaviour while we're at it?
  - currently just in appendix 'for discussion' – says 'not normative'
  - problem: current egress behaviour discards changes to ECT(0) or ECT(1)
    - space for 2 congestion levels (e.g. PCN) but can't use it
    - effectively wastes half a bit of the IP header
  - now written up pros & cons of change (Appx C)
    - convinced myself this change should be in normative part of draft
    - what do you think...?

current egress behaviour



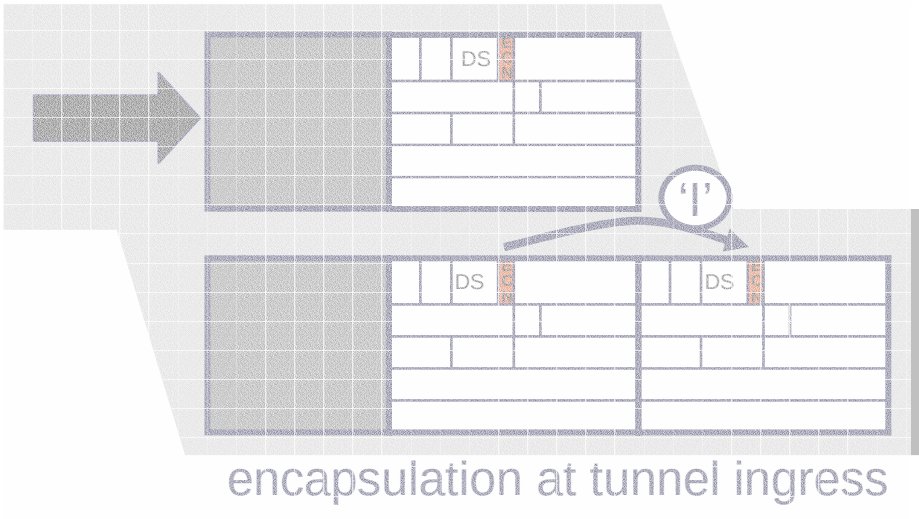
- OK for current ECN
- but any changes to ECT lost
  - effectively wastes ½ bit in IP header
  - again for safety against marginal threat that IPsec decided was manageable
- PCN tried to use ECT(0/1)
  - but having to waste DSCPs instead
  - or a limited scheme where it's arranged for the egress to already know which of ECT(0/1) the ingress originally sent

E

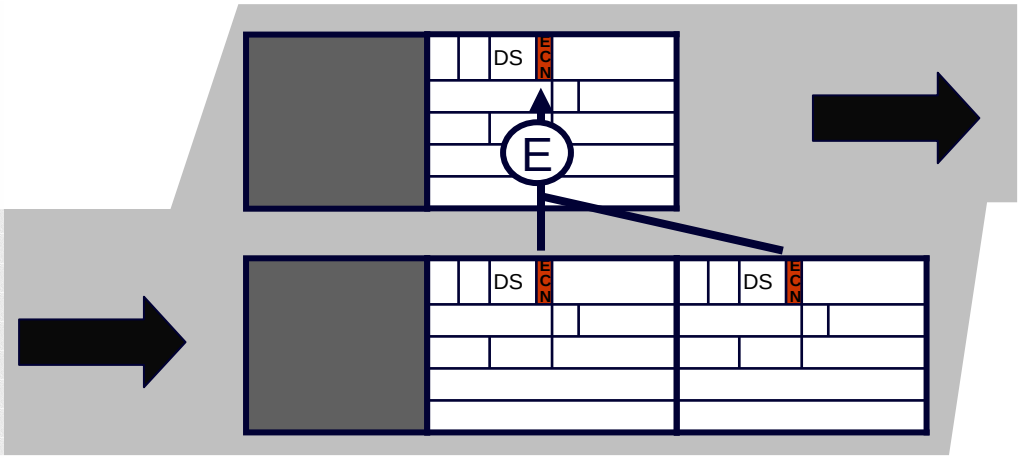
incoming inner	incoming outer			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	drop (**!!!**)	drop (**!!!**)	drop (**!!!**)
ECT(0)	ECT(0)	ECT(0)	ECT(0) (**!!!**)	CE
ECT(1)	ECT(1)	ECT(1) (**!!!**)	ECT(1)	CE
CE	CE	CE	CE (**!!!**)	CE
	Outgoing header (RFC3168 & RFC4301)			

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

'comprehensive' egress rules (only 'for discussion')



encapsulation at tunnel ingress



decapsulation at tunnel egress

- recall: proposed change to ingress
  - brings RFC3168 into line with RFC4301
- if we also changed the egress
  - it would be a new update to *both* RFCs
- but no effect on any existing tunnels
  - adds a new capability using a previously illegal combination of inner & outer
  - only tunnels that need the new capability would need to comply
  - and update, not a fork
- note well: change to egress is currently not in the normative part of this proposal
  - but documented in appendix C 'for discussion'
  - however I'll make it normative if no-one objects

