Resolving Internet capacity sharing and neutrality battles

Bob Briscoe
Chief Researcher, BT

May 2010

This work is partly funded by Trilogy, a research project supported by the European Community
www.trilogy-project.org


simplicity ahead!
can it be QoS?
on exit
check mirrors
– it was QoS

how to share Internet capacity

known problem since early Internet

- tremendous idea
  - anyone can use any link anywhere on the Internet without asking, as much as they like
- when freedoms collide
  - what share does each party get?

- solution keeps
  - one-way datagrams
  - same API
- allows for
  - self-interest & malice
    - of users and of providers
  - evolvability
    - of new app behaviours
    - 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 deep packet inspection
  - solely today's apps frozen into net
  - complex, ugly feature interactions

Internet topology visualization produced by Walrus (Courtesy of Young Hyun, CAIDA)

moving mountains IETF

glossary
IETF Internet Engineering Task Force
IESG Internet Engineering Steering Group
IAB Internet Architecture Board
IRTF Internet Research Task Force

- since 2006 IETF support for TCP capacity sharing has collapsed to zero
  - agree TCP dynamics correct, but sharing goal wrong [Briscoe07]
    - many thought leaders support our new direction – not universally – yet!
  - rewrite of IETF capacity sharing architecture in process
    - IETF delegated process to IRTF design team – eventually IAB
- Oct'09 – May'10
  - formation of IETF working group: "congestion exposure" (ConEx)
  - contentious: requires addition to IP (v4 & v6)
  - IESG now ready to ratify, but not giving up last bit in IPv4 (yet!)
  - >40 offers of significant help on list; individuals from
    - Microsoft, Nokia, Cisco, Huawei, Alcatel-Lucent, NEC, Ericsson, NSN, Sandvine, Comcast, Verizon, …

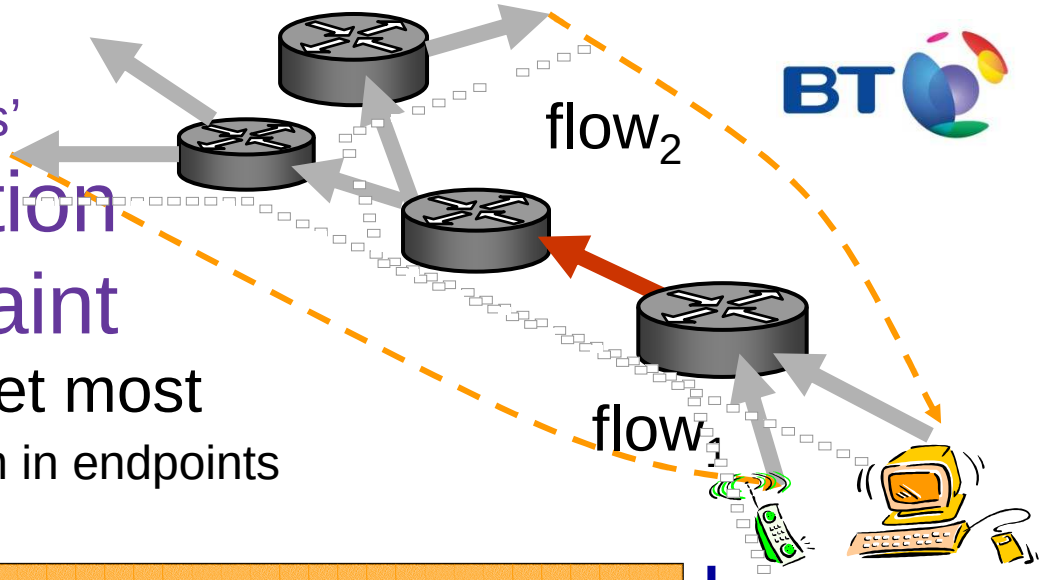
moving mountains ptII
the global ICT industry

- giic.org : ~50 CxOs of the major global ICT corporations
  - particularly interested in neutrality of solution
  - Apr ‘09: then BT CTO proposed GIIC endorses BT solution
  - Sep ’09: expert review: public policy, commercial & technical
  - Jan ‘10: GIIC published favourable assessment report

- technical media coverage (ZDnet, PCWorld, Guardian, c’t, …)
  - prompts near-universally reasonable reader postings
    - on broadband speed, quality, pricing, net neutrality!

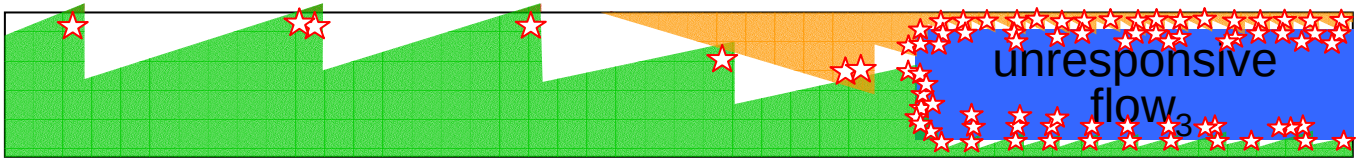
how Internet sharing 'works'

endemic congestion & voluntary restraint



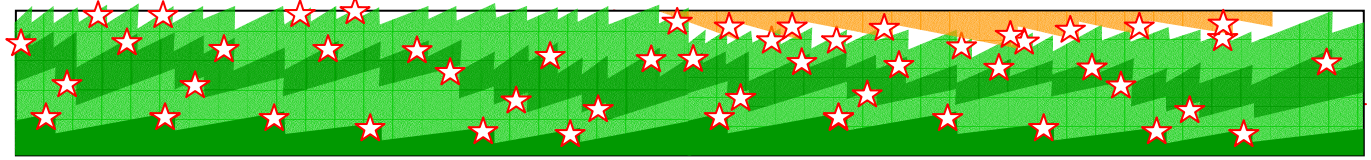
- those who take most, get most
  - voluntarily polite algorithm in endpoints
  - 'TCP-friendliness':

  - a game of chicken – taking all and holding your ground pays



(VoIP, VoD Joost 700kbps)

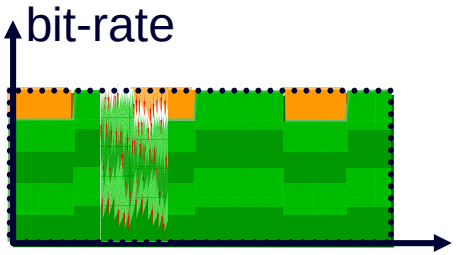
  - or start more 'TCP-friendly' flows than anyone else **(Web: x2, p2p: x5-100)**



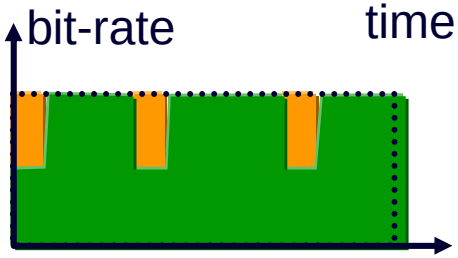
  - or transfer more bytes for longer than anyone else (file transfer x200)
  - net effect of both (p2p: x1,000-20,000 higher traffic intensity)

no traditional sharing approaches harness end-system flexibility… over time

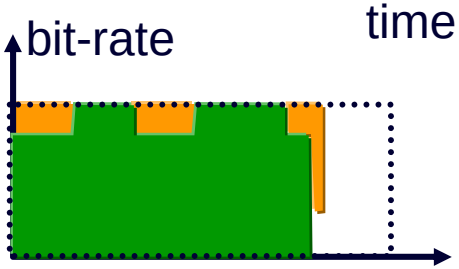
1. TCP



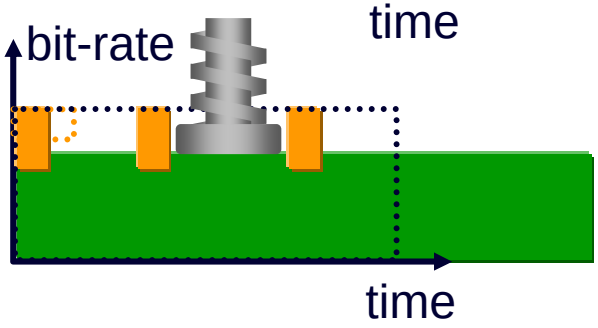
2. WFQ



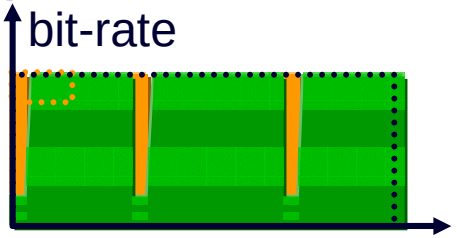
3. volume cap



4. DPI



weighted sharing

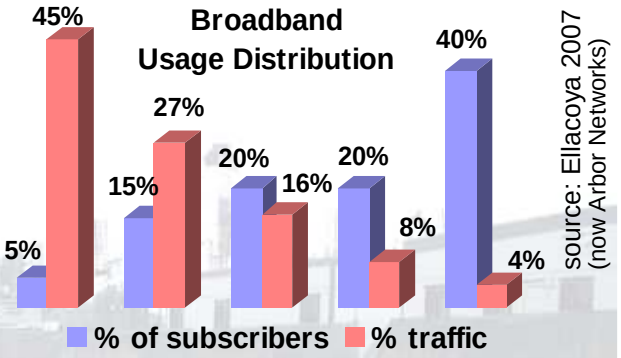


- light usage can go much faster
- hardly affects completion time of heavy usage

NOTE: weighted sharing doesn't imply differentiated network service

Just weighted aggressiveness of end-system's rate response to congestion cf. LEDBAT

