# Re-ECN: Adding Accountability for Causing Congestion to TCP/IP

**Bob Briscoe**, BT & UCL
Arnaud Jacquet, BT
Alessandro Salvatori, BT

CRN DoS w-g, Apr 2006


# solution statement

- re-ECN allows networks to police congestion control at network layer
  - on short and long time-scales
- if attackers can disguise their traffic perfectly
  - by evading any attempt to distinguish it from a flash crowd
  - re-ECN ***can*** ensure attackers cause no more damage than legit users
  - and persistent attackers (incl. zombies) become much less potent than legit users
- conservative networks that want to protect against attacks
  - can make their own users control congestion correctly
  - can make other networks feel the pain they allow their users to cause
    - using penalties (typically financial)
- liberal networks may choose to pay the penalties
  - rather than tightly control their own users (thus attracting the world's attackers)
  - re-ECN doesn't aim to control such attackers (it could, but not scalably)
  - just moves money from networks harbouring attackers to networks harbouring victims

# status Apr 06

- personal draft in IETF Transport Area draft-briscoe-tsvwg-re-ecn-tcp-01
  - presented twice (Oct 05 & Mar 06)
  - draft-01 fixed vulnerability found in draft-00 protocol encoding
  - interest and positive encouragement (mainly off-list)
- considerable interest from other operators
  - including ‘official’ interest channelled through BT a/c mgmt
- net neutrality solution
  - can be used to prevent apps helping themselves to QoS (or account for its use)
    - VoIP, video, p2p file-sharing
    - but not because of what they are, just by their congestion behaviour
  - getting swept into debate around US congressional committee
- looking for partner(s) to take through standards
  - IETF first, later: 3GPP, ETSI TISPAN, etc
  - ideally endpoint OS vendor/policing box vendor, but hits other buttons too

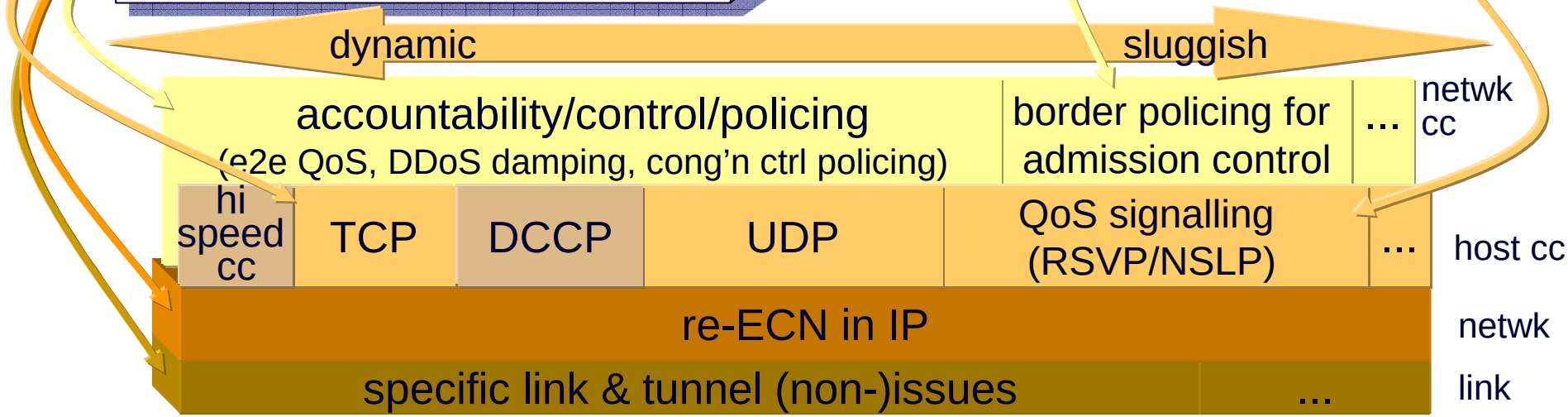
# stack positioning

Re-ECN: Adding Accountability for Causing Congestion to TCP/IP
draft-briscoe-tsvwg-re-ecn-tcp-01
*intent*

§3: overview in TCP/IP
§4: in TCP & others — *stds*
§5: in IP
§6: accountability apps — *inform'l*

