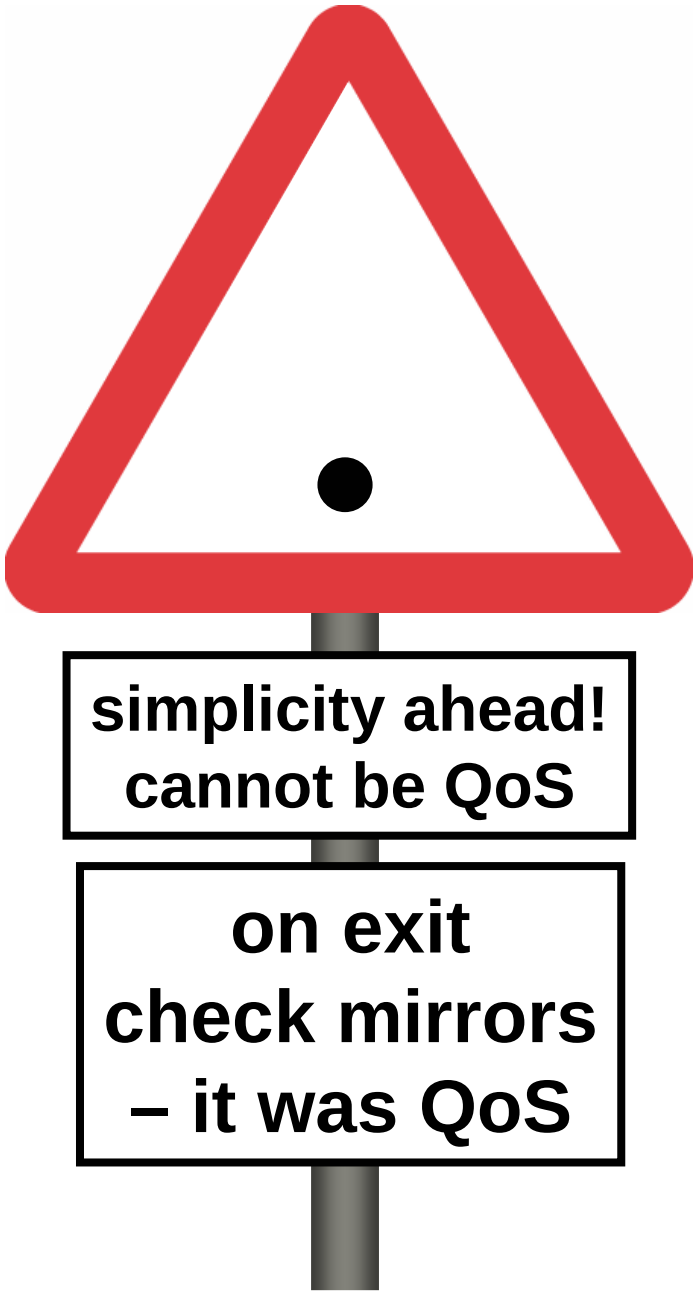
QoS interconnect best without effort

Bob Briscoe
Chief Researcher
BT

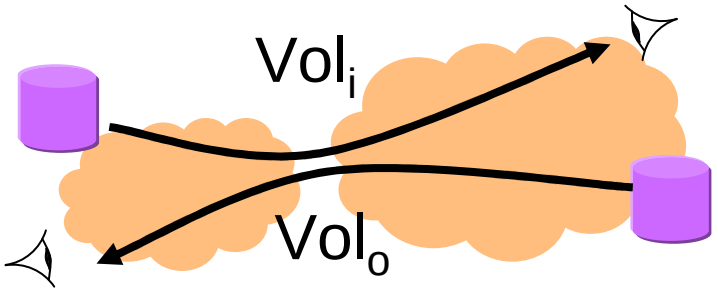
Sep 2009

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

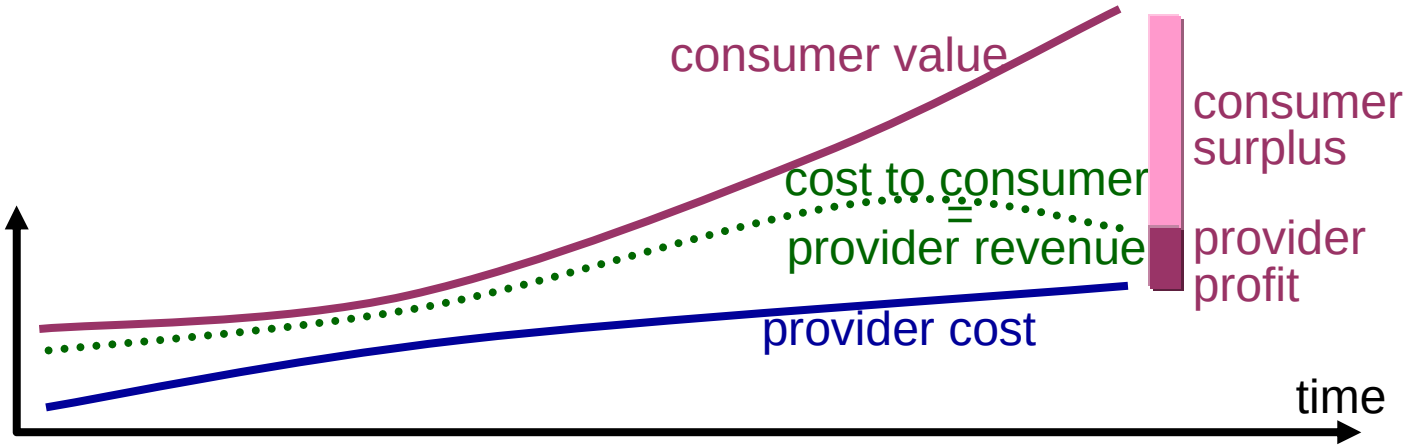


BT

both value *and* cost

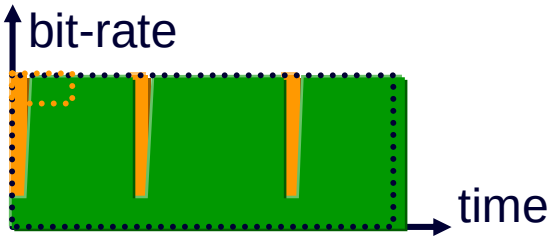


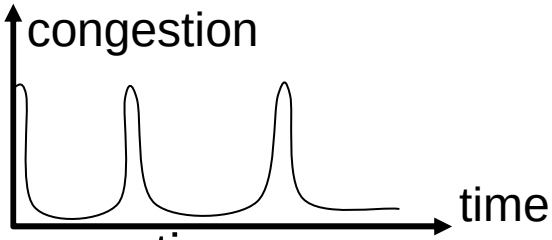
- industry contractual metrics are largely value-based
  - e.g. advertised routes, volume ratio
  - even a CEO should understand both value *and* cost
- competitive market drives revenues down towards provider's marginal cost
  - those who understand marginal costs will succeed



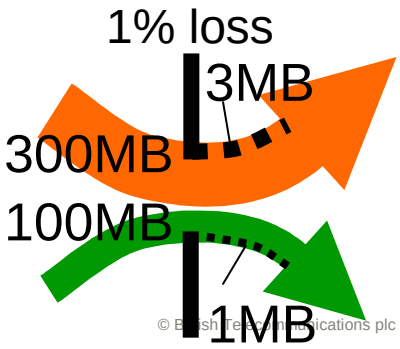
