Internet resource sharing: a way forward?

Bob Briscoe
Chief Researcher, BT
May 2009

This work is partly funded by Trilogy, a research project supported by the European Community
www.trilogy-project.org
fair capacity sharing – a huge responsibility

- getting this right will open a new chapter of Internet innovation
  - freedom for a huge variety of source behaviours
  - so much more than the TCP-friendly monoculture
  - rate response to congestion still important, but not the basis of capacity sharing

- getting it wrong leaves ISPs no choice but to close off the future
  - TCP/IP suite wasn't designed for ISPs to even see congestion
  - without visibility of correct metric, ISPs resort to app analysis
  - getting impossible to deploy a new use of the Internet
  - must negotiate the arbitrary blocks and throttles en route

- grudging acceptance of proverb: "good fences make good neighbours"
  - not natural for most of us to design fences
  - but lacking a good fence design, the industry is building bad ones
    - cf. lack of an IETF/IRTF firewall architecture
    - goal: a building block for fences that doesn't encourage fence-mentality
design team's top level research agenda?

• statement of ultimate target
  • metrics & deprecated metrics
  • structure & deprecated structure
  • enduring concepts

• standards agenda
  • weighted congestion controls
  • ECN gaps
  • re-ECN

• deployment scenarios
  • unilateral
  • co-ordinated
statement of ultimate target

- metrics
  - congestion-volume
    - volume of marked bits
      - $\equiv \sum_i \int p(t)x_i(t) \, dt$
    - $\neq$ volume
      - $\equiv \sum_i \int x_i(t) \, dt$
  - congestion-bit-rate
    - rate of lost / marked bits
      - $\equiv \sum_i p(t)x_i(t)$
    - $\neq$ aggr. bit-rate
      - $\equiv \sum_i x_i(t)$
- deprecated metrics
  - hi-speed flows competing with low is perfectly ok
  - relative flow sizes at a resource not relevant to fairness
  - blocking exceptionally high flow rates becomes a sin
- competition with legacy
  - s/equal windows within an order of magnitude
    - /avoid legacy flow starvation & ratchet down effects/
  - shift from relative rates to sufficient absolute legacy rate
"deprecated"?

- "design for tussle" doesn't mean no design principles
  - setting architectural direction is important
  - blessing or deprecating interim steps is important too
  - as long as everyone's interests can be satisfied

- per-flow bit-rate policing ≠ per-user bit-rate policing
  - ultimately share access networks by congestion-bit-rate
  - until then, per-user rate policing closes off nothing
    - just as if a shared link were multiple separate links
  - but per-flow rate policing closes off a lot of future flexibility
    - and it's unnecessary to satisfy anyone's interests
bottleneck policers: active research area since 1999
- detect flows causing unequal share of congestion
- located at each potentially congested router
- takes no account of how active a source is over time
- nor how many other routers the user is congesting
- based on cheap pseudonyms (flow IDs)

re-ECN / ECN
- reveals congestion caused in all Internet resources by all sources (or all sinks) behind a physical interface, irrespective of addressing
- accumulates over time
- no advantage to split IDs
- like counting volume, but ‘congestion-volume’
- focus of fairness moves from flows to packets
Initial results measured on Naples Uni network feeding numerous residential networks.

Each point is a user correlation coefficient: 0.43

WARNING: Requires validation with more sample data.
Acceptable Use Policy
'congestion-volume' allowance: 1GB/month
@ £15/month
Allows ~70GB per day of data in typical conditions

• incentive to avoid congestion
• simple invisible QoS mechanism
  • apps that need more, just go faster
• side-effect: stops denial of service
• only throttles traffic when your contribution to congestion in the cloud exceeds your allowance

...but it can't
• the Internet wasn't designed this way
• path congestion only visible to end-points, not to network
enduring concepts, but nuanced

• random congestion signals (drops or marks) from undifferentiated FIFO queues
  • marks preferred – network can't measure whole-path drop
  • holy grail if feasible – new cc with old AQM?
  • has to work well enough, optimisation can be piecemeal
• end point congestion control (rate response)
  • with weights added
    & network encourages weights to be set sparingly
design team's top level research agenda?