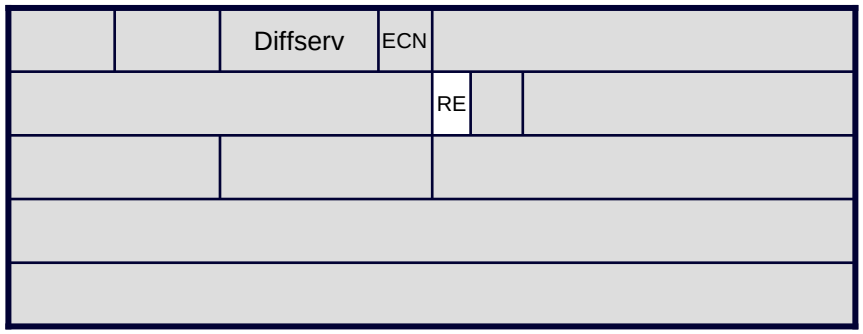
Emulating Border Flow Policing using Re-ECN on Bulk Data
draft-briscoe-tsvwg-re-ecn-border-cheat-00
*intent: informational*

RSVP Extensions
for Admission Control over Diffserv
using Pre-congestion Notification
draft-lefaucheur-rsvp-ecn-00
*intent*
adds congestion f/b to RSVP — *stds*



# re-ECN

- proposed change to IP header
  - re-ECN Extension (RE) flag
  - using last unused bit in IP header




- once flow established
- sender re-inserts ECN feedback into forward data (“re-ECN”) as follows
  - re-ECN sender usually sends `grey` packets
  - on transport layer (e.g. TCP) feedback of every red packet (network layer congestion)
    sender sends **`black`**
    else sends `grey`
- conceptually, ‘worth’ of packet as shown in matrix
- aim for zero balance of worth in flow

[RFC3168]
ECN marking router (debit)

sender (credit)

| RE / ECN | 0 | +1 | 0 |
|---|---|---|---|
| 1 | | 0 | -1 |
| | | ECT(1) 01 | CE 11 |

worth

# extended ECN codepoints: summary

- extra semantics backward compatible with previous ECN codepoint semantics

| ECN code-point | ECN [RFC3168] codepoint | RE flag | Extended ECN codepoint | re-ECN meaning | `worth' |
|---|---|---|---|---|---|
| 00 | not-ECT | 0 | Not-RECT | Not re-ECN capable transport | |
| | | 1 | FNE | Feedback not established | +1 |
| 01 | ECT(1) | 0 | Re-Echo | Re-echo congestion event | +1 |
| | | 1 | RECT | Re-ECN capable transport | 0 |
| 10 | ECT(0) | 0 | --- | 'Legacy' ECN use | |
| | | 1 | --CU-- | Currently unused | |
| 11 | CE | 0 | CE(0) | Congestion experienced with Re-Echo | 0 |
| | | 1 | CE(-1) | Congestion experienced | -1 |

# flow bootstrap

- feedback not established (**FNE**) codepoint; RE=1, ECN=00
  - sent when don't know which way to set RE flag, due to lack of feedback
  - 'worth' **+1**, so builds up credit when sent at flow start
- after idle >1sec
  next packet MUST be **green**
  - enables deterministic flow state mgmt (policers, droppers, firewalls, servers)
- **green** packets are ECN-capable
  - routers MAY ECN mark, rather than drop
  - strong condition on deployment (see draft)

- **green** also serves as state setup bit [Clark, Handley & Greenhalgh]
  - protocol-independent identification of flow state set-up
  - for servers, firewalls, tag switching, etc
  - don't create state if not set
  - may drop packet if not set but matching state not found
  - firewalls can permit protocol evolution without knowing semantics
  - some validation of encrypted traffic, independent of transport
  - can limit outgoing rate of state setup
- considering I-D [Handley & Greenhalgh]
  - state-setup codepoint independent of, but compatible with, re-ECN
- **green** is 'soft-state set-up codepoint' (idempotent), to be precise

# brief romp through re-ECN draft (65pp)

- easter egg added **: )**
- re-ECN in TCP fully spec'd (§4)
- network layer (§5)
  - OPTIONAL router forwarding changes → next slide
  - control and management section added
- accountability/policing applications (§6)
  - incentive framework
    - example ingress policers & egress dropper, pseudo-code TBA
  - DDoS mitigation explained
  - enables simpler ways to do e2e QoS, traffic eng, inter-domain SLAs
- incremental deployment (§7) → next slide but one
- architectural rationale (§8)
- security considerations (§10) → next slide but two

# incremental deployment (§7: 5½pp)

- brings together reasoning for wire protocol choices
  - during deployment period networks can throttle down goodput for legacy hosts
  - can't attack by using legacy behaviours
- deployment scenarios & incentives
  - everyone who needs to act, must have strong incentive to act
  - and incentives must arise in the order of required deployment
- main messages
  - **first step** to break ECN deployment deadlock
    - edge-edge PCN for end-to-end controlled load (CL) QoS
  - **next step:** greed and fear motivators
    - help TCP (naively friendly) against greedy (streaming) apps
    - probably vertically integrated (conservative) operators first
    - 3GPP devices leak deployment to other networks by roaming
  - unilateral deployment per network ...