BT

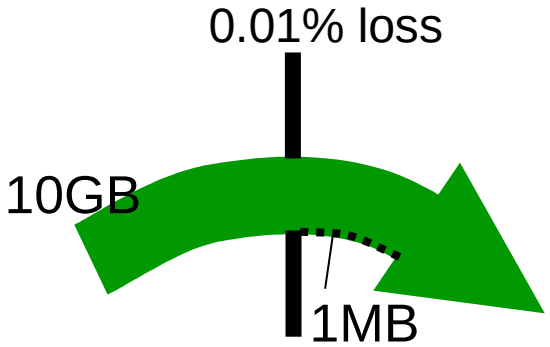
marginal cost of network usage?



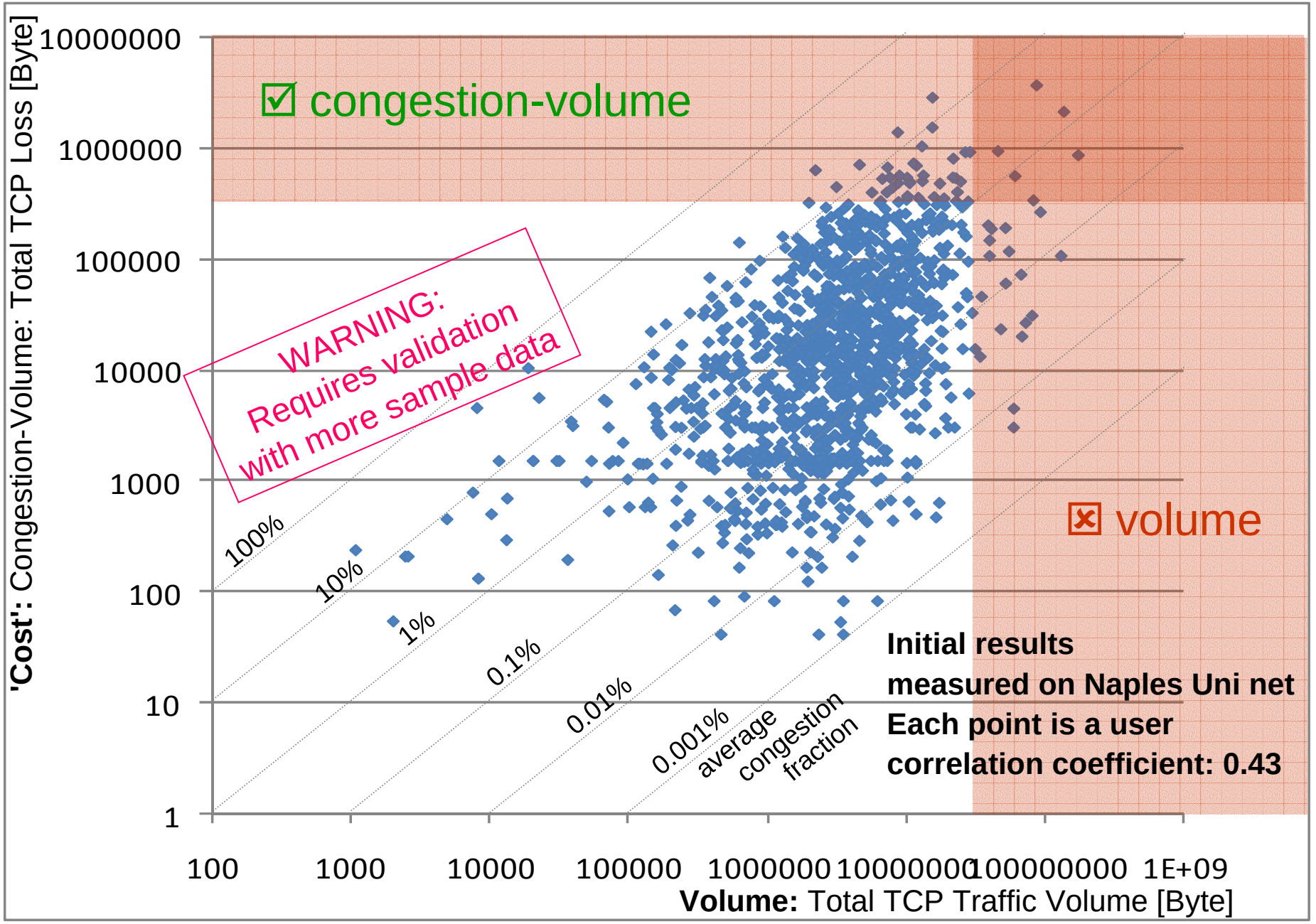


- volume is NOT a good measure
- green user yields whenever detects high congestion
  - very high volume but very low cost to others
  - e.g. LEDBAT (BitTorrent's low extra delay background transport) or weighted TCP
- by counting volume, ISPs kill nice behaviour
  - not just file transfers, e.g. congestion-sensitive video codec transfers >100% more videos thru same capacity (same MoS)
- correct measure: congestion-volume
  - volume weighted by congestion when it is forwarded
  - easily measured by a host
  - bytes sent x loss fraction = bytes lost





