Note Well

• Any submission to the IETF intended by the Contributor for publication as all or part of an IETF Internet-Draft or RFC and any statement made within the context of an IETF activity is considered an "IETF Contribution". Such statements include oral statements in IETF sessions, as well as written and electronic communications made at any time or place, which are addressed to:
  – the IETF plenary session,
  – any IETF working group or portion thereof,
  – the IESG, or any member thereof on behalf of the IESG,
  – the IAB or any member thereof on behalf of the IAB,
  – any IETF mailing list, including the IETF list itself, any working group or design team list, or any other list functioning under IETF auspices,
  – the RFC Editor or the Internet-Drafts function

• All IETF Contributions are subject to the rules of RFC 3978 (updated by RFC 4748) and RFC 3979. Statements made outside of an IETF session, mailing list or other function, that are clearly not intended to be input to an IETF activity, group or function, are not IETF Contributions in the context of this notice.

• Please consult RFC 3978 (and RFC 4748) for details.

• A participant in any IETF activity is deemed to accept all IETF rules of process, as documented in Best Current Practices RFCs and IESG Statements.

• A participant in any IETF activity acknowledges that written, audio and video records of meetings may be made and may be available to the public.
other administrivia

• note taker
  • notes & slides will appear with links to background at
  • <http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/>

• wireless, pls:
  • don’t have your adapter in ad hoc mode
  • cell phones to silent
agenda of unofficial BoF, 21 March 1510-1640 Karlin I, Prague Hilton

re-ECN architectural intent

- Start 15:10
  - [ 5] Administrivia
  - [30] Architectural intent of re-ECN (incl simple abstraction of how it works)
    Questions for clarification only
  - [20] Questions & Answers
  - [10] community interest?
    - IETF or IRTF?
    - How to change architecture
    - Next Steps
  - [10] break for cookies & drinks if required
  - [15 (squeezable/stretchable)] More questions & discussion

- End 16:40

- not covered in main talk, but open to questions on these
  - protocol, algorithm and implementation detail
  - conflict with ECN nonce
  - likely outcomes / implications
  - fairness, net neutrality & welfare maximisation
  - simplifying border adm ctrl in PCN
  - simplifying generalised QoS
  - flexibility for hi-speed cc, DCCP etc
  - potential for load balanced routing
  - tunnelling & layering
re-ECN $\iff$ cost fairness

- re-ECN is a low level architectural enabler (in IP)
  - designed to solve an information visibility problem
  - not a solution to fairness in itself
  - but a step to shape evolutionary change
- all the IETF needs to do is standardise a protocol like re-ECN
  - policers, customer contracts, border contracts, etc are just scenarios
  - merely what will probably happen (existence proof that protocol is robust)
- re-ECN is not limited to cost fairness, but motivated by it
  - re-ECN appendix shows how to police TCP (flow rate fairness)
  - fairness I-D shows how other forms of fairness can sit within cost fairness
- could have cost fairness with an alternative to re-ECN
  - but no other practical schemes (yet)
re-ECN architectural intent
a step to shape evolutionary change
<draft-briscoe-tsvwg-re-ecn-tcp-03.txt>

Bob Briscoe
Chief Researcher, BT Group
unofficial Birds of a Feather at IETF-68
Mar 2007
known problem since early days

- how to share all the parts of a huge, multi-provider packet multiplexer between competing processes
- keeping one-way datagrams
- allowing for
  - self-interest & malice
    - of users and of providers
  - evolvability
    - of new rate dynamics from apps
    - of new business models
  - viability of supply chain
  - simplicity

- if we do nothing
  - the few are ruining it for the many
  - massive capacity needed to keep interactive apps viable
  - poor incentives to invest in capacity
  - operators are kludging it with DPI
  - solely today’s apps frozen into net
  - complex, ugly feature interactions
solution step #1: ECN
make congestion visible to network layer

- packet drop rate is a measure of congestion
  - but how does network at receiver measure holes? how big? how many?
  - can’t presume network operator allowed any deeper into packet than its own header
  - not in other networks’ (or endpoints’) interest to report dropped packets

- solution: Explicit Congestion Notification (ECN)
  - mark packets as congestion **approaches** - to avoid drop
  - already standardised into IP (RFC3168 – 2001)
  - implemented by most router vendors – very lightweight mechanism
  - but rarely turned on by operators (yet) – mexican stand-off with OS vendors
new information visibility problem
ECN is not enough