closing off the future

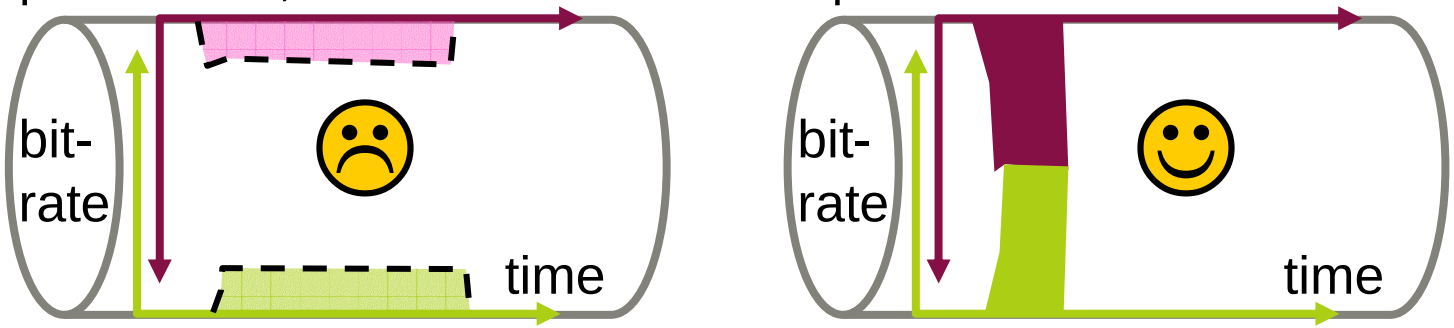


- ISPs must have a role in bandwidth sharing
  - minimally, incentivise end-systems to manage congestion
  - can't today, because ISPs can't see path congestion
- without correct metric, ISPs resort to application analysis
  - getting impossible to deploy a new use of the Internet
  - must negotiate the arbitrary blocks and throttles en route
- two confusable motives
  - fairer cost sharing
  - competitive advantage to own services

- how to deconfuse: make cost of usage transparent
  - fixing Internet technology should avoid need for legislation

congestion is not evil
congestion signals are healthy

- no congestion across whole path ⇒ feeble transport protocol
  - to complete ASAP, transfers should sense path bottleneck & fill it



the trick

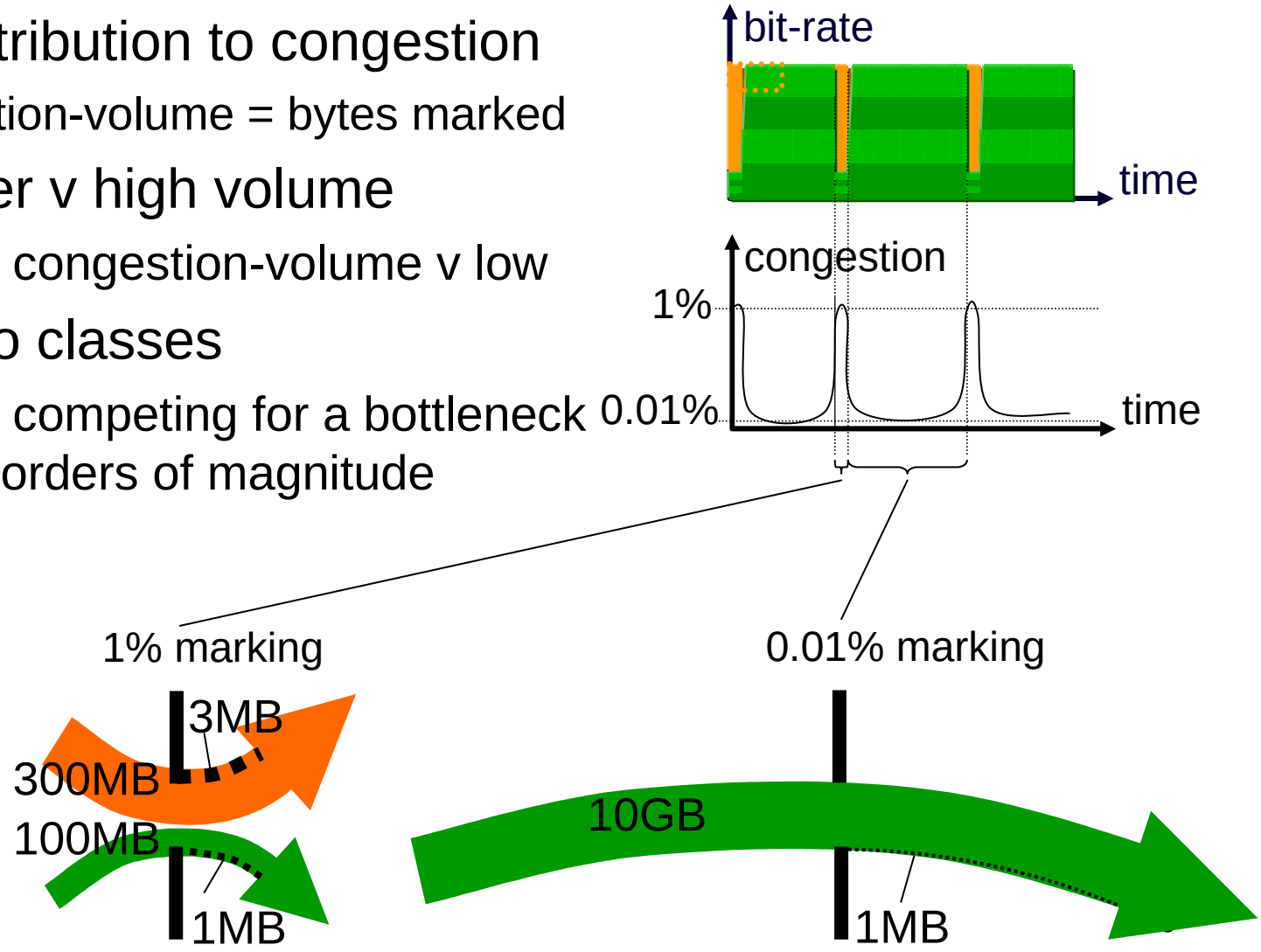
congestion signal *without* impairment

- explicit congestion notification (ECN); update to IP (2001)
  - mark more packets as queue builds
  - then tiny queuing delay and tiny loss for all traffic
  - no need to avoid congestion signals to prevent impairment
- original ECN: gain too small to overcome deployment barriers