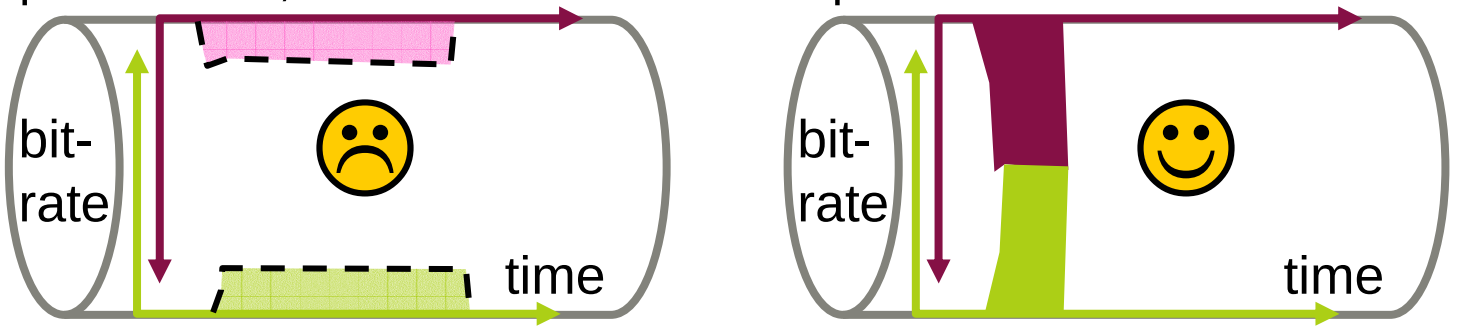
BT


'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]

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 in 2001: mark more packets as queue builds
- then tiny queuing delay and tiny tiny loss for all traffic

- no need to avoid congestion (whether core, access or borders) to prevent impairment

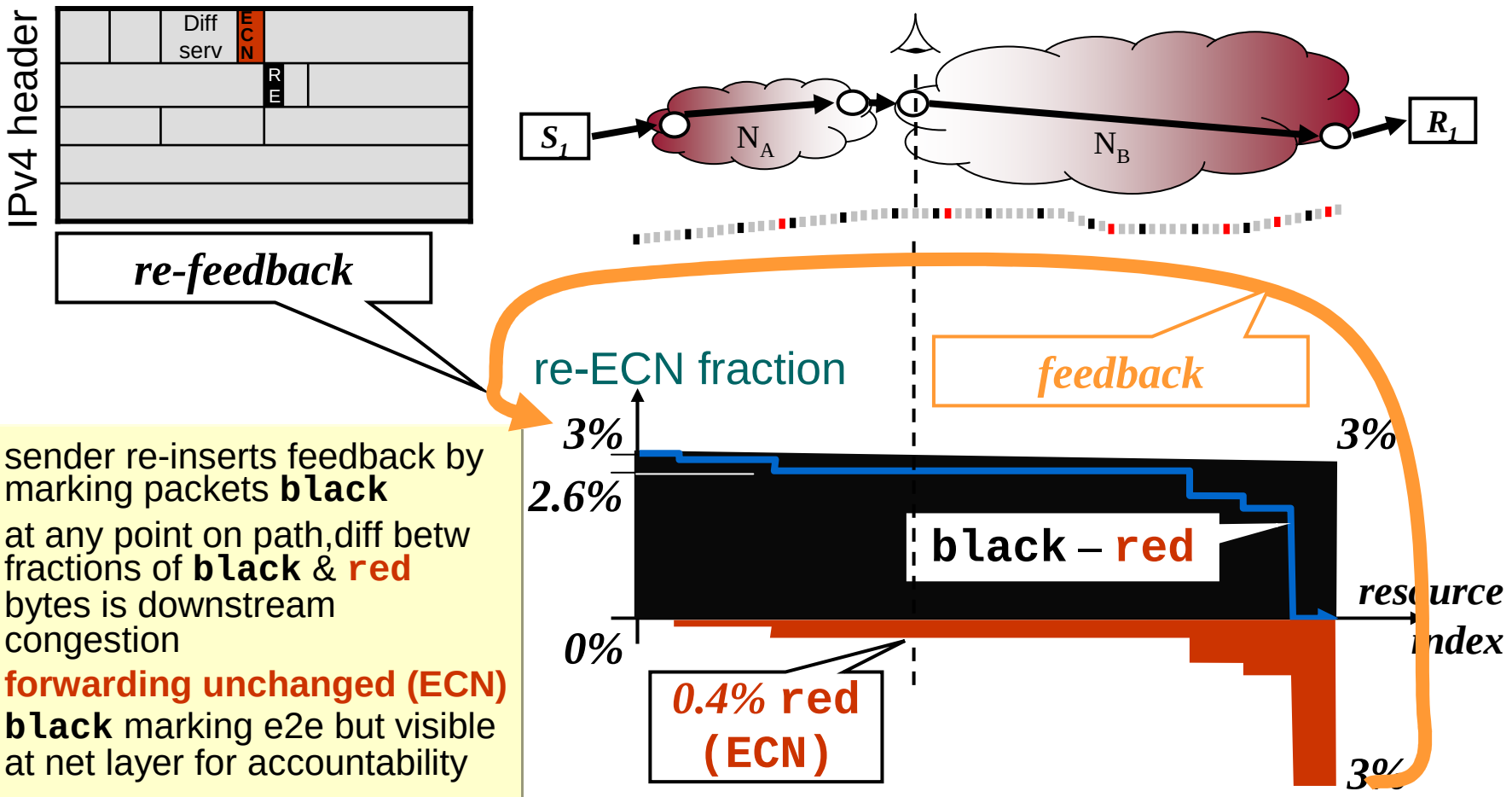
BT

congestion exposure

- by Internet design, endpoints detect & handle losses
  - v hard for networks to see losses (marginal costs)
- proposed IETF working group: “congestion exposure”
  - protocol for sender to mark IP headers to expose congestion
  - to measure traffic cost as easily as we measure volume
  - just count volume of marked packets in aggregate
  - >40 offers of help just in the last fortnight
- named re-ECN (re-inserted ECN)
  - builds on explicit congestion notification (ECN [RFC3168])