• path congestion only measurable at exit

• can’t measure path congestion at entry
  – can’t presume allowed deeper into feedback packets
re-ECN in brief

- reinsert feedback
- packets arrive at each router predicting downstream path
- incremental deployment + upgrade incentive knob
- hangs new capabilities on ECN deployment, not just performance
- a simple idea for the Internet’s accountability architecture

![Diagram showing the before and after states of re-ECN]
measurable downstream congestion

solution step #2

- sender re-inserts feedback by marking packets **black**
- at any point on path, diff between fractions of **black** & **red** bytes is downstream congestion
- ECN routers unchanged
- **black** marking e2e but visible at net layer for accountability
flow bootstrap ‘pre-feedback’

• at least one green packet(s) at start of flow or after >1sec idle
  • means “feedback not established”
  • ‘credit’ for safety due to lack of feedback
  • a green byte is ‘worth’ same as a black byte

• lots of powerful uses for a different colour from black
  • distinguishes conservatism from expected congestion based on experience
  • ability to vary the expected cost of jump-starting (research needed)
  • gives deterministic flow state mgmt (policers, droppers, firewalls, servers)
proposed re-ECN service model

- to encourage sender (or proxy) to indicate sufficient expected congestion...
- Internet won’t try to deliver packet flows beyond the point where more congestion has been experienced than expected
  - if sender wants to communicate, has to reveal expected congestion
  - even if sender not trying to communicate (e.g. DoS) packets can be dropped rather than enqueued before they add to congestion
Policing Congestion using Re-ECN

non-interactive long flows (e.g. P2P, ftp)

interactive short flows (e.g. Web, IM)

100% capacity

Flow Bandwidth

Time

Animation requires Office XP or equivalent
congestion policer – one example: per-user policer

solution step #4

two different customers, same deal

non-interactive long flows (e.g. P2P, ftp)

interactive short flows (e.g. Web, IM)

congestion volume allowance

overdraft
- 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)
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

![Diagram of network showing inter-domain accountability for congestion]

Legend:
- downstream congestion marking [%]
- bit rate
- highly congested link

Legend for usage charges:
- flat (e.g. monthly) charges

Legend for currency:
- ¥
- £
- $
re-ECN partial deployment

interconnect penalties

re-ECN fraction

feedback

unpoliced (liberal) network

policed (conservative) network

resource index

0.4% red

red

2.6%

3%

N_A

S

N_B

R

S

1

S

1
deployment incentives
bootstrap then chain reaction

- deployment effectively involves architectural change
  1. (minor) change to sender’s Internet stack
  2. network deploys edge/border incentive functions
     - breaking the stand-off between 1 & 2 requires strong incentives
- re-feedback solves ISPs’ main cost control problem
  - third party services competing with ISP pay below network cost
  - ISP has to compete *while* paying balance of competitor’s costs
  - hits big fear button and big greed button
  - but keeps moral high ground
     - net neutral: managing congestion not app discrimination
- first movers: vertically integrated cellular operators?
  - 3GPP devices leak deployment to other networks by roaming
- 2nd movers (NGNs?) continue chain reaction
  - adopters’ incoming border charges focus on non-adopters
outstanding issues

• technical
  ✗ a lot more verification of all the claims to do
  ✗ community found a few nasty vulnerabilities over last two years
    ✓ fixed (added minor complexity in only one case)
  ✗ connection spoofing attack still outstanding
    ✓ possible solution recently brainstormed

• religious
  ✗ underlying problem has been dogma that equal flow rates are fair
    ✓ groundswell change in community thinking since mid Oct‘06
    ✗ dismantling a religion not so easy

• community
  ✗ a lot of passive support
  but consensus needs a lot more active interest

  a change to IP needs to be ‘owned’ by Internet community
  please take it, break it, analyse it, re-design it, work out implications
conclusions

• resolution of tensions in fairness / net neutrality debate
  • freedom to use the Internet, until you congest freedom of others
  • proportionate restriction of freedom during congestion

• an architectural change with grand implications
  • simple management and control of fairness & QoS
  • naturally mitigates DDoS
  • generates correct capacity investment incentives and signals