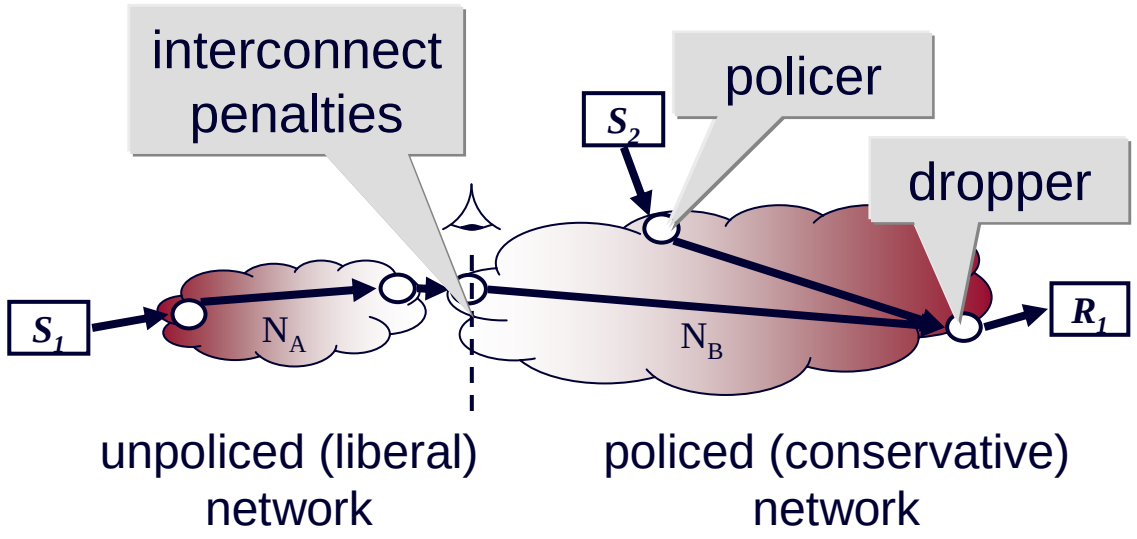
# how to allow *some* networks to police - NGN *and* Internet

- conservative networks
  - might want to throttle if unresponsive to congestion (VoIP, video, DDoS)
- middle ground
  - might want to cap congestion caused per user (e.g. 24x7 heavy sources)
- liberal networks
  - open access, no restrictions
  - evolution of hi-speed/different congestion control,... new worms

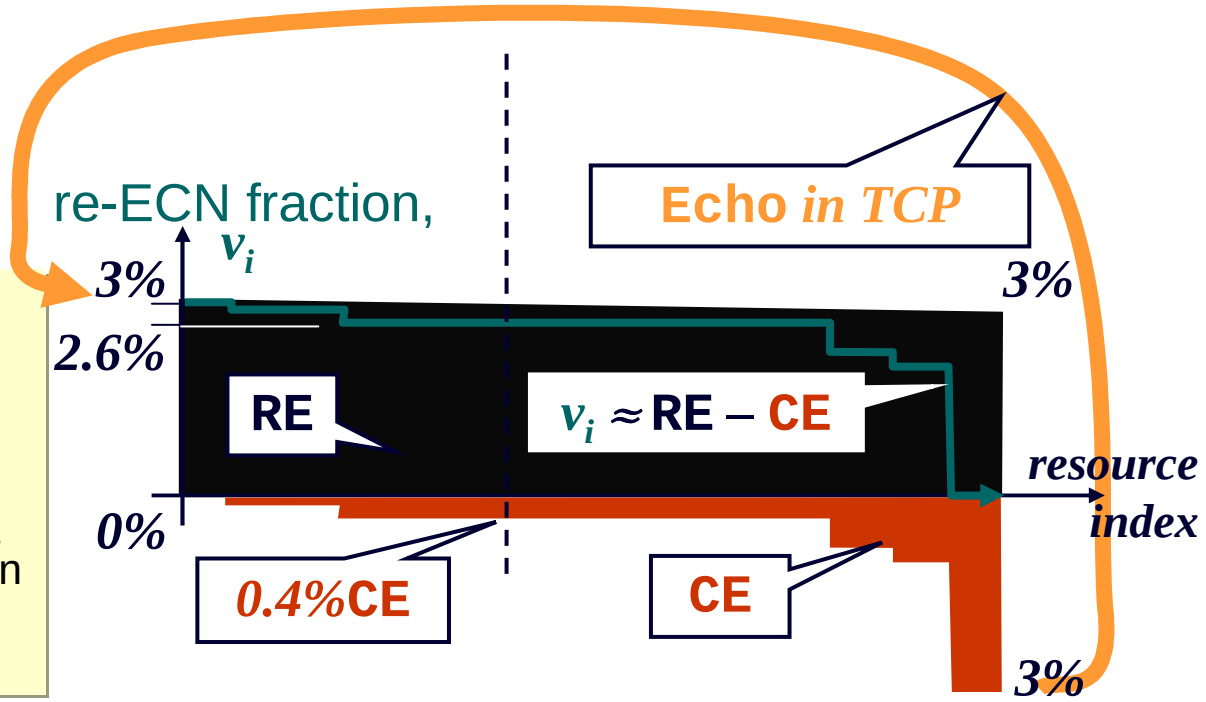
- many believe Internet is broken
  - not IETF role to pre-judge which is right answer to these socio-economic issues
  - Internet needs all these answers – balance to be determined by natural selection
  - 'do-nothing' doesn't maintain liberal status quo, we just get more walls
- re-ECN goals
  - just enough support for conservative policies without breaking 'net neutrality'
  - manage evolution of new congestion control, even for liberal → conservative flows
  - nets that allow their users to cause congestion in other nets, can be held accountable