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

BT

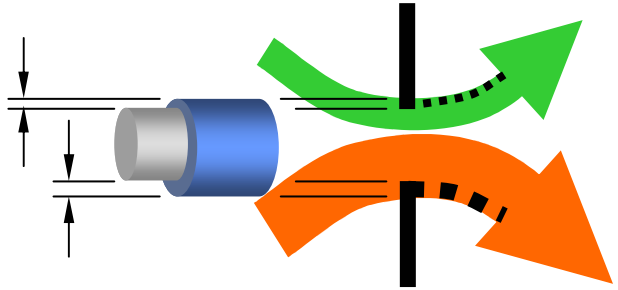
congestion-volume metric

dual demand & supply role

- a resource accountability metric
  1. of customers to ISPs (too much traffic)
  2. and ISPs to customers (too little capacity)

  1. cost to other users of my traffic
  2. the marginal cost of upgrading equipment
     - so it wouldn't have been congested

- competitive market matches 1 & 2



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

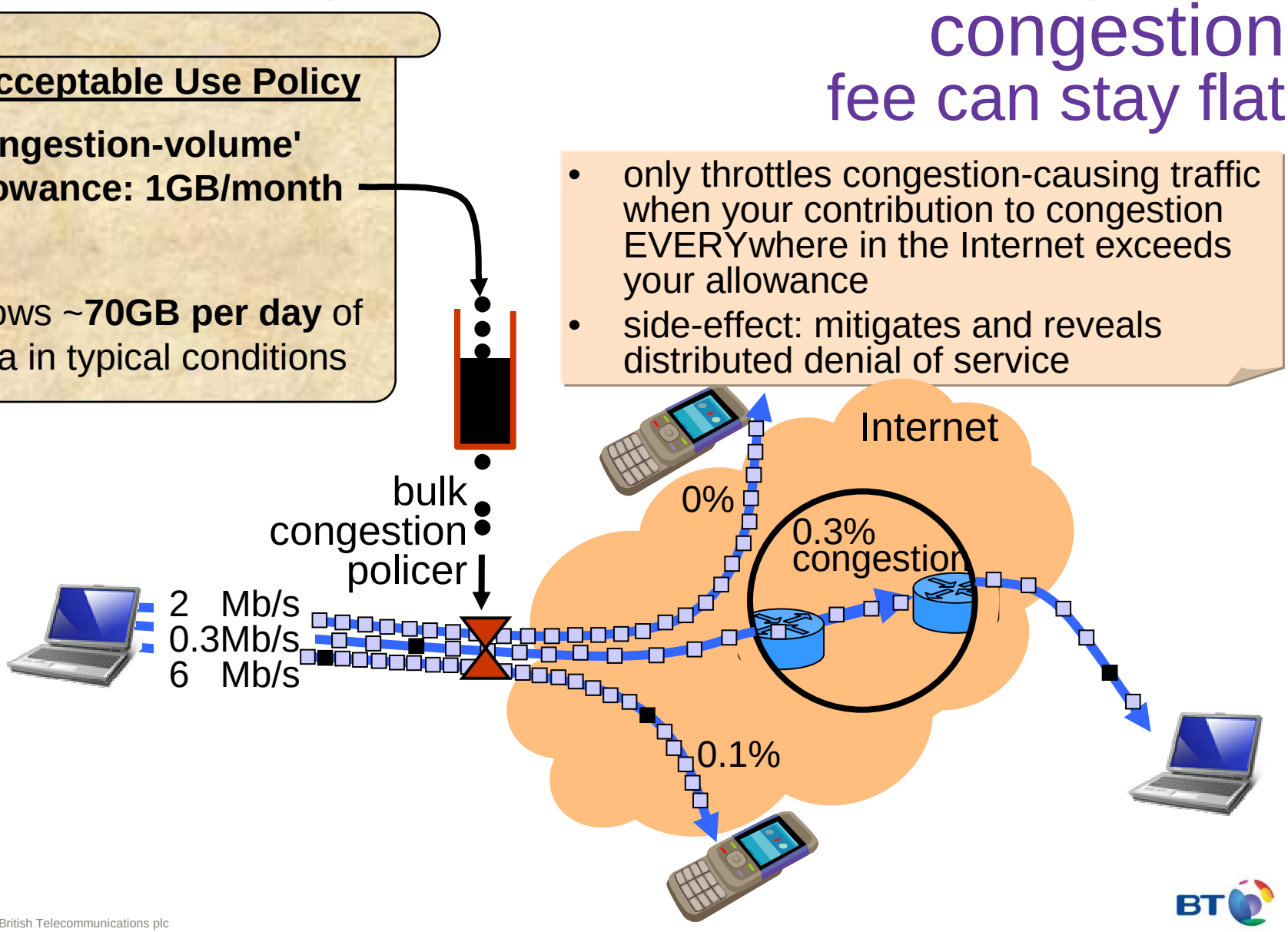
example consumer use of exposed congestion
fee can stay flat