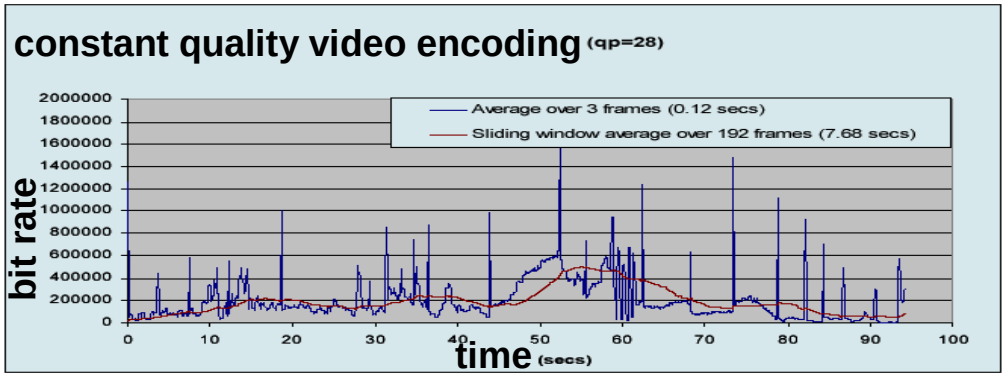
measuring contribution to congestion

- user's contribution to congestion

  congestion-volume = bytes marked
- can transfer v high volume
  - but keep congestion-volume v low
- not just two classes
  - file sizes competing for a bottleneck span ~7 orders of magnitude



harnessing flexibility
within streamed video



- 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
- equitable quality streamed video
  - aim for constant mean opinion score
  - deliver over MulTCP
    - TCP with weight parameter, ***n***
  - adjust ***n*** to 'hardness' of video [Mulroy09]
  - near constant MOS and 216% more videos over backhaul than today's constant bit-rate [Crabtree09]
- services want freedom & flexibility
  - access to a large shared pool, not a pipe
- when freedoms collide
  - many services can adapt to congestion
  - shift around resource pool in time/space

powerful resource accountability metric

congestion-volume

- volume
  weighted by congestion when sent

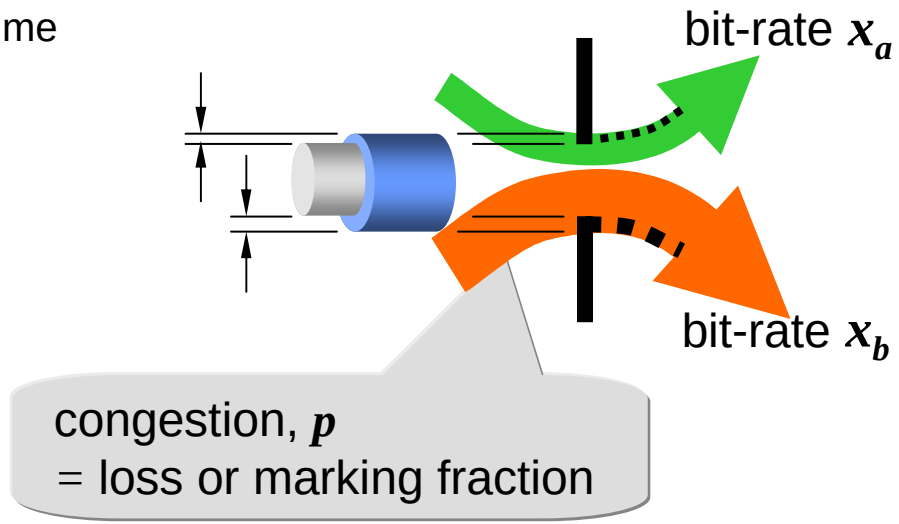
- intuition
  - contribution to congestion
  - some ISPs count volume only during peak
  - like counting (100% x volume) during peak
    and (0% x volume) otherwise
  - congestion-volume counts $\boldsymbol{p} \cdot \boldsymbol{x}_i$ over time

- a dual metric
  - of customers to ISPs (too much traffic)
  - and ISPs to customers (too little capacity)

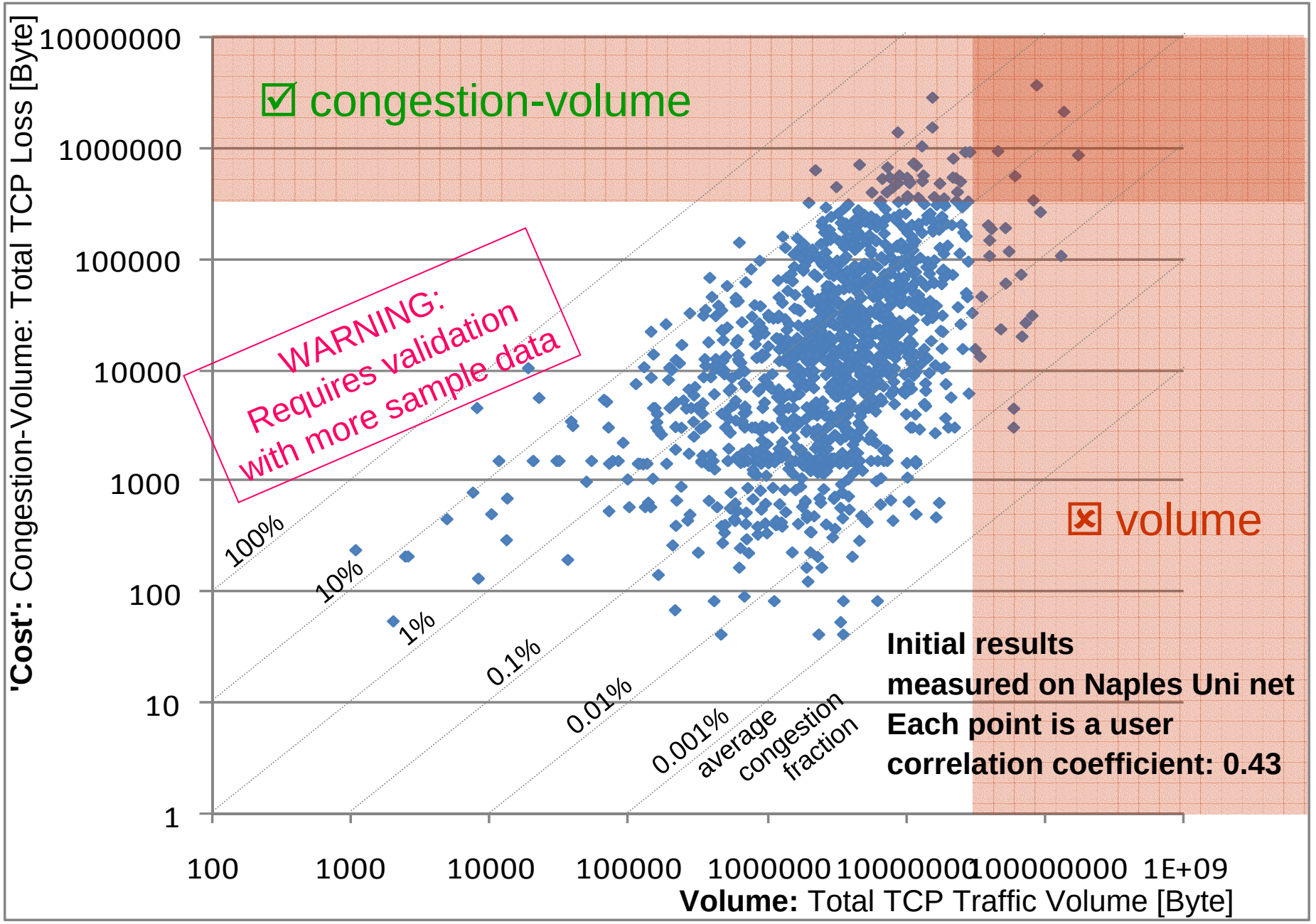
a) cost to other users of your traffic

b) marginal cost of equipment upgrade
  - so it wouldn't have been congested
  - so traffic wouldn't have affected others

- competitive market matches a) & b)



note: diagram is conceptual
congestion volume & capital cost of equipment would be accumulated over time


'Cost': Congestion-Volume: Total TCP Loss [Byte]
10000000
1000000
100000
10000
1000
100
10
1
☑ congestion-volume
WARNING:
Requires validation
with more sample data
☒ volume
100%
10%
1%
0.1%
0.01%
0.001%
average
congestion
fraction
Initial results
measured on Naples Uni net
Each point is a user
correlation coefficient: 0.43
100
1000
10000
100000
1000000
10000000
100000000
1E+09
Volume: Total TCP Traffic Volume [Byte]

if only... ingress net could see congestion...