# re-ECN partial deployment

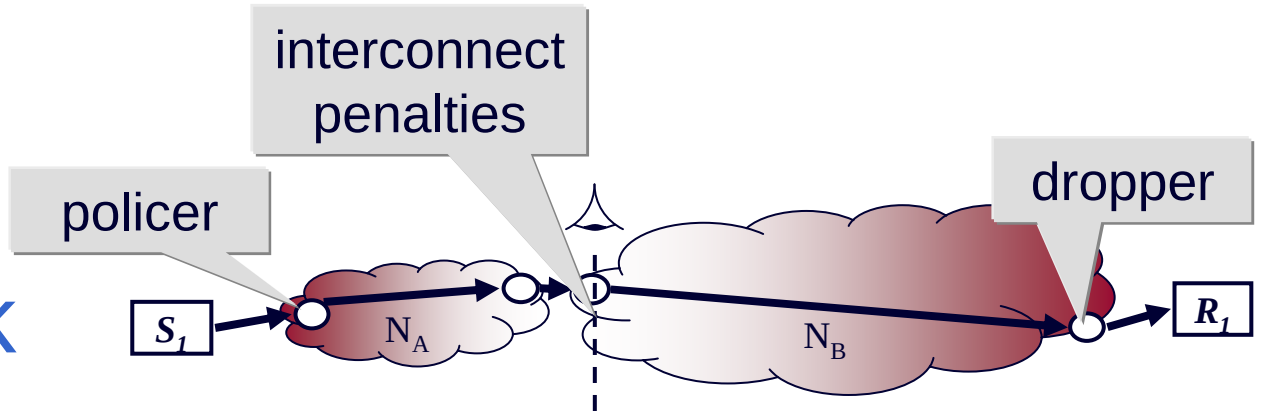
| 0 | +1 worth 0 | 0 |
|---|---|---|
| 1 | 0 | -1 |
| RE / ECN | ECT(1) 01 | CE 11 |