Acceptable Use Policy

'congestion-volume' allowance: 1GB/month

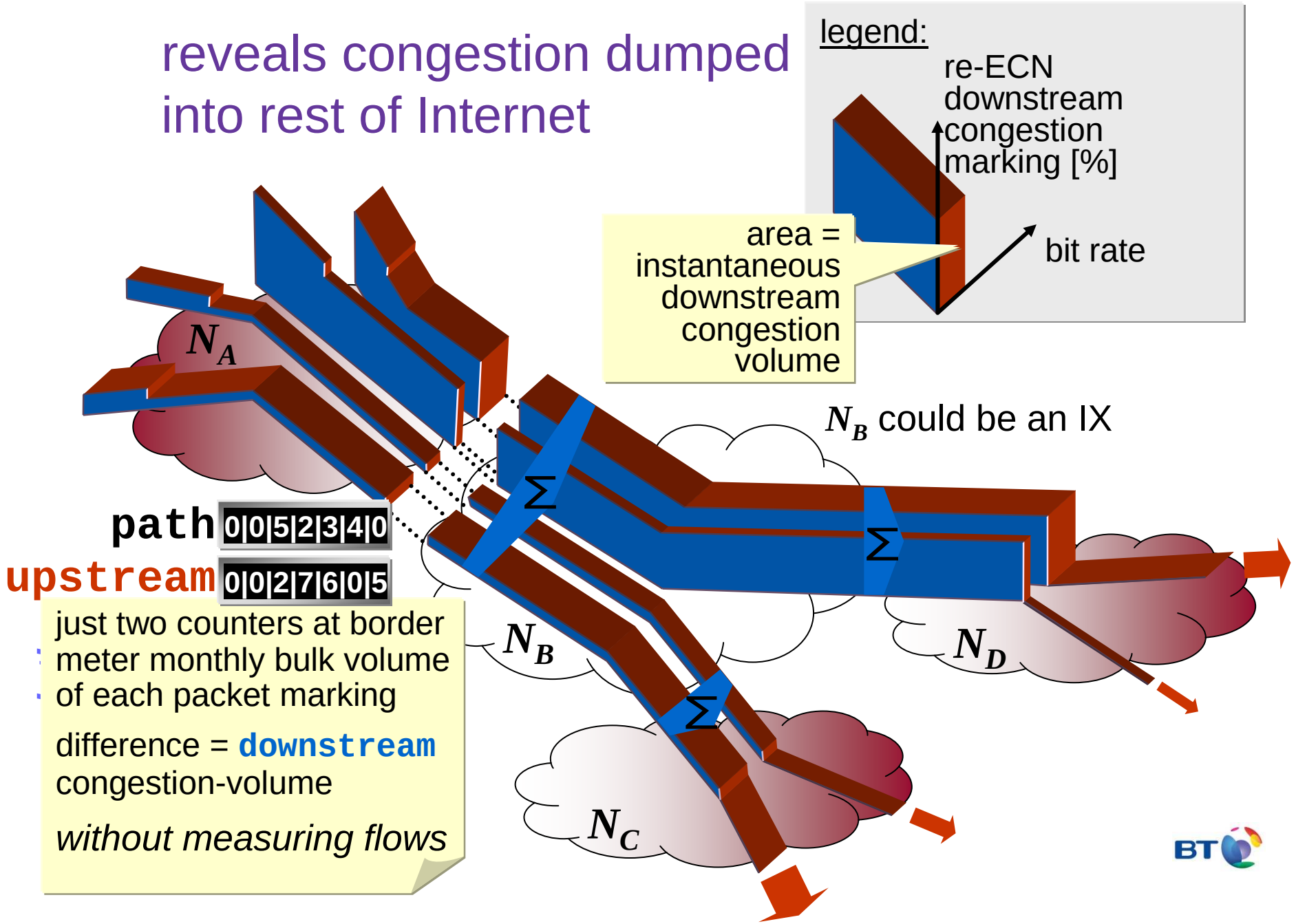
Allows ~**70GB per day** of data in typical conditions

- only throttles congestion-causing traffic when your contribution to congestion EVERYwhere in the Internet exceeds your allowance
- side-effect: mitigates and reveals distributed denial of service



BT

reveals congestion dumped into rest of Internet



I’m a conservative, get me out of here!

- if we don’t listen to the economics, we’re all dead
  - shift from value-based to cost-based is unstoppable
    - competition
  - bit transport needs to be viable on its own

(another talk)

- as cost pressures grow
- existing capacity sharing methods feed an arms race
  - TCP doesn’t share capacity fairly by any means
    - recent unanimous consensus in IETF Transport Area
  - ISPs have quietly been fighting TCP with piecemeal tools
    - WFQ, volume capping, deep packet inspection
- with congestion in IP header, wouldn’t need to look deeper

BT

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

- the invisible hand of the market
  - favours ISPs that get their customers to manage their traffic in everyone else‘s best interests
- incentives to cooperate across Internet value chain
  - content industry, CDNs, app & OS authors, network wholesalers & retailers, Internet companies, end-customers, business, residential

- if you want this, vote early and vote often!
  - re-ecn@ietf.org list
  - IETF, Hiroshima, Nov’09

BT

more info...

- White paper – the whole story in 7pp
  - Internet: Fairer is Faster, Bob Briscoe (BT), BT White Paper TR-CXR9-2009-001 (May 2009)
    - an abridged version of this article appeared in IEEE Spectrum, Dec 2008
- Inevitability of policing
  - The Broadband Incentives Problem, Broadband Working Group, MIT, BT, Cisco, Comcast, Deutsche Telekom / T-Mobile, France Telecom, Intel, Motorola, Nokia, Nortel (May '05 & follow-up Jul '06) <cfp.mit.edu>
- Stats on p2p usage across 7 Japanese ISPs with high FTTH penetration
  - Kenjiro Cho et al, "The Impact and Implications of the Growth in Residential User-to-User Traffic", In Proc ACM SIGCOMM (Oct '06)
- Slaying myths about fair sharing of capacity
  - 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)
- Understanding why QoS interconnect is better understood as a congestion issue
  - Bob Briscoe and Steve Rudkin "Commercial Models for IP Quality of Service Interconnect" BT Technology Journal 23 (2) pp. 171--195 (April, 2005)
- Re-architecting the Internet:
  - The Trilogy project
- Re-ECN & re-feedback project page:
  <http://bobbriscoe.net/projects/refb/>
  <trac.tools.ietf.org/area/tsv/trac/wiki/re-ECN>