flat fee congestion policing

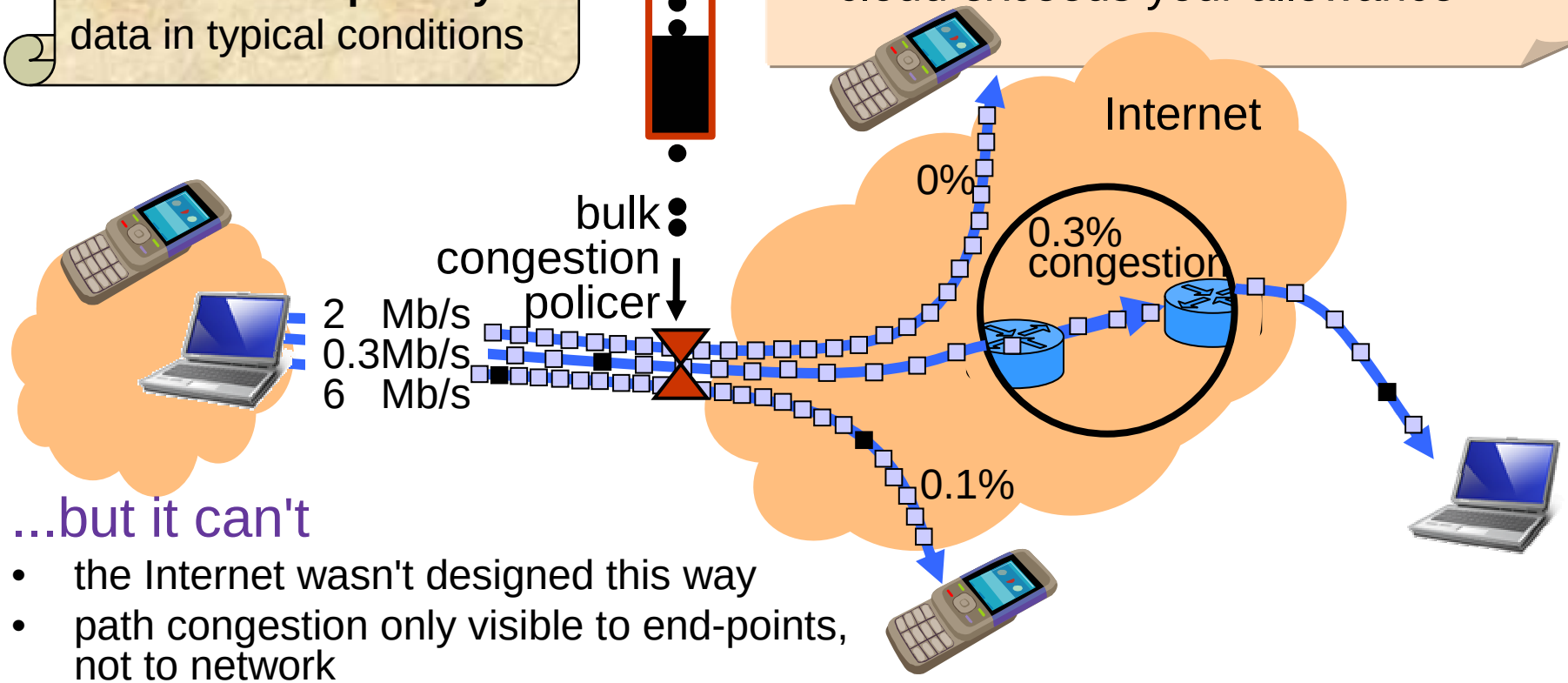
Acceptable Use Policy

'congestion-volume' allowance: 1GB/month

@ €15/month

Allows ~**70GB per day** of data in typical conditions

- incentive to avoid congestion
- policing only necessary at edge
- only throttles traffic when your contribution to congestion in the cloud exceeds your allowance

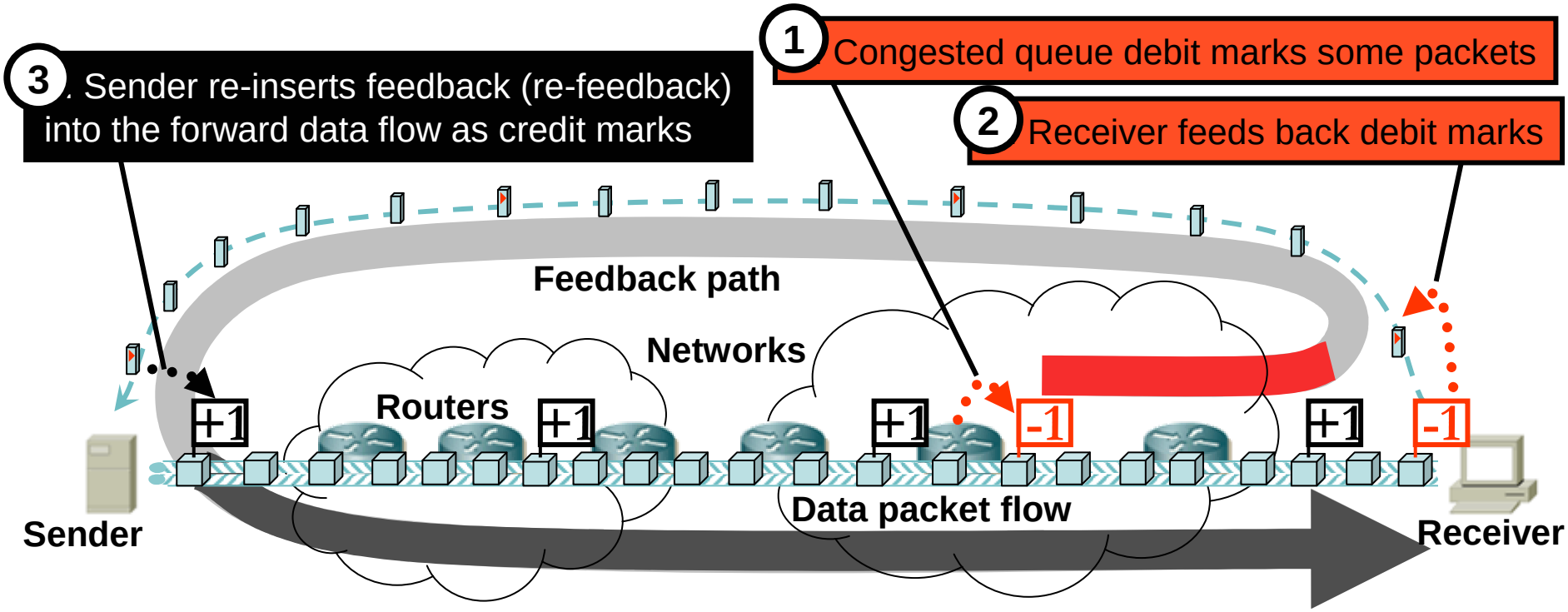


...but it can't

- the Internet wasn't designed this way
- path congestion only visible to end-points, not to network