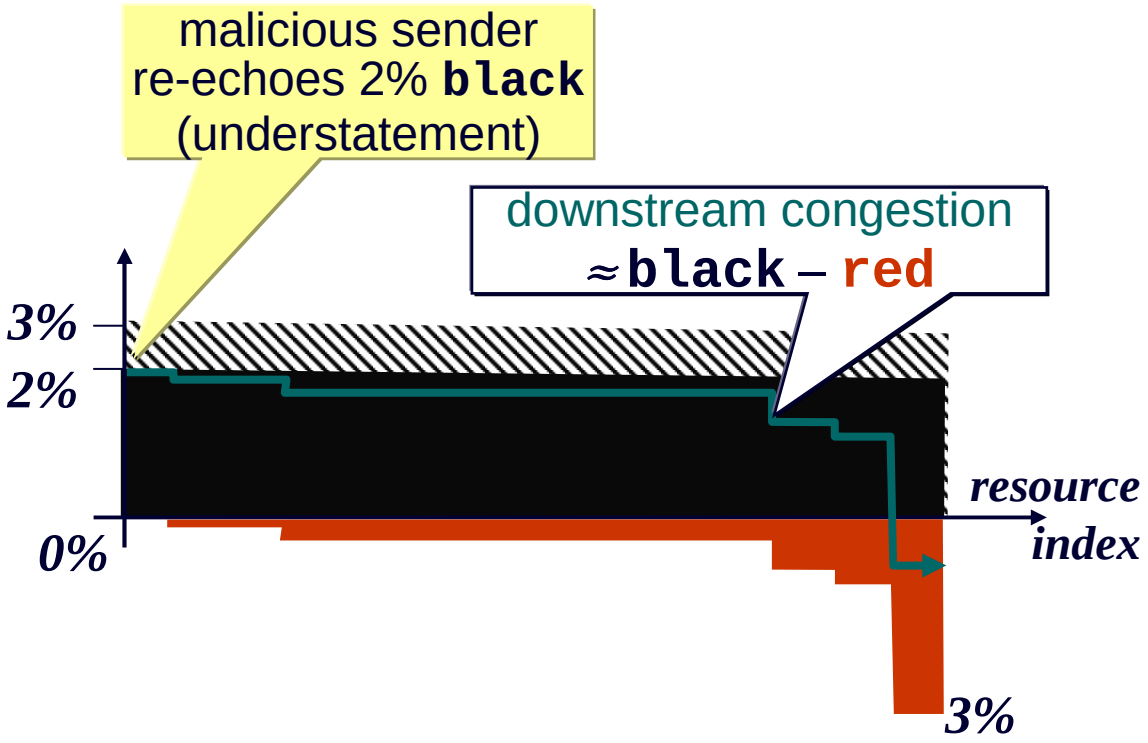

- on every **congestion event** from TCP, sender sends **`black`**, else sets **`grey`**
- at any point on path, diff betw fractions of **`black`** & **`red`** is downstream congestion
- routers unchanged



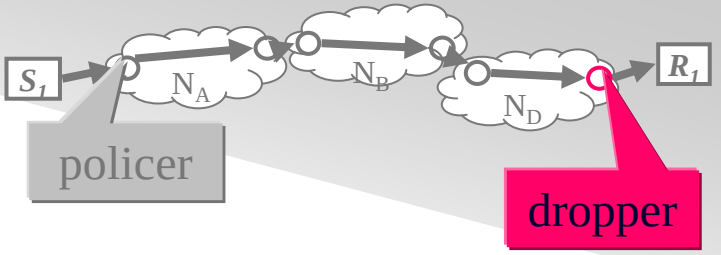
# incentive framework

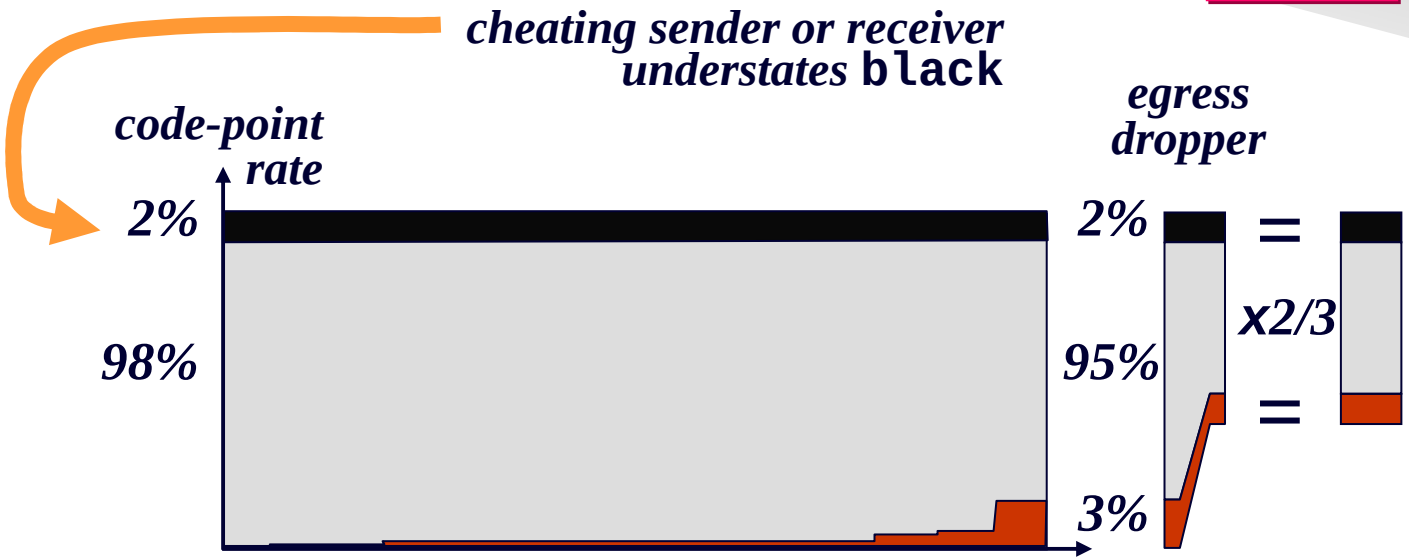



- packets carry view of downstream path congestion to each router
- using path congestion declared by sender
  - can police rate response
  - or enforce congestion quotas
- won't sender or rcvr just understate congestion?
  - egress drops negative balance (next slide )