BT

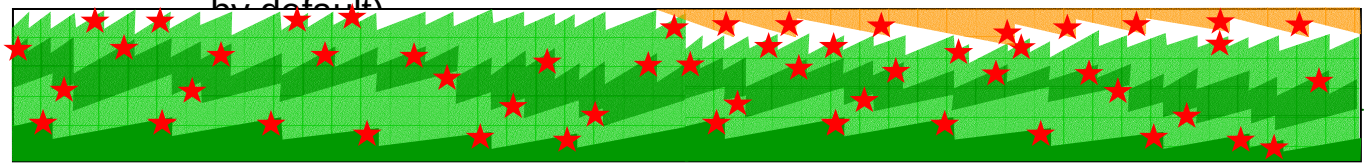
best without effort QoS interconnection

Q&A...

problems using congestion in contracts

	1. loss	2. ECN	3. re-ECN
can't justify selling an impairment	☹	☺	☺
absence of packets is not a contractible metric	☹	☺	☺
congestion is outside a customer's control	☹	☹	☺
customers don't like variable charges	☹	☹	☺
congestion is not an intuitive contractual metric	☹	☹	☹

1. **loss:** used to signal congestion since the Internet's inception
   - computers detect congestion by detecting gaps in the sequence of packets
   - computers can hide these gaps from the network with encryption
2. **explicit congestion notification (ECN):** standardised into TCP/IP in 2001
   - approaching congestion, a link marks an increasing fraction of packets
   - implemented in Windows Vista (but off by default) and Linux, and IP routers (off by default)

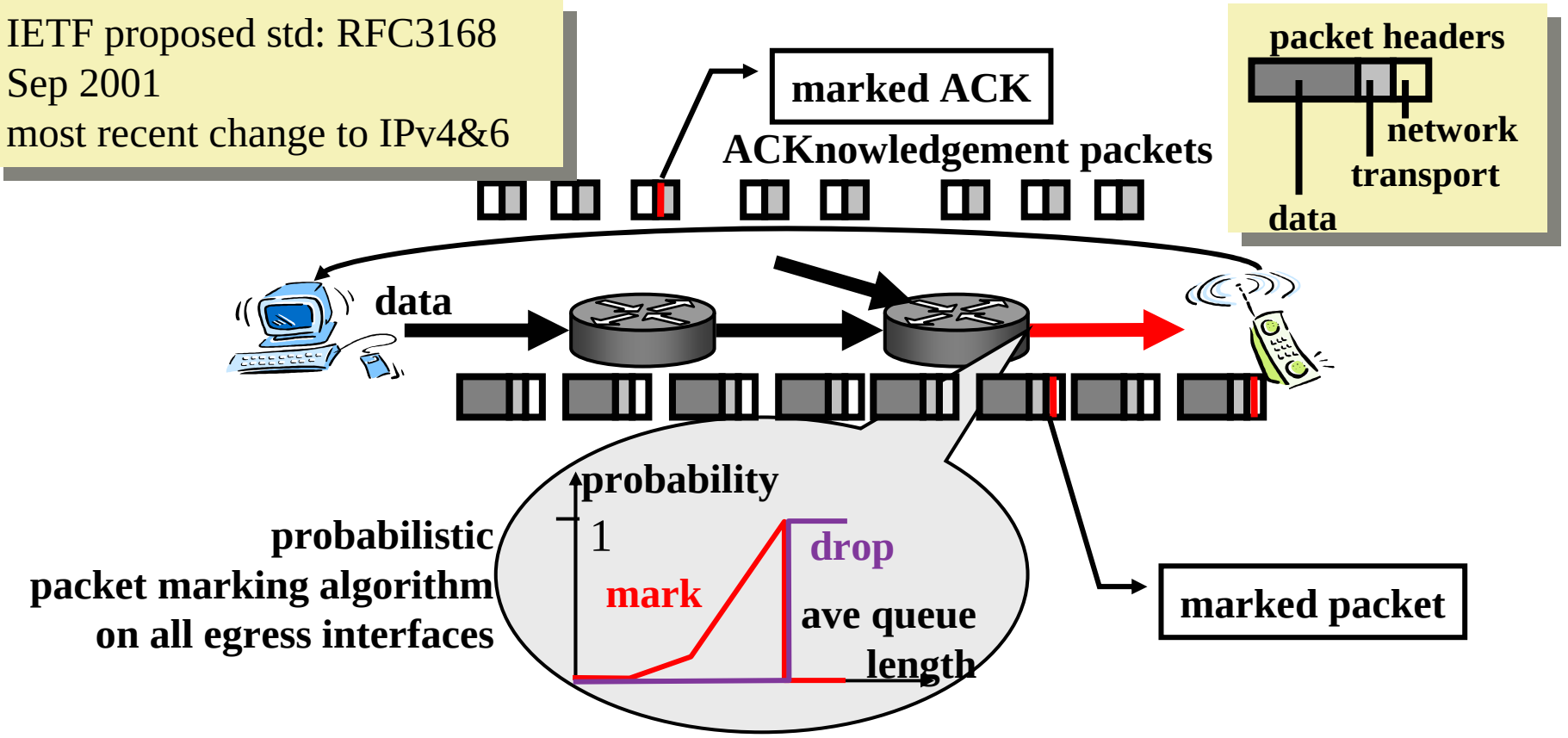


3. **re-inserted ECN (re-ECN):** standards proposal since 2005
   - packet delivery conditional on sender declaring expected congestion
   - uses ECN equipment in the network unchanged

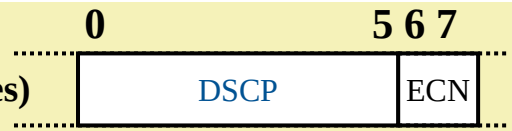
BT

explicit congestion notification (ECN)