- statement of ultimate target
  - metrics & deprecated metrics
  - structure & deprecated structure
  - enduring concepts a basis for consensus?

- standards agenda
  - weighted congestion controls
  - ECN gaps
  - re-ECN

- deployment scenarios
  - unilateral
  - co-ordinated
Standards agenda
weighted congestion controls

1. TCP
- Light usage can go much faster
- Hardly affects completion time of heavy usage

2. WFQ

3. Volume cap

4. DPI

NOTE: Weighted sharing doesn’t imply differentiated network service
- Just weighted aggressiveness of end-system's rate response to congestion
- LEDBAT: a fixed example
standards agenda
weighted congestion controls

- toy models
  - don't fret over numbers
  - p: loss/marking fraction (log scale)
- weighted w-Relentless TCP \(w={1/25}\)
  - on every mark/loss \(W = 25\)
  - just FIFO queues
- Reno gets 'enough' over range
  - would hardly do better alone
  - if it's not enough, upgrade
Reno vs. w-Relentless
no less flow starvation than TCP-friendly

<table>
<thead>
<tr>
<th>usage type</th>
<th>no. of users</th>
<th>activity factor</th>
<th>ave. simul flows /user</th>
<th>TCP bit rate /user</th>
<th>vol/day (16hr) /user</th>
<th>traffic intensity /user</th>
</tr>
</thead>
<tbody>
<tr>
<td>attended</td>
<td>80</td>
<td>5%</td>
<td></td>
<td>417kbps</td>
<td>150MB</td>
<td>21kbps</td>
</tr>
<tr>
<td>unattended</td>
<td>20</td>
<td>100%</td>
<td></td>
<td>417kbps</td>
<td>3000MB</td>
<td>417kbps</td>
</tr>
</tbody>
</table>
standards agenda
weighted congestion controls

• important to enable $w < I$, negates weight inflation
• add weight to all(?) new congestion controls
  • LEDBAT, mTCP, SCTP, Relentless ...
• new app parameter overloading socket API
  • also app & policy integration
• timing relative to ability to police is tricky
  • change to IP will take much longer than new cc algos
  • perhaps have weighting in cc algo, but hard-code a value without an API until later
standards agenda

ECN gaps

• turn it on
  • hosts (particularly servers) should be on-by-default
  • performance delta wasn't sufficient motivation for ISPs
  • monitoring ECN for traffic control could motivate them
• ECN in L2 technologies
  • esp IEEE802 (being drafted)
• ECN in various transports
  • RTP/RTCP [RTP-ECN, RTCP-ECN]
  • ...
standards agenda
re-ECN

- source reveals congestion to net in IP header
- work to get to standards track
  - re-ECN in IPv6
  - re-ECN in IPv4 (experimental)
    - in controlled environments (e.g. GENI slice)
  - re-ECN in various transports
  - tunnelling IPv6 re-ECN in IPv4?

- the work that will take longest ought to finish first
  - Transport Area, Network Area, Security Area, etc.
  - should we take a punt before agreeing the way forward
    - Congestion Transparency (re-ECN) BoF in Stockholm?
design team's top level research agenda

- statement of ultimate target
  - metrics & deprecated metrics
  - structure & deprecated structure
  - enduring concepts
- standards agenda
  - weighted congestion controls
  - ECN gaps
  - re-ECN
- deployment scenarios
  - unilateral
  - co-ordinated
unilateral deployment scenarios
(non-TCP-friendly, ECN, re-ECN)

• no congestion transparency (not in protocols)
  • operator uses local congestion-volume metric in place of volume (e.g. on traffic control boxes)
  • end-host acts as if congestion-volume is limited
  • appears as voluntary as TCP, but unlikely to happen?
  • cf. BitTorrent, Microsoft & LEDBAT

• congestion transparency
  • re-ECN sender proxy
deployment scenarios
(non-TCP-friendly, ECN, re-ECN)

• academic networks and hi-speed data transfer
  • start with no policing & just conservatively weighted cc?
  • require IPv6 to have congestion policing framework?
  • sufficient proof of concept to move v4 from experimental?
  • remove of ad hoc controls when add congestion policing