# egress dropper (sketch)




- drop enough traffic to make fraction of **red** = **black**
  - understatement allows gain through policer, but dropper always fully cancels it out
  - goodput best if rcvr & sender honest about feedback & re-feedback
- understate congestion to attack routers?
  - given overloaded routers, honest senders will be sending nearly all **black**
  - overloaded routers preferentially drop **grey** and **red** (next slide)
- important principle: attack traffic does no harm until it congests a router
  - re-ECN drops attack at first congested router (no push-back, no new attack vector)

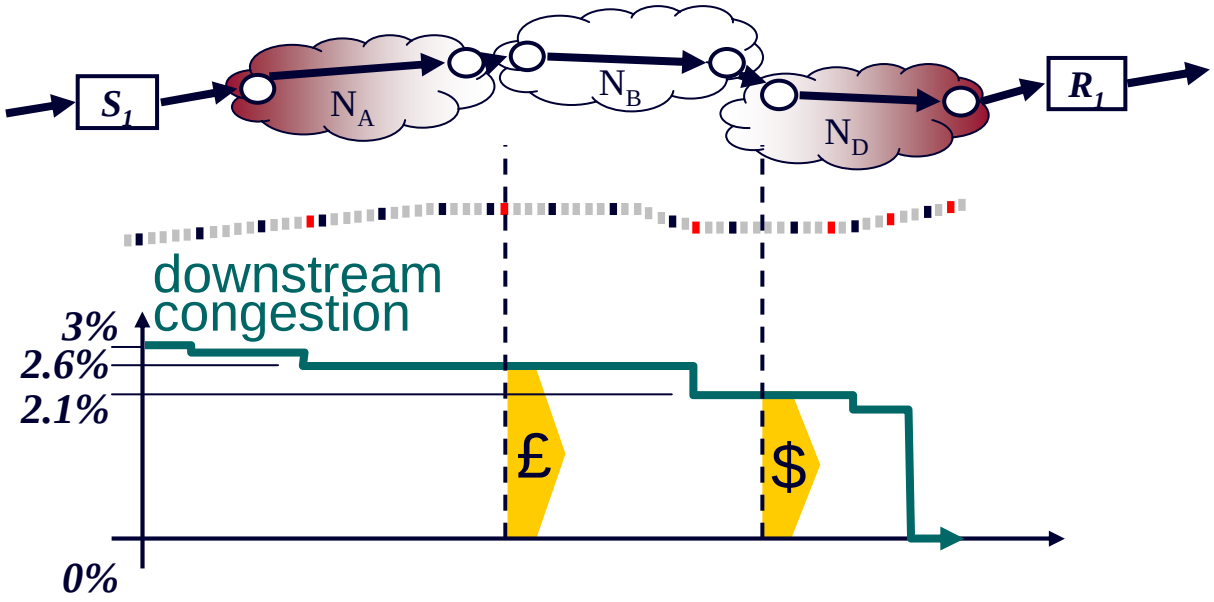
# OPTIONAL router forwarding changes

- preferential drop: improves robustness against DDoS
- **green** can be ECN marked rather than dropped (with caveat)

| ECN code-point | ECN [RFC3168] codepoint | RE flag | Extended ECN codepoint | re-ECN meaning | `worth' | pref drop (1=drop 1st) |
|---|---|---|---|---|---|---|
| 00 | not-ECT | 0 | Not-RECT | Not re-ECN capable transport | | 1 |
| | | 1 | FNE | Feedback not established | +1 | 3 |
| 01 | ECT(1) | 0 | Re-Echo | Re-echo congestion event | +1 | 3 |
| | | 1 | RECT | Re-ECN capable transport | 0 | 2 |
| 10 | ECT(0) | 0 | --- | 'Legacy' ECN use | | 1 |
| | | 1 | --CU-- | Currently unused | | 1 |
| 11 | CE | 0 | CE(0) | CE with Re-Echo | 0 | 2 |
| | | 1 | CE(-1) | Congestion experienced | -1 | 2 |