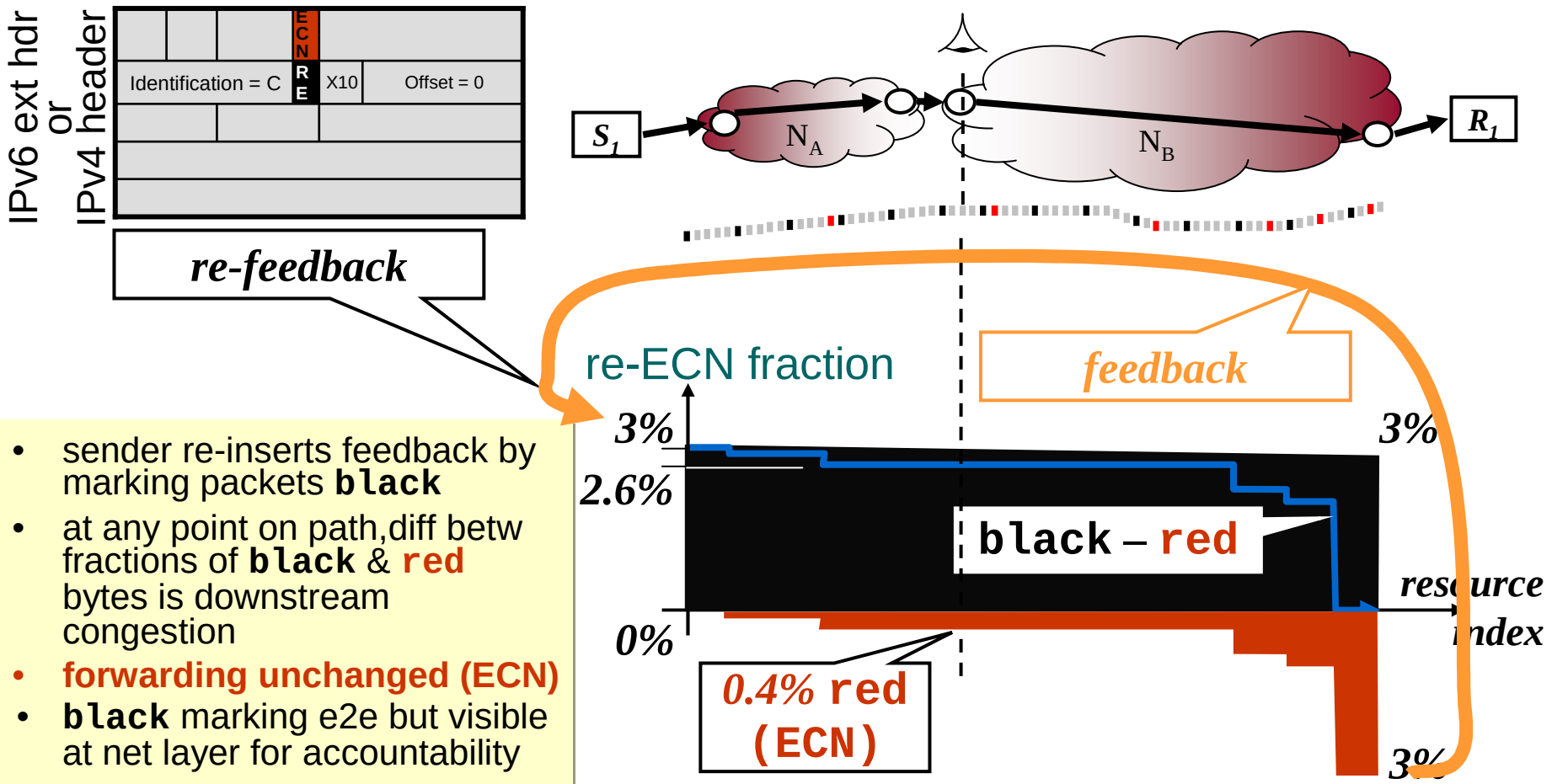
congestion exposure

standard ECN + re-inserted feedback (re-feedback) = re-ECN



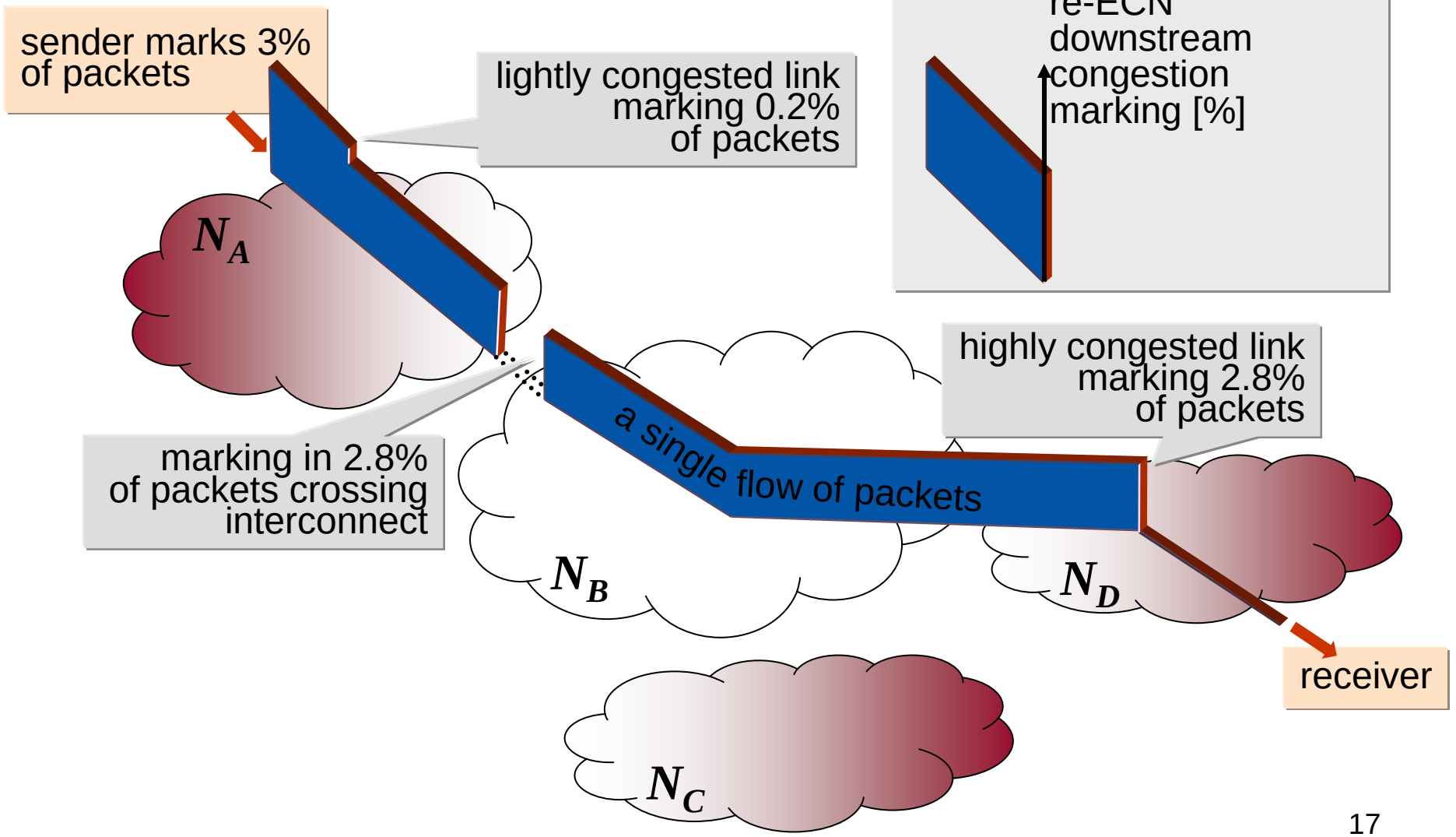
no changes required to IP data forwarding

congestion exposure with ECN & re-ECN
measurable upstream, downstream and path congestion



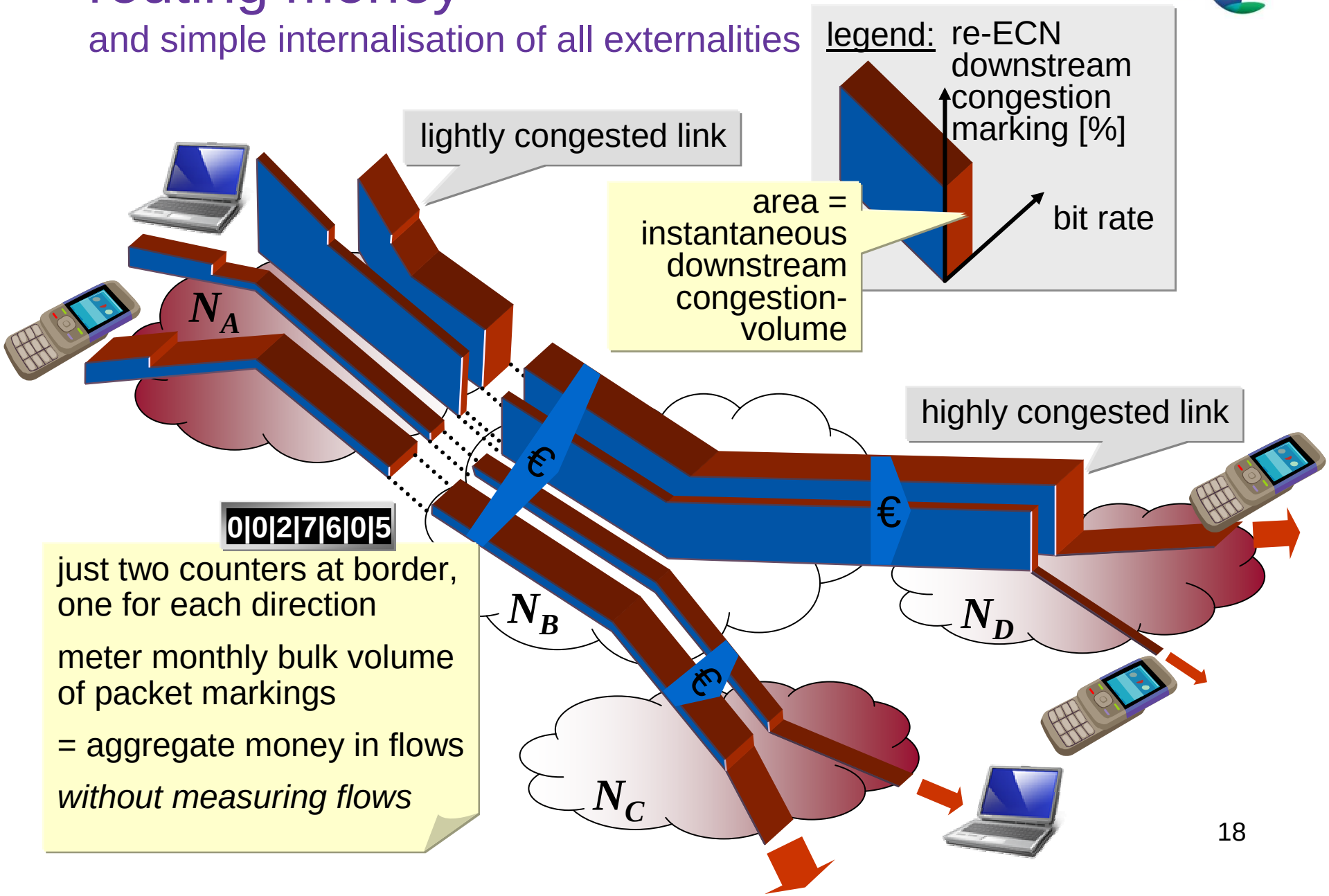
- sender re-inserts feedback by marking packets **black**
- at any point on path,diff betw fractions of **black** & **red** bytes is downstream congestion
- **forwarding unchanged (ECN)**
- **black** marking e2e but visible at net layer for accountability