• but conceptually simple and trivial to implement

• strong deployment incentives
  • bootstrap and onward chain reaction

• where’s the catch?
  • invite you to analyse it, break it, re-design it
Re-ECN: Adding Accountability for Causing Congestion to TCP/IP

draft-briscoe-tsvwg-re-ecn-tcp-03

intent

§3: overview in TCP/IP
§4: in TCP & other transports stds
§5: in IP (v4 & v6)
§6: accountability apps inform’l

Emulating Border Flow Policing using Re-ECN on Bulk Data
draft-briscoe-tsvwg-re-ecn-border-cheat-02

intent: informational

•more papers (PCN, QoS, DDoS etc): <http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/>

accountability/control/policing (e2e QoS, DDoS damping, cong’n ctrl policing)

border policing for admission control

QoS signalling (RSVP/NSLP)

re-ECN in IP

specific link & tunnel (non-)issues

TCP | SCTP | DCCP | UDP | QoS signalling (RSVP/NSLP) | ...

high speed cc | border cc | network cc | network link
re-ECN architectural intent

more motivating problems
more architectural motivation
• (non)issues with layering & tunnelling
• bottleneck policing harmful
• independence from identifiers
mechanism
• IPv4 & v6 wire protocol
• drop preference semantics
• conflict with the ECN nonce

uses
• simplifying generalised QoS
• flexibility for hi-speed cc, DCCP etc
• adding re-ECN to various transports: TCP, SCTP, DCCP, PCN, UDP
• DDoS mitigation
• potential for load balanced routing

incentives and security
• attacks on re-ECN and fixes
• IPR

Q&A
next steps

• build community
• simulations, implementation continues
• Official IETF BoF?
• IRTF Internet Congestion Control research group?
designed for tussle

- current Internet gives freedom but no fairness
  - the more you take, the more you get; the more polite you are, the less you get
  - but we don’t want to lose freedom by enforcing fairness

- solution: allow ISPs to enforce user-specific congestion control fairness

  liberal acceptable use policies
  - open access, no restrictions

  middle ground
  - might want to cap congestion caused per user (e.g. 24x7 heavy p2p sources, DDoS)
  - evolution of different congestion control (e.g. hi-dynamics; rate adaptive VoIP, video)

  conservative acceptable use policies
  - might want to throttle if unresponsive to congestion (VoIP, video, DDoS)

- engineers shouldn’t pre-judge 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 middlebox kludges

- re-ECN at network layer: goals
  - just enough support for conservative policies without breaking ‘net neutrality’
  - nets that allow their users to cause congestion in other nets can be held accountable
designed for tussle
Internet needs all these answers – market selection finds balance

demand side – freedom to degrade others

- the Internet is all about the freedom to get what I want (within my line rate)
  limited by how much I impinge on the freedom of others
  - enforceable congestion control

freedom within fairness

- differentiated quality of service
you’ll get what you contract to have
- you’ll get what we infer you want given what you’re doing

supply side – freedom to degrade competitors

re-ECN allows extremes but doesn’t help them and provides handles for the market to make it very hard for them
summary

• Internet needs to be able to discriminate
  • against bits limiting the freedom of others – bits causing congestion
  • then wouldn’t need to discriminate against apps causing congestion

• operators can choose not to limit their users’ freedoms
  • but they take responsibility for congestion their users cause in other nets

• if operators do discriminate against apps
  • customers need enough choices to be able to switch operators
  • or apps can often obfuscate themselves anyway

• these economic effects require change to the Internet Protocol
  • making IP more suitable as the basis of a converged architecture
freedom how Internet sharing ‘works’

• those who push most, get most
  • restraint: the other ingredient of early Internet success
  • reliant on voluntary politeness of endpoint algorithms (TCP)
  • a game of chicken – taking all and holding your ground pays
  • or starting more ‘TCP-fair’ flows than anyone else
  • or for much longer than anyone else (p2p file-sharing)

(browsers x4, bitTorrent x50)

(VoIP, video streaming)
congestion cap auto-adjusts
volume cap always a hard compromise

No cap or loose volume cap  Tight volume cap  Congestion allowance

High capacity

Low capacity
capacity growth will prevent congestion?

