QoS interconnect best without effort

Bob Briscoe
Chief Researcher
BT
Sep 2009
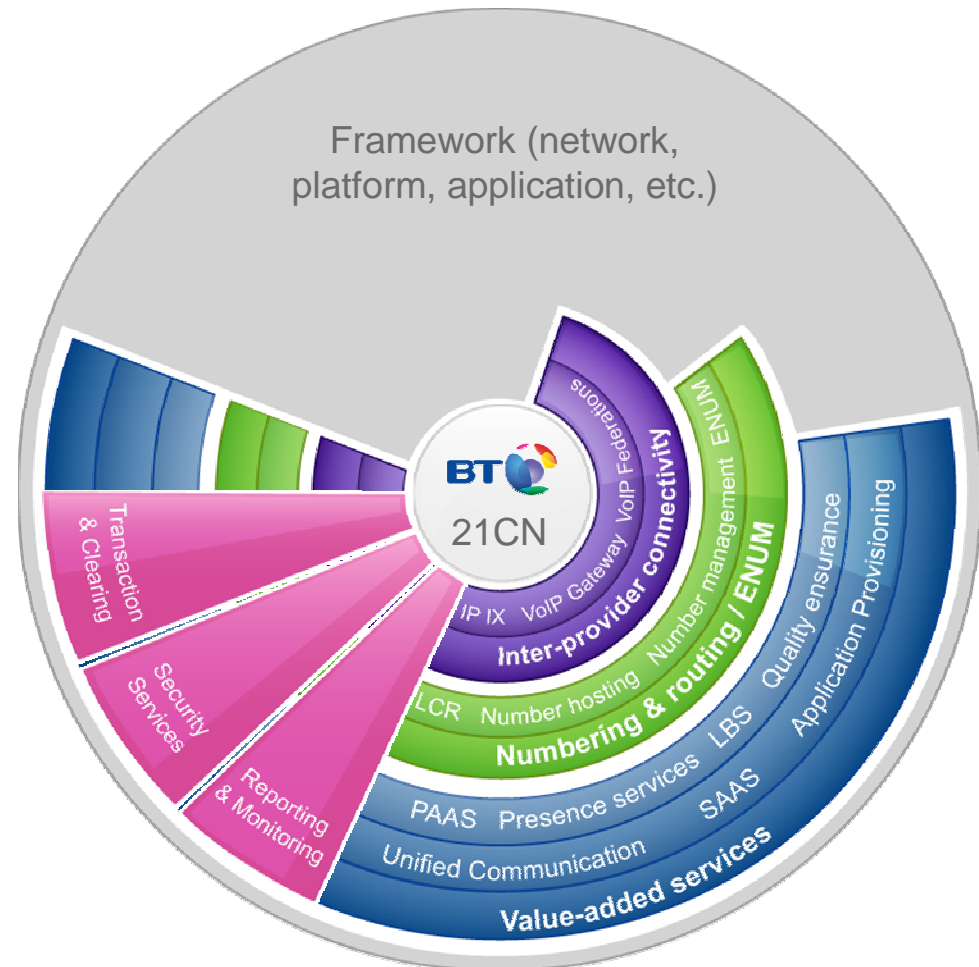
BT's Interoperability: Bespoke approach to customers

- Connect to one, connect to many

- BT provides transaction and clearing, security, reporting and monitoring to optimise all services on each layer

- BT provides a building-block approach to match all CP and SP needs from inter-provider connectivity to value-added services

- BT provides specific services based on customers' needs, allowing them to pick and choose from the different layers what fits their market ambitions

Complete range of IP-based services relying on BT 21CN and BT Global IPX in an open and flexible environment where CPs and SPs can seamlessly select BT IP Interoperability services adapted to their needs and benefit from the IP convergence world

Framework (network, platform, application, etc.)

BT 21CN

Transaction & Clearing

Security Services

Reporting & Monitoring

VoIP Federations
IP IX VoIP Gateway
Inter-provider connectivity

Number management ENUM
Numbering & routing / ENUM
LCR Number hosting

Quality assurance
Application Provisioning
LBS
Presence services SAAS
PAAS
Unified Communication
Value-added services

BT

BT's Interoperability – IP Exchange (IPX)

- IP Exchange is at the heart of BT's Interoperability portfolio.