network can now see which packets won't fit



routing money

and simple internalisation of all externalities



just two counters at border, one for each direction

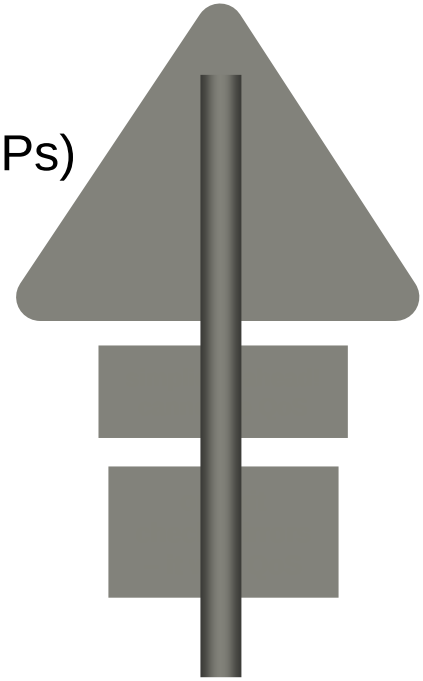
meter monthly bulk volume of packet markings

= aggregate money in flows

without measuring flows

best without effort

- did you notice the interconnected QoS mechanism?
  - *endpoints* ensure tiny queuing delay & loss for all traffic
  - if your app wants more bit-rate, it just goes faster
  - effects seen in bulk metric at every border (for SLAs, AUPs)
  - simple – and all the right support for operations
- neutral
  - no permission needed for QoS
  - application agnostic

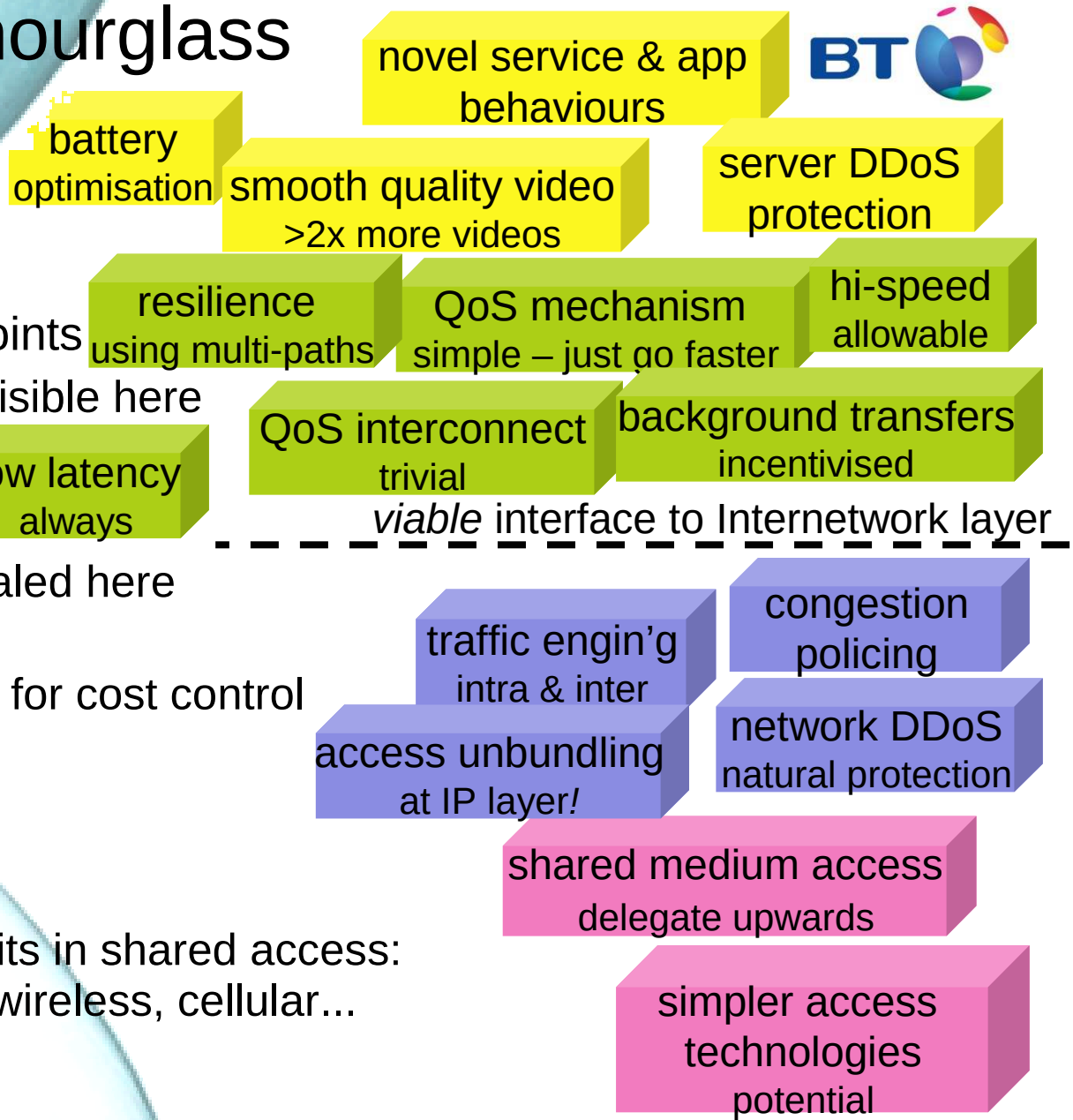


- long-held beliefs left by the wayside
  - QoS requires SLAs and guarantees
  - low latency needs special QoS treatment
  - QoS needs a new API

  - fairness = same instantaneous rates
  - people want to buy fixed bit-rate

the neck of the hourglass ...but for control

- applications & services
- transport layer on end-points
  - usage costs currently visible here
- internetwork layer
  - once usage costs revealed here
  - ISPs won't need deep packet inspection for cost control
- link layer
  - can remove bit-rate limits in shared access: passive optical, cable, wireless, cellular...



summary
network and host co-operation

- congestion-volume
  - a metric to express and resolve conflicting interests
  - robust to self-interest and malice
- neutral
- ambitious but simple
- new horizons for the Internet if we take the challenge

more info...