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



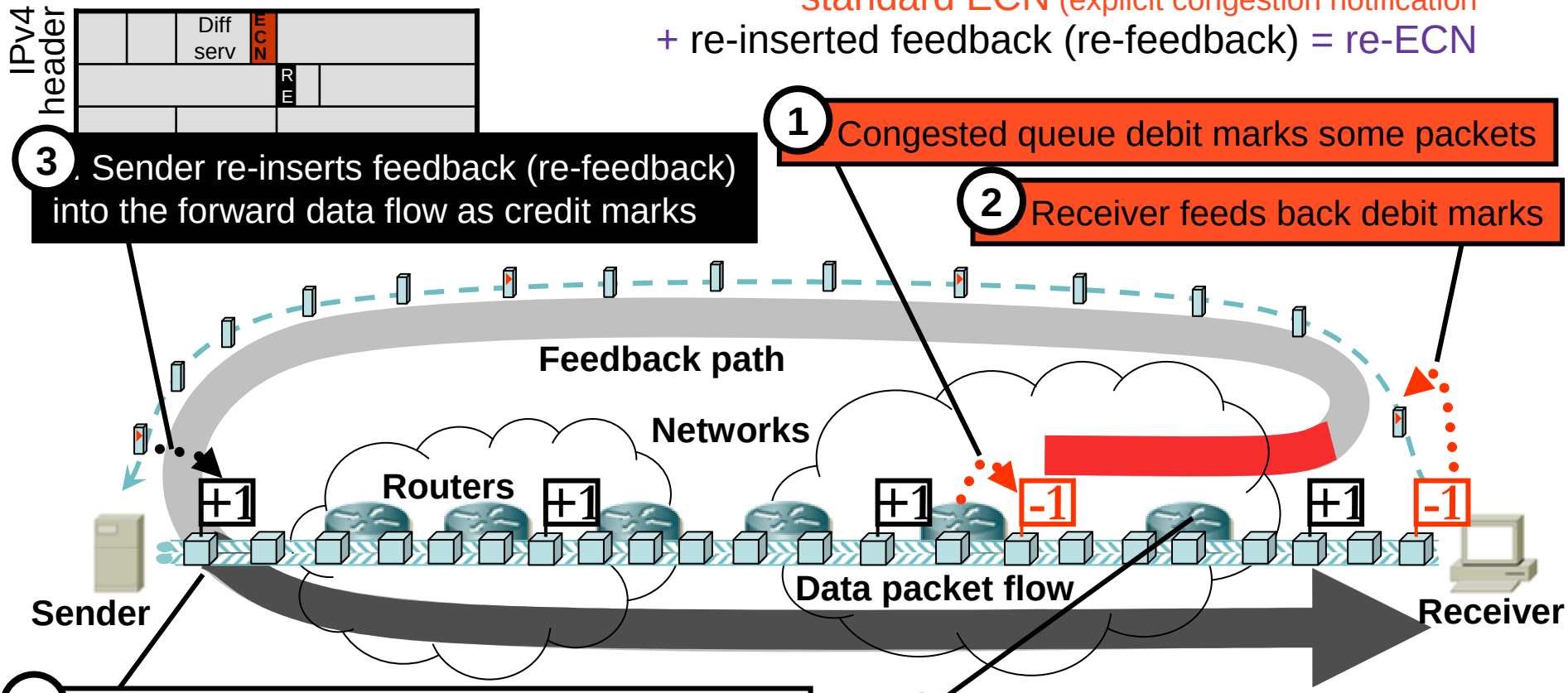
00:	**Not ECN Capable Transport (ECT)**
01 or 10:	**ECN Capable Transport - no Congestion Experienced (sender initialises)**
11:	**ECN Capable Transport - and Congestion Experienced (CE)**



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

congestion exposure in one bit

standard ECN (explicit congestion notification
+ re-inserted feedback (re-feedback) = re-ECN



no changes required to IP data forwarding

BT

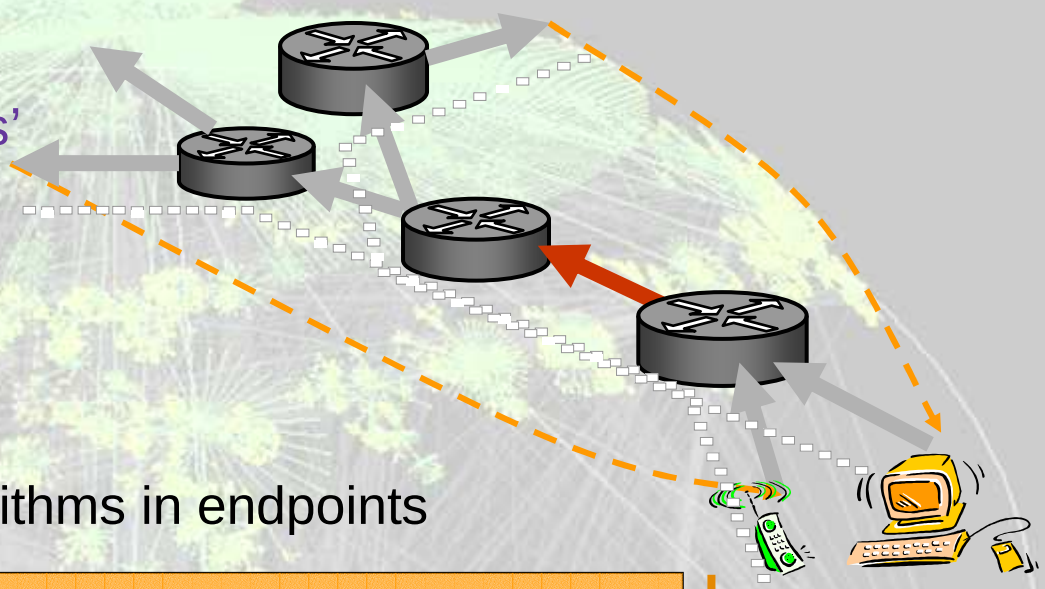
main steps to deploy re-feedback / re-ECN