# inter-domain accountability for congestion

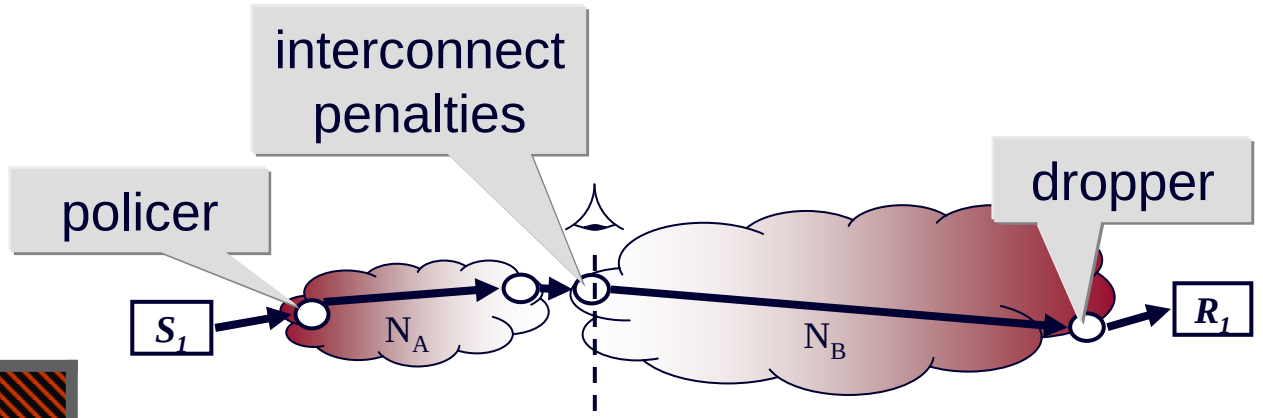
- metric for inter-domain SLAs or usage charges
  - $N_B$ applies penalty to $N_A$ in proportion to bulk volume of **`black`** less bulk volume of **`red`** over, say, a month
  - could be tiered penalties, directly proportionate usage charge, etc.
  - flows de-aggregate precisely to responsible networks

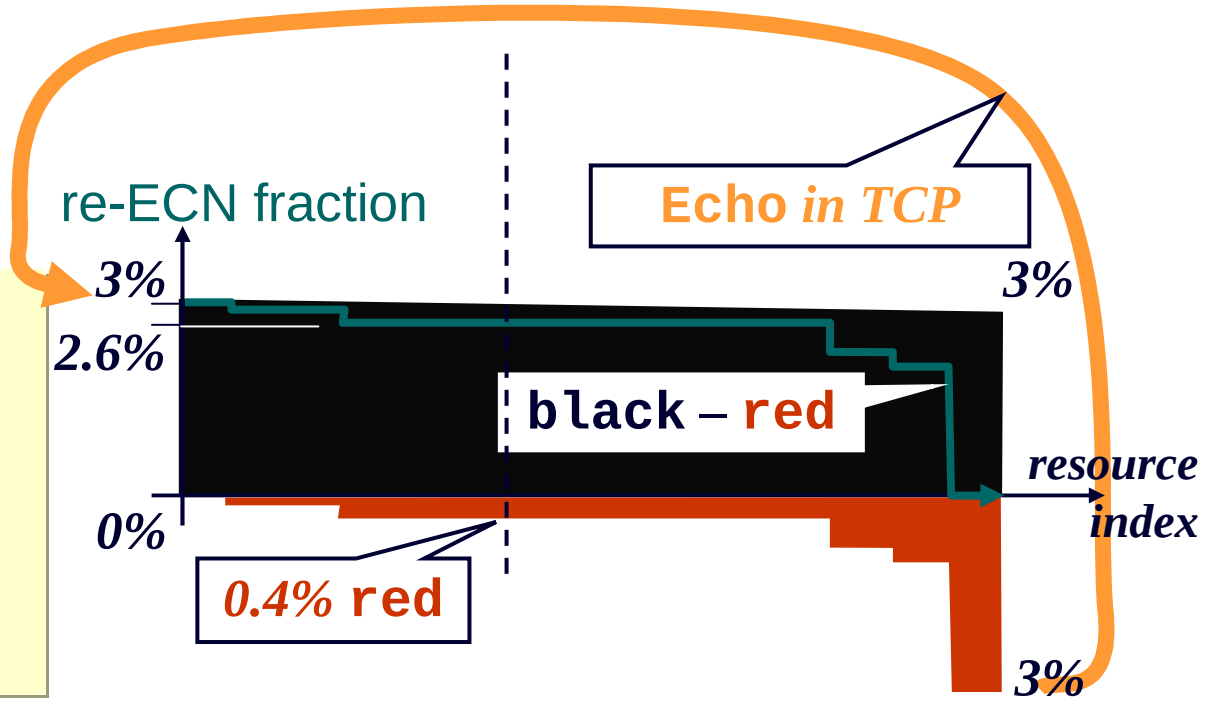


# incentive framework

| 0 | +1 | 0 |
|---|---|---|
| 1 | 0 | -1 |
| RE / ECN | ECT(1) 01 | CE 11 |

worth




- at any point on path, diff betw fractions of **black** & **red** is downstream congestion
- routers unchanged