• cellular networks
  • terminals & networks standardised monolithically
  • operators motivated to police heavy users [re-ECN06, re-ECN09]
  • mobile devices cross-fertilise fixed networks
  • requires radio resource control to trigger L3 ECN [Siris03]

• co-ordination
  • top-down: Global Information Infrastructure Commission (GIIC) & Internet Governance Forum (IGF)
    • as a way to distinguish net neutral behaviour from not
  • bottom-up: MIT interconnection w-g

• sticking points are bound to appear under each one
guaranteed bit-rate? or much faster 99.9% of the time? harnessing flexibility

- the idea that humans want to buy a known fixed bit-rate
  - comes from the needs of media delivery technology
  - hardly ever a human need or desire
- services want freedom & flexibility
  - access to a large shared pool, not a pipe
  - when freedoms collide, congestion results
  - many services can adapt to congestion
  - shift around resource pool in time/space

---

**Constant Bit Rate** 100%
**Constant Quality** 125%
**Equitable Quality** 216%

sequences encoded at same average of 500kb/s

% figures = no. of videos that fit into the same capacity

[Crabtree09]
bringing information to the control point

- no control without information
  - re-ECN packets reveal real-time cost
- flat fee policer was just one example...
- huge space for business &
  technical innovation at the control point
  - cost based, value-cost based
  - bulk, per flow, per session
  - call admission control
  - policing, charging
  - tiers, continuous
  - wholesale, retail

- truly converged architecture
  - can apply different industry cultures
  - through policies at the control point
  - not embedded in each technology
a design team needs a name

- some potential keywords
  - Internet
  - resource/capacity sharing
  - beyond TCP-friendly
  - fair
  - congestion
Re-architecting the Internet:
The Trilogy project <www.trilogy-project.org>
re-ECN & re-feedback project page:
http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/
These slides <www.cs.ucl.ac.uk/staff/B.Briscoe/present.html>
bob.briscoe@bt.com
deployment incentives
[re-ECN06] Using Self-interest to Prevent Malice; Fixing the Denial of Service Flaw of the Internet, Bob Briscoe (BT & UCL), The Workshop on the Economics of Securing the Information Infrastructure (Oct 2006)
[re-ECN] <draft-briscoe-tsvwg-re-ecn-tcp>
[re-ECN09] <draft-briscoe-tsvwg-re-ecn-tcp-motivation>
ECN @ L2
ECN @ L4-7
[RTP-ECN] draft-carlberg-avt-rtp-ecn
[RTCP-ECN] draft-carlberg-avt-rtcp-xr-ecn

ECN @ L2
ECN @ L4-7
[RTP-ECN] draft-carlberg-avt-rtp-ecn
[RTCP-ECN] draft-carlberg-avt-rtcp-xr-ecn
Internet resource sharing: a way forward?

discuss...
main steps to deploy re-feedback / re-ECN

• network
  • turn on explicit congestion notification in data forwarding
    – already standardised in IP & MPLS
    – standards required for meshed network technologies at layer 2
      (ECN in IP sufficient for point to point links)
  • deploy simple active policing functions at customer interfaces
    around participating networks
  • passive metering functions at inter-domain borders
• terminal devices
  • (minor) addition to TCP/IP stack of sending device
  • or sender proxy in network
• then new phase of Internet evolution can start
  • customer contracts & interconnect contracts
  • endpoint applications and transports
• requires update to the IP standard (v4 & v6)
  • started process in Autumn 2005
  • using last available bit in IPv4 header or IPv6 extension header
one bit opens up the future
standard ECN (explicit congestion notification) + re-inserted feedback (re-feedback) = re-ECN

1. Congested queue debit marks some packets
2. Receiver feeds back debit marks
3. Sender re-inserts feedback (re-feedback) into the forward data flow as credit marks
4. Outcome:
   End-points still do congestion control
   But sender has to reveal congestion it will cause
   Then networks can limit excessive congestion
5. Cheaters will be persistently in debt
   So network can discard their packets
   (In this diagram no-one is cheating)

no changes required to IP data forwarding