Distribution of customers’ daily traffic into & out of a Japanese ISP (Feb 2005)

(5GB/day equivalent to 0.46Mbps if continuous)

Changing technology shares of Japanese access market

100Mbps fibre to the home (FTTH 46.4%)

digital subscriber line (DSL 53.6%)

(9%, 2.5GB)
(4%, 5GB)

Courtesy of Kenjiro Cho et al
The Impact and Implications of the Growth in Residential User-to-User Traffic, SIGCOMM’06
hang on! solution step #0: what’s congestion got to do with the problem?

- can’t solve a sharing problem without sharing costs
  - congestion is the cost of usage
  - cost of your behaviour on others

- NOTE WELL
  - IETF needs to provide the cost metric
  - don’t need metric for value – leave that for industry to guess

- it’s not what you get
  - it’s what you unsuccessfully tried to get
    - the bits each you contributed to excess load
    - loss/marking fraction \( p(t) \) times your bit rate \( x_i(t) \)
    - only need dimensionless loss fraction \( p(t) \) in wire protocol (ECN)
    - it subtly communicates your excess rate, because your own rate \( x_i(t) \) is visible

- excess bits accumulate simply and correctly
  - over time, over flows and over network paths
  - congestion volume = bits of dropped/marked data you sent

\[
\text{congestion or loss (marking) fraction [\%] } \quad p(t) \equiv \frac{\text{excess load}}{\text{offered load}}
\]
calibrating ‘cost to other users’

- a monetary value can be put on ‘what you unsuccessfully tried to get’
  - the marginal cost of upgrading network equipment
    - so it wouldn’t have marked the volume it did
    - so your behaviour wouldn’t have affected others
- competitive market matches...
  - the cost of congestion volume
  - with the cost of alleviating it

- **congestion volume is not an extra cost**
  - part of the flat charge we already pay
  - but we can’t measure who to blame for what
  - if we could, we *might* see pricing like this...

<table>
<thead>
<tr>
<th>access link</th>
<th>congestion volume allow’ce</th>
<th>charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>100Mbps</td>
<td>50MB/month</td>
<td>€15/month</td>
</tr>
<tr>
<td>100Mbps</td>
<td>100MB/month</td>
<td>€20/month</td>
</tr>
</tbody>
</table>

- **NOTE WELL**
  - IETF provides the metric
  - industry does the business models
**re-feedback summary**

- reinsert feedback to align path characterisations at receiver
- packets arrive at each router predicting downstream path
- arranged for dominant strategy of all parties to be honesty
- incremental deployment + upgrade incentive knob
- hangs new capabilities on ECN deployment, not just performance
- a simple idea for the Internet’s accountability architecture

- democratises path information
  - either network or source can control (control requires timely information)
  - designed for tussle: preserves e2e principle, but endpoint control optional
(non-)issues with layering & tunnels

• general non-issue
  • RE flag shouldn’t change once set by sender (or proxy)
  • policers merely read RE to compare with CE introduced so far
  • OK as long as CE represents congestion since same origin that set RE

• IP in IP tunnels
  • OK if tunnel entry copies RE and CE to outer header
  • but full functionality RFC3168 ECN tunnel resets CE in outer header
    – RFC3168 only said reset because security folks thought copy might leak info
    – concern has been resolved – updated IPSec RFC4301 (Dec 05) copies ECN at ingress
    – RFC3168 tunnelling section needs updating to reflect later security thinking and practice

• IP payload encryption (e.g. IPSec ESP)
  • non-issue – re-ECN designed to work only in network layer header
  • flow-ID obfuscation also non-issue – re-ECN only uses flow ID uniqueness, if at all

• layer 2 congestion notification (ATM, Frame, ... MPLS, 802.3ar)
  • non-issue given IP layer should accumulate CE from each ‘L2 network’ into ECN
bottleneck policing harmful to evolvability
...and bypass-able anyway

- bottleneck policers: active research area since 1999
  - detect misbehaving flows causing ‘unfair’ share of congestion
  - located at each potentially congested routers
  - what right have these policers to assume a specific congestion response for a flow?
    - if they could police accurately, new congestion control evolution would require
      per-flow authorisation from all policers on the path (cf. IntServ)
  - malicious sources can bypass them by splitting flow IDs
    - even splitting flow across multiple intermediate hosts (or src address spoofing)