- network
  - turn on explicit congestion notification in routers (already available)
  - 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
- customer contracts
  - include congestion cap
- oh, and first we have to update the IP standard
  - started process in Autumn 2005
  - using last available bit in the IPv4 packet header
  - proposal for new working group, Nov 2009 IETF

BT

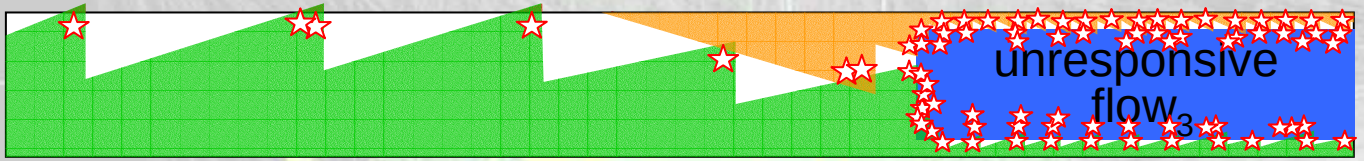
how Internet sharing 'works'

TCP-friendliness



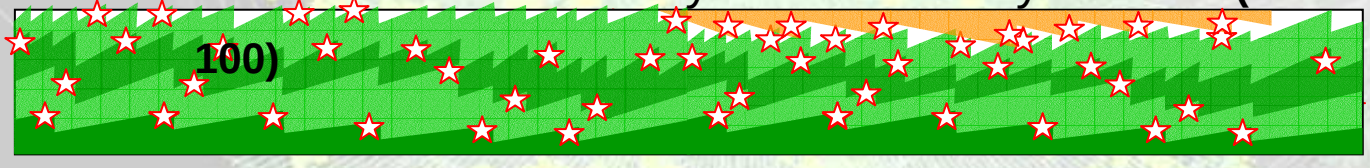
- endemic congestion
- voluntarily restraint by algorithms in endpoints

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



(VoIP, VoD)

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



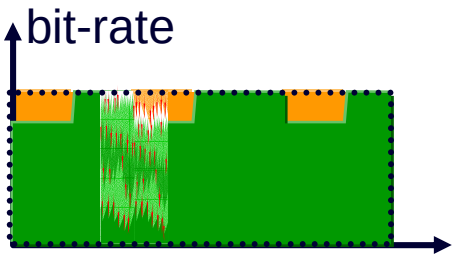
or for much longer than anyone else (p2p file-sharing x200)

- net effect of both (p2p: x1,000-20,000 higher traffic intensity)

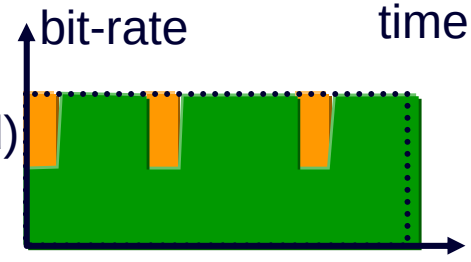
none of these

harness end-system flexibility

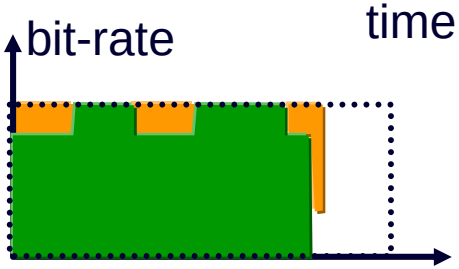
1. TCP



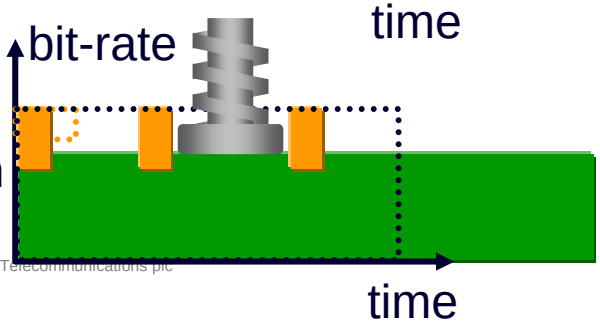
2. (weighted) fair queuing

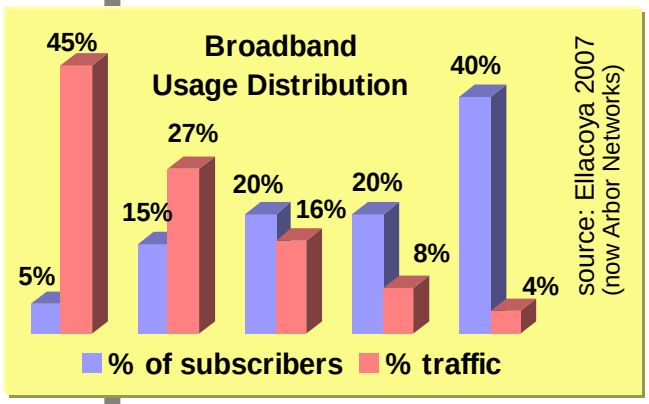


3. volume caps



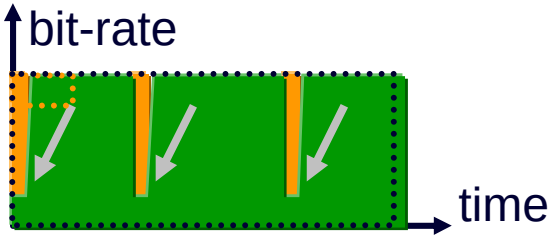
4. deep packet inspection (DPI)

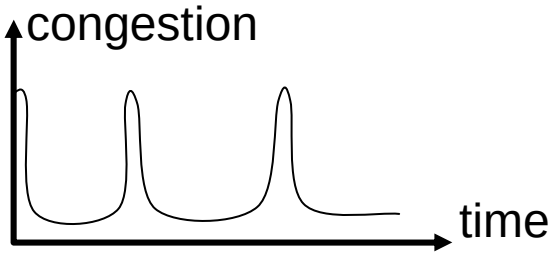




simpler & better...

weighted TCP 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

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