- The whole story in 7 pages
  - Bob Briscoe, “Internet Fairer is Faster", BT White Paper (Jun 2009) ...this formed the basis of:
  - Bob Briscoe, "A Fairer, Faster Internet Protocol", IEEE Spectrum (Dec 2008)
- Slaying myths about fair sharing of capacity
  - [Briscoe07] Bob Briscoe, "Flow Rate Fairness: Dismantling a Religion" ACM Computer Communications Review 37(2) 63-74 (Apr 2007)
- How wrong Internet capacity sharing is and why it's causing an arms race
  - Bob Briscoe et al, "Problem Statement: Transport Protocols Don't Have To Do Fairness", IETF Internet Draft (Jul 2008)
- re-ECN protocol spec
  - Bob Briscoe et al, “Adding Accountability for Causing Congestion to TCP/IP” IETF Internet Draft (Mar 2009)
- Equitable quality streaming of video
  - [Crabtree09] B. Crabtree, M. Nilsson, P. Mulroy and S. Appleby “Equitable quality video streaming” Computer Communications and Networking Conference, Las Vegas, (January 2009)
  - [Mulroy09] Mulroy, P., Appleby, S., Nilsson, M. & Crabtree, B., "The Use of MulTCP for the Delivery of Equitable Quality Video," In: *Proc. Int'l Packet Video Wkshp (PV'09)* IEEE (May 2009)

Re-architecting the Internet:

- The Trilogy project <www.trilogy-project.org>

IRTF Internet Capacity Sharing Architecture design team

<http://trac.tools.ietf.org/group/irtf/trac/wiki/CapacitySharingArch>

re-ECN & re-feedback project page:

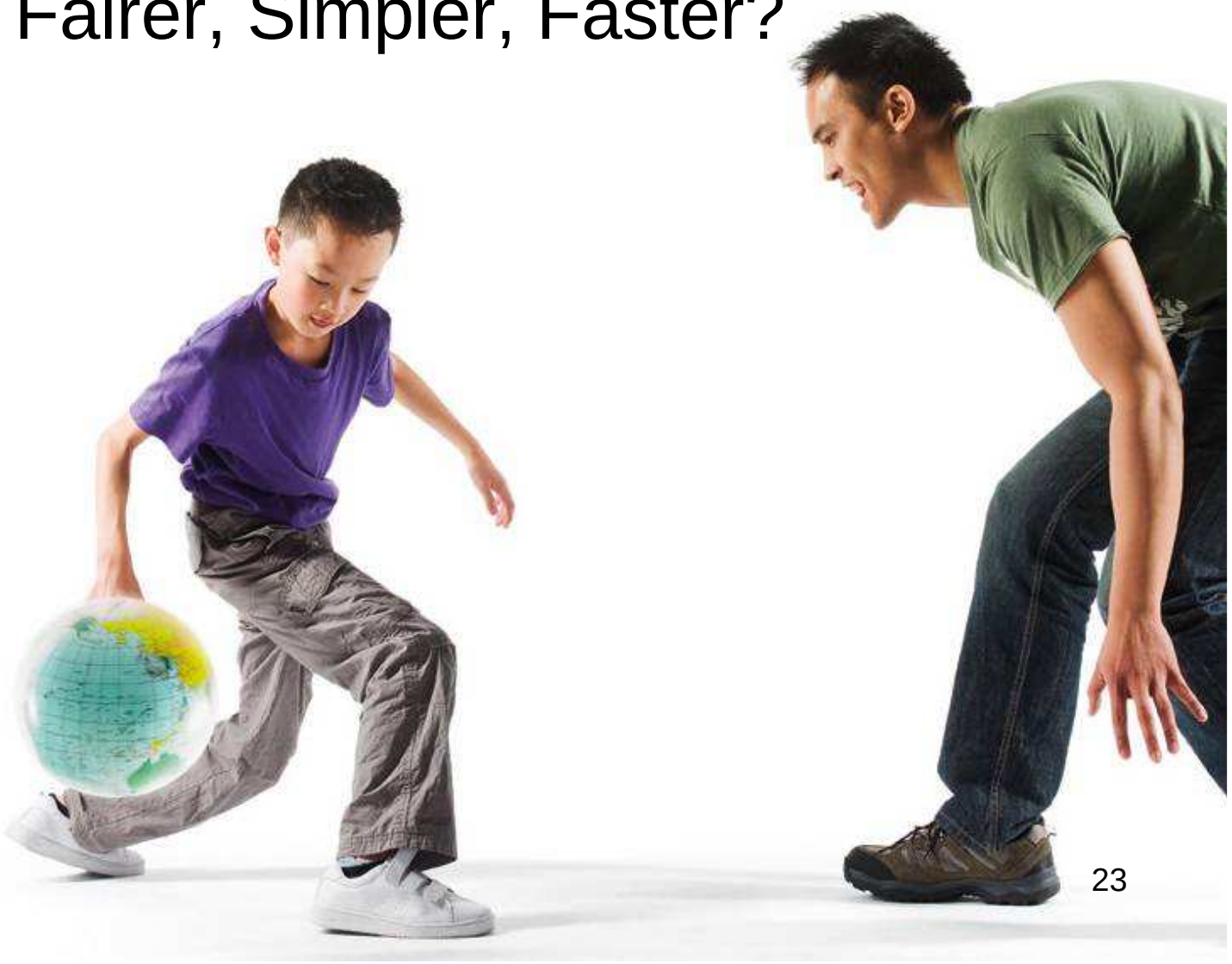
<http://bobbriscoe.net/projects/refb/>

Congestion Exposure (ConEx) IETF ‘BoF’: <http://trac.tools.ietf.org/area/tsv/trac/wiki/re-ECN>
subscribe: <https://www.ietf.org/mailman/listinfo/re-ecn>, post: re-ecn@ietf.org

implementation (linux or ns2) bob.briscoe@bt.com

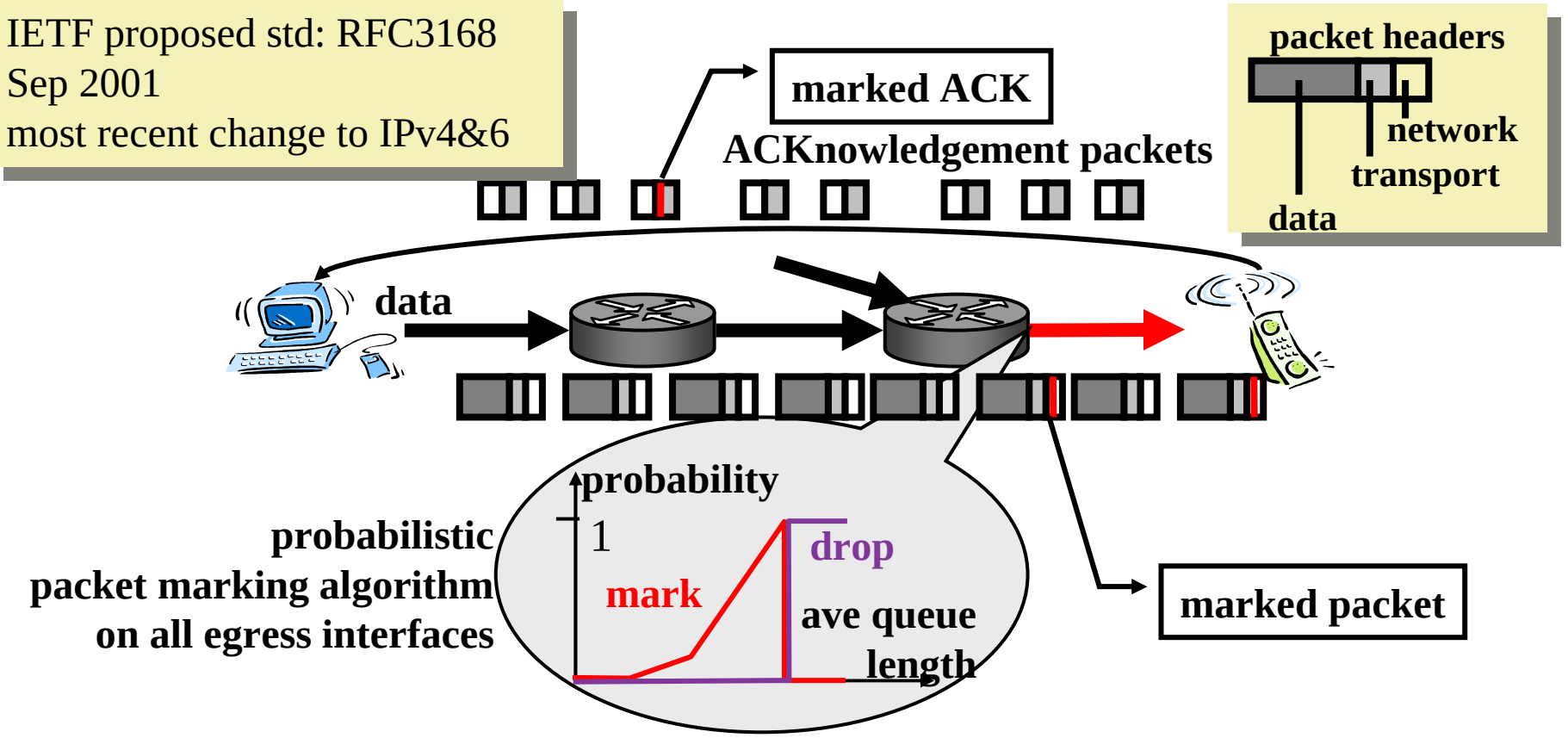
Internet capacity sharing: Fairer, Simpler, Faster?

discuss...