- re-ECN policing
  - polices congestion caused by all sources behind a physical interface, irrespective of addressing
  - within that, can also choose to police per-flow, per flow setup, per-destination etc.
  - evolution of new behaviours by bilateral agreement with first ingress, if at all
  - dropper uses flow IDs, but no advantage to split IDs
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
extended ECN codepoints: summary

- extra semantics backward compatible with previous ECN codepoint semantics

<table>
<thead>
<tr>
<th>ECN code-point</th>
<th>ECN [RFC3168] codepoint</th>
<th>RE flag</th>
<th>Extended ECN codepoint</th>
<th>re-ECN meaning</th>
<th><code>worth</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>not-ECT</td>
<td>0</td>
<td>Not-RECT</td>
<td>Not re-ECN capable transport</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>FNE</td>
<td>Feedback not established</td>
<td>+1</td>
</tr>
<tr>
<td>01</td>
<td>ECT(1)</td>
<td>0</td>
<td>Re-Echo</td>
<td>Re-echo congestion event (ECN nonce conflict)</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>RECT</td>
<td>Re-ECN capable transport</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>ECT(0)</td>
<td>0</td>
<td>---</td>
<td>‘Legacy’ ECN use</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>--CU--</td>
<td>Currently unused</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CE</td>
<td>0</td>
<td>CE(0)</td>
<td>Congestion experienced with Re-Echo</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>CE(-1)</td>
<td>Congestion experienced</td>
<td>-1</td>
</tr>
</tbody>
</table>
**re-ECN wire protocol in IPv4 (§3)**

- propose Re-ECN Extension (RE) flag
  - for IPv4: propose to use bit 48 (was reserved)
  - set by sender, unchanged e2e

- once flow established
- sender re-inserts ECN feedback into forward data ("re-ECN") as follows
  - re-ECN sender always sets ECT(1)
  - on every congestion event from transport (e.g. TCP)

<table>
<thead>
<tr>
<th>DiffServ</th>
<th>ECN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- conceptually, ‘worth’ of packet depends on 3 bit ‘codepoint’
- aim for zero balance of worth in flow

---

37
IPv6 re-ECN protocol encoding

- IPv6 hop-by-hop options header extension
  - new Congestion hop-by-hop option type
    - action if unrecognized (AIU) = 00 ‘skip and continue’
    - changeable (C) flag = 1 ‘may change en route’
      - even tho RE flag shouldn’t change en route (AH would just tell attackers which packets not to attack)
  - seems wasteful for 1 bit, but we plan:
    - future hi-speed congestion control I-D using multi-bit congestion field
    - other congestion-related fields possible
      - e.g. to distinguish wireless loss and per-packet vs per-bit congestion
OPTIONAL router forwarding changes

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

<table>
<thead>
<tr>
<th>ECN code-point</th>
<th>ECN [RFC3168] codepoint</th>
<th>RE flag</th>
<th>Extended ECN codepoint</th>
<th>re-ECN meaning</th>
<th><code>worth</code></th>
<th>pref drop (1=drop 1st)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>not-ECT</td>
<td>0</td>
<td>Not-RECT</td>
<td>Not re-ECN capable transport</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>FNE</td>
<td>Feedback not established</td>
<td>+1</td>
<td>3</td>
</tr>
<tr>
<td>01</td>
<td>ECT (1)</td>
<td>0</td>
<td>Re-Echo</td>
<td>Re-echo congestion event</td>
<td>+1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>RECT</td>
<td>Re-ECN capable transport</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>ECT (0)</td>
<td>0</td>
<td>---</td>
<td>‘Legacy’ ECN use</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>--CU--</td>
<td>Currently unused</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CE</td>
<td>0</td>
<td>CE (0)</td>
<td>CE with Re-Echo</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>CE (-1)</td>
<td>Congestion experienced</td>
<td>-1</td>
<td>2</td>
</tr>
</tbody>
</table>
new appendix “Argument for holding back the ECN nonce” (§AI)

ECN nonce usefulness

• re-ECN and a transport layer nonce defend against wide range of attacks
  – ECN nonce defends against a small subset
  – and only one outside re-ECN’s range (*)
    • a sender that uses network ECN to allocate its own resources, can limit a lying receiver
    • sender can contain this attack without nonce