E

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

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

new comprehensive decap rules

pros & cons of ways to introduce them

		within tsvwg-ecn-tunnel stds track	new pcn-tunnel-... expt track
		Disadv: may never need change	Disadv: eventually extra mode of tunnel to be compatible with
Default for all PHBs	Adv: no config as old behaviour was unusable	Recommended. Can fall back on expt track if stall	More likely to get through
For PHBs that need it	Disadv: no motivation for unused fork	reject	reject

next steps

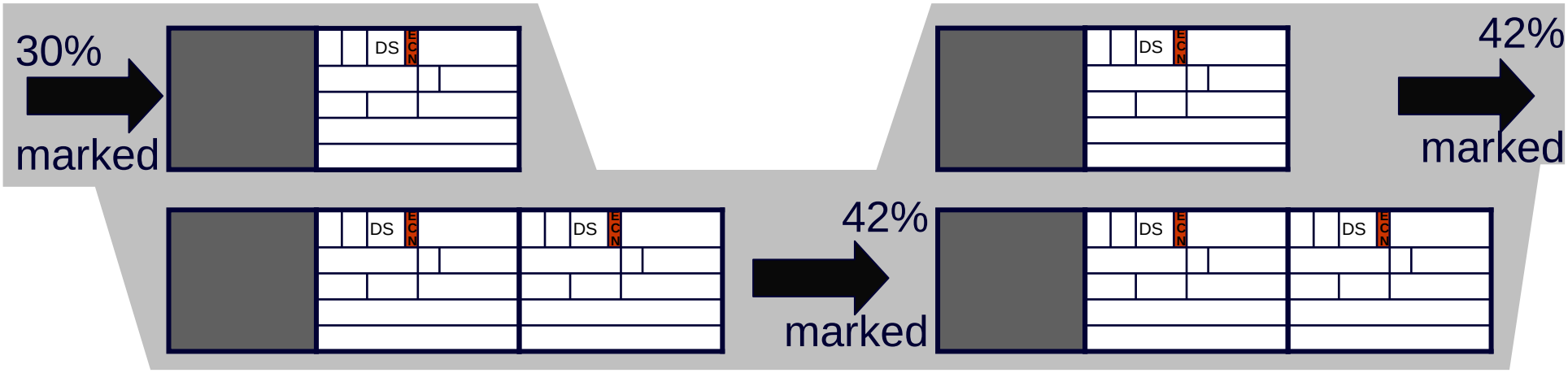
- should we change the egress at the same time?
  - tunnel stuff makes people's heads hurt
  - needs careful list discussion
  - remember, these are nuances to the behaviour of the neck of the hour-glass
  - will need to assure IPsec folks that they don't have to change (again)
  - I'll only make comprehensive egress rules normative if consensus to do so
  - I'll also add reasoning for original egress behaviour (requested in Anil Agarwal's rvw)
- plan to split out guidelines for new ECN encapsulations
  - for those adding congestion notification to alternate PHBs or to layer 2 technologies (incl. non-IETF, e.g. IEEE 802.1)
  - better in a separate (informational) I-D – just stds track IPinIP stuff in this one
  - and improve structure of this draft at same time (Michael Menth's comments)
- need people to sign up to review this draft
  - will need reviews once all the above settled

Layered Encapsulation of Congestion Notification

draft-briscoe-tsvwg-ecn-tunnel-01.txt

Q&A

contribution to congestion across tunnel

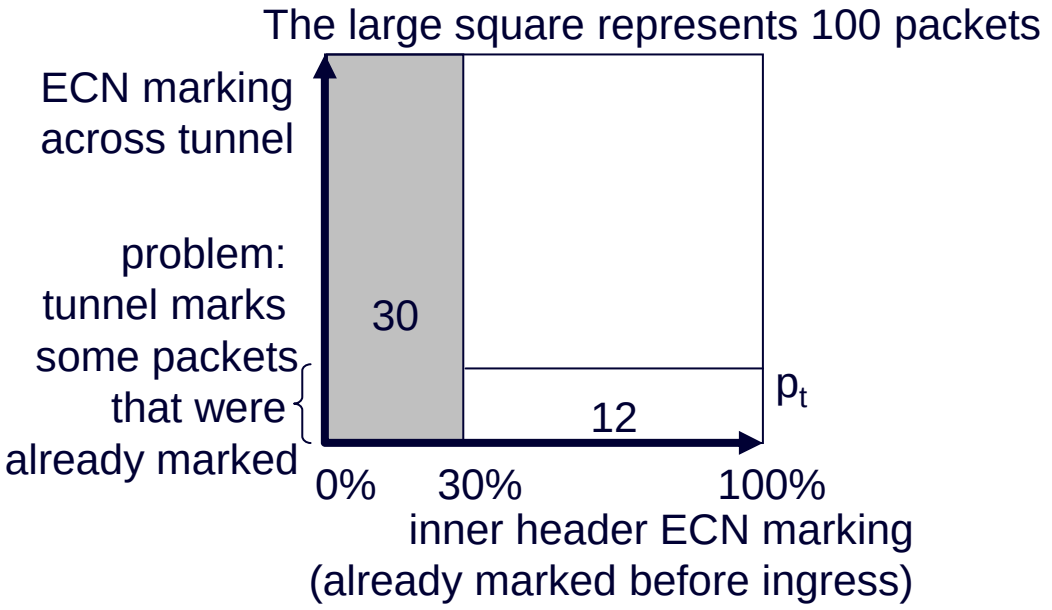


complaint:

- if CE copied at ingress, operators can't distinguish congestion added since tunnel ingress
- it's not 12%

new method in Appendix B

- it's $= {}^{12}/_{(100-30)}$
  $\approx 17\%$
- just monitor the 70 packets without the inner header marked



backward & forward compatibility

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