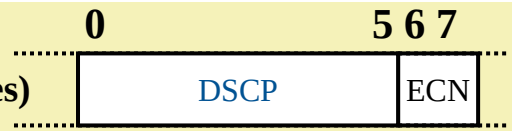


explicit congestion notification (ECN)

IETF proposed std: RFC3168
Sep 2001
most recent change to IPv4&6



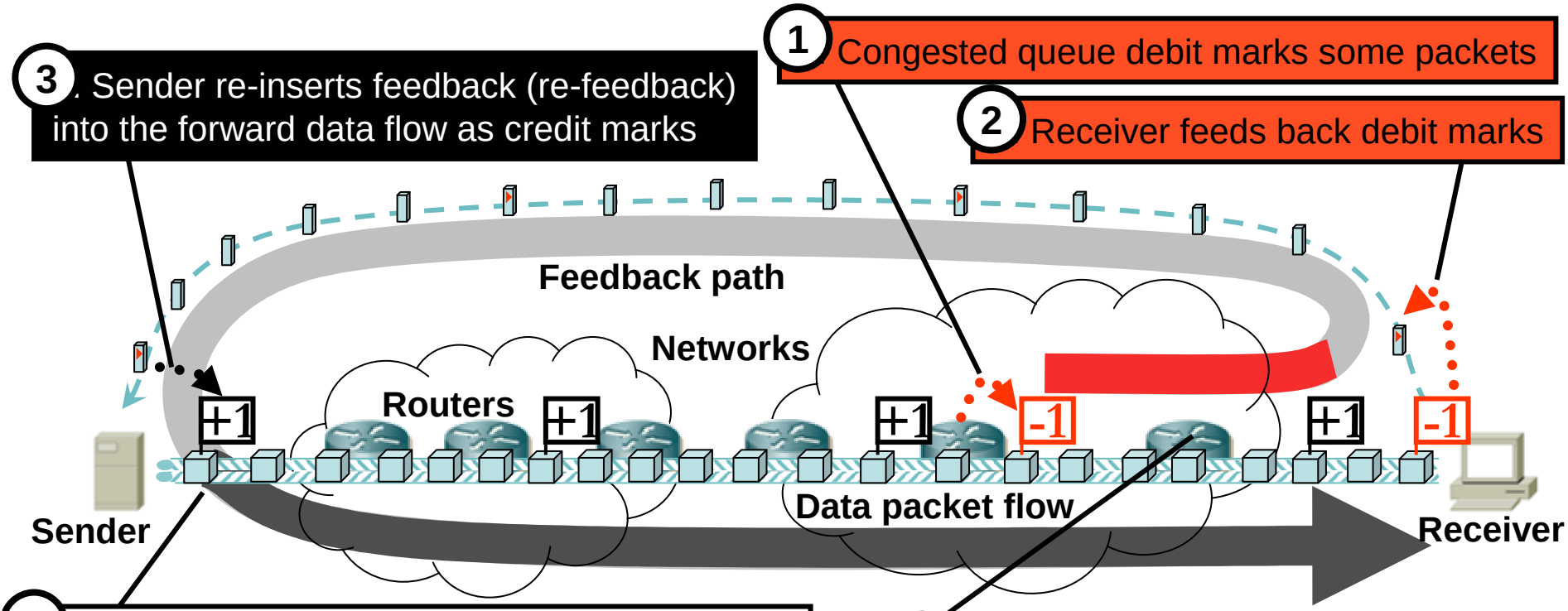
00:	**Not ECN Capable Transport (ECT)**
01 or 10:	**ECN Capable Transport - no Congestion Experienced (sender initialises)**
11:	**ECN Capable Transport - and Congestion Experienced (CE)**



bits 6 & 7 of **IP** DS byte

congestion exposure

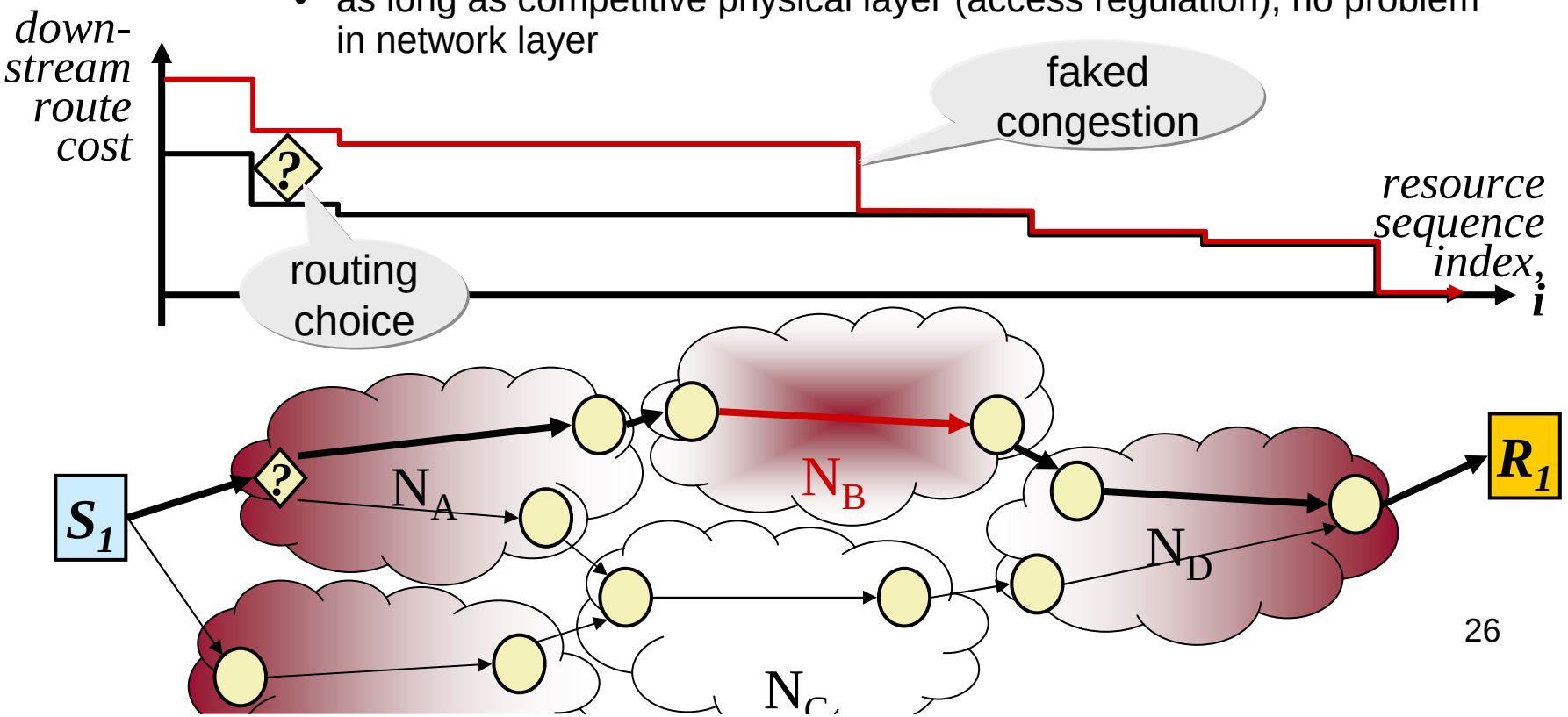
standard ECN + re-inserted feedback (re-feedback) = re-ECN



no changes required to IP data forwarding

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



main steps to deploy re-feedback / re-ECN

summary
rather than control sharing in the access links,
pass congestion info & control upwards

- hosts
  - (minor) addition to TCP/IP stack of sending device
  - or sender proxy in network
- 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
- new phase of Internet evolution starts
  - customer contracts & interconnect contracts
  - endpoint applications and transports
- requires update to the IP standard (v4 & v6)
  - in progress at IETF
  - using bits in IPv4 header or IPv6 extension header