• IP header bits used to do this:
  – ECN nonce $\frac{1}{4}b$ (leaving last bit)
  – re-ECN $\frac{3}{8}b$ (using last bit)

• one common codepoint
  – re-ECN negotiates its use, but ECN nonce doesn’t

• propose to hold back ECN nonce
  – to see if we can find a coding to do both
  – to see if we can prevent (*) another way
  – develop a transport layer solution
flow bootstrap

- at least one **green** packet(s) at start of flow or after >1 sec idle
  - means “feedback not established”
  - ‘credit’ for safety due to lack of feedback
  - a **green** byte is ‘worth’ same as a **black** byte
- a different colour from black
  - distinguishes expected congestion based on experience from based on conservatism
  - gives deterministic flow state mgmt (policers, droppers, firewalls, servers)
  - rate limiting of state set-up
  - congestion control of memory exhaustion

**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

- to be precise **green** is ‘idempotent soft-state set-up codepoint’
guidelines for adding re-ECN to other transports

• main focus of <draft-briscoe-tsvwg-re-ecn-tcp-03>
  – IP (§5)
  – TCP (§4.1)
• added very brief sections giving guidelines for
  – DCCP (§4.2.3)
  – SCTP (§4.2.4)
    – spec would have to be a new I-D in each case
• focus of <draft-briscoe-tsvwg-re-ecn-border-cheat-01>
  – RSVP/NSIS transports (‘re-PCN’)
  – proposed technique to extend PCN-based admission control
    • Internet wide (edge-edge) – many untrusting domains
• our current focus
  – controlling fairness between current transports & hi-speed congestion control
border anti-cheating solution

3% Re-Echo into PCN data

3% CE feedback (RSVP)
### re-PCN

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>+1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>RE</td>
<td></td>
<td>ECT(1)</td>
<td>CE</td>
</tr>
<tr>
<td>ECN</td>
<td>01</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

- Ingress gateway blanks **RE**, in the same proportion as the fraction of **CE** arriving at egress.
- At any point on the path, bulk difference between fractions of **RE** and **CE** is downstream congestion.
- Routers are unchanged.

---

**3% Re-Echo (black) into data**

**3% Congestion Level Estimate in RSVP extension**

**bulk marking monitor**

**downstream congestion**

**resource index**
solution rationale

- <0.01% packet marking at typical load
  - addition of any flow makes little difference to marking
- penalties to ingress of each flow appear proportionate to its bit rate
  - emulates border flow rate policing
- as load approaches capacity
  - penalties become unbearably high (~1000x typical)
  - insensitive to exact configuration of admission threshold
  - emulates border admission control
- neither is a perfect emulation
  - but should lead to the desired behaviour
  - fail-safes if networks behave irrationally (e.g. config errors) – see draft
differential quality of service (QoS) control without all the complicated stuff

- QoS only relevant when there’s a risk of congestion
- enforcing congestion control is equivalent to QoS (and to paying for it)
  - allowing one app’s rate to slow down less than others in response to incipient congestion (ie. still low delay)
  - is equivalent to giving scheduling priority on routers*
- even if user pays a flat monthly fee
  - better QoS for some apps leaves less congestion ‘quota’ for rest
- purely by local (sender↔ingress) arrangement
  - no authorisation on any other network elements (equal marking)
- other networks reimbursed automagically
  - by inter-domain congestion pricing (SLA model also possible)
- incredible simplification of mechanisms for QoS control & mgmt
  - and, unlike other QoS mechanisms
  - it also prevents users ‘stealing’ QoS at everyone else’s expense

* except within a round trip time – implies two priority classes would be sufficient
(can also determine relative congestion marking rates of each class using economics)
incentive framework

downstream path congestion

Index

bulk congestion charging

flat fees not shown (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)

malicious sender re-echoes 2% black (understatement)

downstream congestion ≈ black - red

resource index

3%

2%

0%
aggregation
internalisation of externalities

area = instantaneous downstream congestion

legend
downstream congestion marking [%]

bit rate

large step implies highly congested link

total area = aggregate downstream congestion

metering per month: bulk volume black – red
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
  - as long as competitive physical layer (access regulation), no problem in network layer