# re-ECN security considerations (§10) and incentive framework limitations (§6.3)

- egress dropper
  - robust against attack that plays-off against ingress policing
  - robust against state exhaustion attacks (by design of **`green`**)
  - write-up of state aggregation implementation TBA
  - believe new protocol allows dropper to be robust against dynamic attacks
- collateral damage attack still possible → next slide
- re-ECN deliberately designed not to rely on crypto

# independence from identifiers

- controls congestion crossing any physical interface
  - user-network, network-network
  - congestion from network layer down to physical
  - not from a source address
- does have a dependency on source addresses
  - not to ***identify*** sources, merely to treat each flow separately
  - outstanding vulnerability
    - attacker spoofs another source's flow
    - deliberately brings down their joint average causing high drop

# re-ECN summary

- neutralises attacks indistinguishable from flash crowd
  - or bankrupts (?) networks that harbour attackers
- simple architectural fix
  - generic accountability hook per datagram
  - requires one bit in IP header
  - can separate out `feedback not est'd` flag (≡ state set-up)
- driven by big greed buttons, not just fear (DoS)
  - enables 'net neutral' policing of causes of congestion
- fixed vulnerabilities so far by making it simpler
  - working on robustness to new attacks
- detailed incremental deployment story
  - liberal networks can choose not to police, but still accountable

# Re-ECN:
# Adding Accountability for Causing Congestion to TCP/IP

draft-briscoe-tsvwg-re-ecn-tcp-01.txt

# Q&A


# previous re-ECN protocol (IP layer)

| ECN code-point | standard designation |
|---|---|
| 00 | not-ECT |
| 10 | ECT(0) |
| 01 | ECT(1) |
| 11 | CE |

- sender re-inserts congestion feedback into forward data: “re-feedback”

on every **Echo-CE** from transport (e.g. TCP)

sender sets **ECT(0)**
else sets ECT(1)

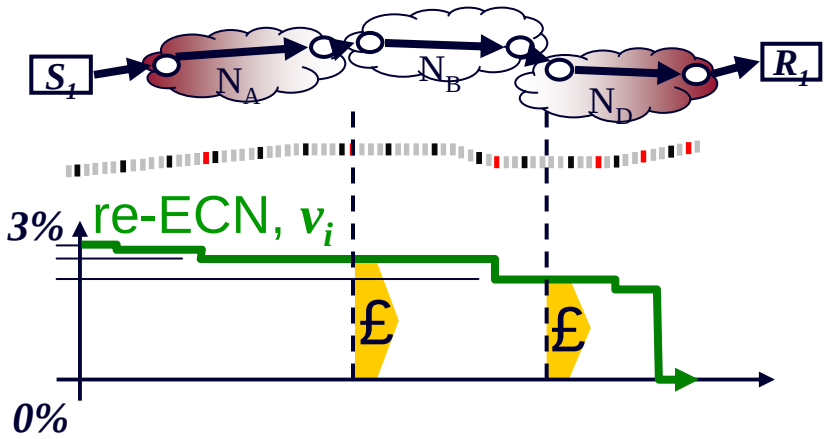
- Feedback-Established (FE) flag

| IPv4 control flags | | |
|---|---|---|
| FE | DF | MF |

accountability for congestion

# other applications

- congestion-history-based policer (congestion cap)
  - throttles causes of past heavy congestion (zombies, 24x7 p2p)
- DDoS mitigation
- QoS & DCCP profile flexibility
  - ingress can unilaterally allow different rate responses to congestion
- load sharing, traffic engineering
  - multipath routers can compare downstream congestion
- bulk metric for inter-domain SLAs or charges
  - bulk volume of **ECT(0)** less bulk volume of **CE**
  - upstream networks that do nothing about policing, DoS, zombies etc will break SLA or get charged more



# congestion competition – inter-domain routing

- if congestion → profit for a network, why not fake it?
  - upstream networks will route round more highly congested paths
  - $N_A$ can see relative costs of paths to $R_1$ thru $N_B$ & $N_C$
- the issue of monopoly paths
  - incentivise new provision
  - collusion issues require market regulation