![Diagram](image-url)
DDoS mitigation
just managing (extreme) congestion control

- two differences from congestion control
  - malice, not self-interest
    - sender doesn’t care about goodput
      1. need droppers sampling for negative flows at borders
      2. pushes beyond incipient congestion into heavy loss
  - need preferential drop on routers

- provides incentives to deploy complementary DDoS solutions
distributed denial of service (DDoS) attack
strategy #1

Web Server

BOT Agent

BOT Agent

BOT Agent

Web Client

animation requires Office XP or equivalent
per-user congestion policer
DDoS attack strategy #1

BOT agent attack traffic

interactive short flows (e.g. Web, IM)

animation requires Office XP or equivalent
distributed denial of service (DDoS) attack
strategy #2

animation requires Office XP or equivalent
attacks on re-ECN & fixes

• recap: why two codepoints worth 0?
  • when no congestion send neutral (0)
  • packet marked ‘cancelled’ if network happens to mark a packet (-1) which the sender used to re-echo congestion (+1); +1 – 1 = 0
  • in draft 00, congestion marking of +1 packet turned it to -1 not 0, but networks could cheat by focusing marking on +1 (see §B)

• but now can’t attacker just send cancelled packets?
  • immune from congestion marking
  • simple fix: policer counts cancelled with +1 towards path congestion
    – should have specified this anyway, as both represent path congestion
    – also check proportion of cancelled to +1 packets same as -1 to neutral

• set of attacks using persistently negative dummy traffic flows
  • see next presentation for border policing fix

• one remaining known vulnerability if attacker can spoof another flow ID
  • known since early on – plan to focus effort on fixing this next
dummy traffic attacks on re-ECN

- sanctions against persistently negative flows may not discourage dummy traffic

- various attacks ([Salvatori, Bauer] see draft), eg.
  - a network sends negative dummy traffic with just enough TTL to cross border [Salvatori]
    - offsets penalties from other positive traffic

- fix is to estimate contribution from negative flows crossing border by sampling
  - inflate penalties accordingly – removes attack motivations
  - see draft for details and example algorithm in appendix
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
load balanced routing support?

- automate inter-domain traffic engineering (damped)?
- validate route adverts?
  - re-balances info asymmetry
BT IPR related to draft-briscoe-.tsvwg-re-ecn-tcp-00.txt

- See IPR declaration at https://datatracker.ietf.org/public/ipr_detail_show.cgi?&ipr_id=651 which overrides this slide if there is any conflict

1) WO 2005/096566 30 Mar 2004 published
2) WO 2005/096567 30 Mar 2004 published
3) PCT/GB 2005/001737 07 May 2004 published
4) GB 0501945.0 (EP 05355137.1) 31 Jan 2005 published
5) GB 0502483.1 (EP 05255164.5) 07 Feb 2005 published

- BT hereby grants a royalty-free licence under any patent claims contained in the patent(s) or patent application(s) disclosed above that would necessarily be infringed by implementation of the technology required by the relevant IETF specification ("Necessary Patent Claims") for the purpose of implementing such specification or for making, using, selling, distributing or otherwise lawfully dealing in products or services that include an implementation of such specification provided that any party wishing to be licensed under BT’s patent claims grants a licence on reciprocal terms under its own Necessary Patent Claims.
more info...

- Fixing mindset on fairness
  - Flow Rate Fairness: Dismantling a Religion IETF Internet draft (Mar 2007)
- Overall re-feedback idea, intention, policing, QoS, load balancing etc
  - Policing Congestion Response in an Inter-Network Using Re-Feedback (SIGCOMM’05 – mechanism outdated)
- Protocol Spec and rationale
  - Re-ECN: Adding Accountability for Causing Congestion to TCP/IP IETF Internet Draft (Oct 2006)
- Using re-ECN with pre-congestion notification (PCN)
  - Emulating Border Flow Policing using Re-ECN on Bulk Data IETF Internet draft (Jun 2006)
- Relation between re-ECN and inelastic QoS
  - Commercial Models for IP Quality of Service Interconnect BT Technology Journal (Apr 2005)
- Mitigating DDoS with re-ECN
  - Using Self-interest to Prevent Malice: Fixing the Denial of Service Flaw of the Internet Workshop on the Economics of Securing the Information Infrastructure (Oct 2006)
- more related papers and all the above:
  <http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/>