- IP Exchange enables communication between different infrastructures handling voice and data traffic to ensure that users can communicate to each other irrespective of
  - the service
  - the location
  - the device or network

- IP Exchange ensures that all parties involved in a service usage can be billed and charged appropriately
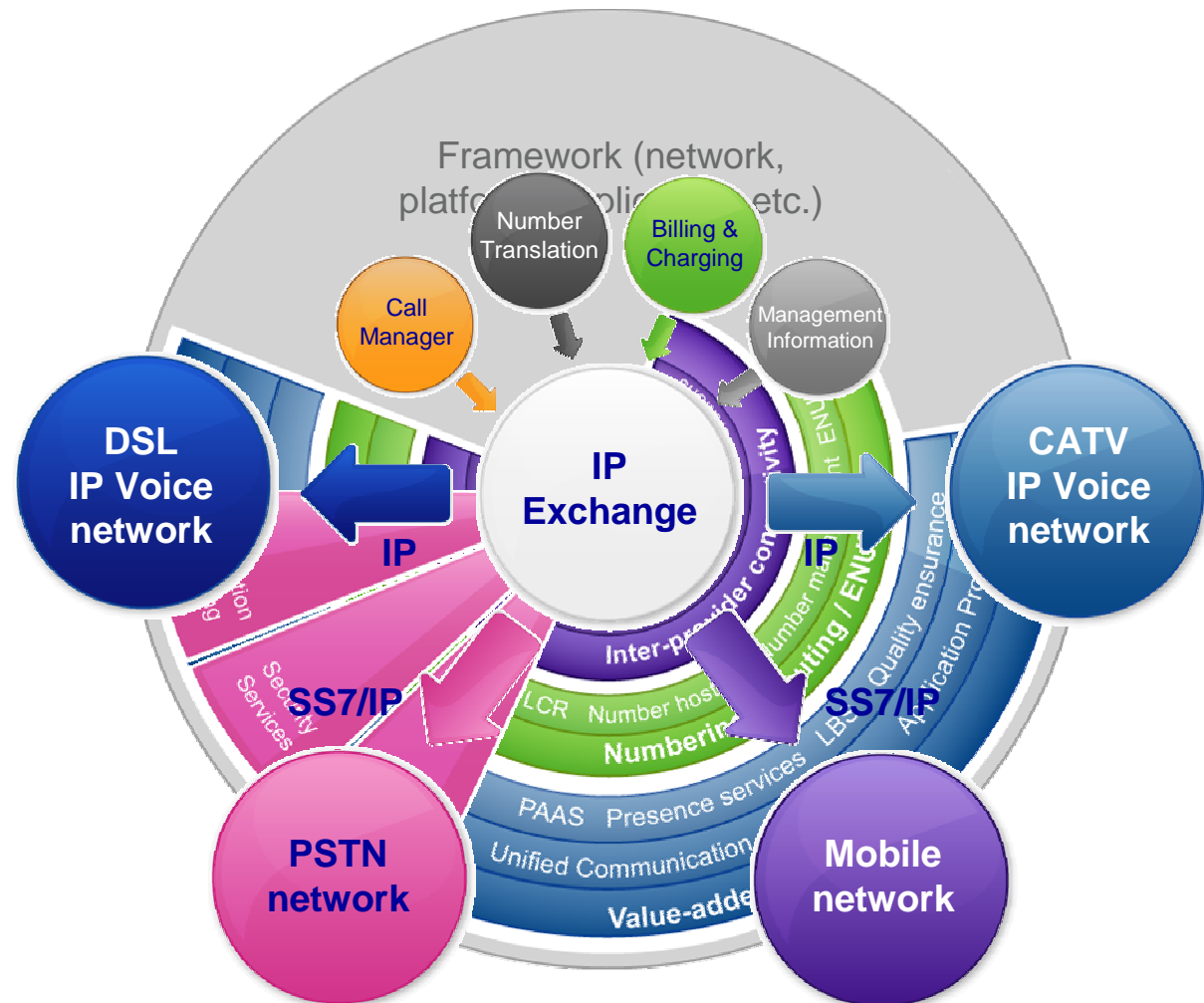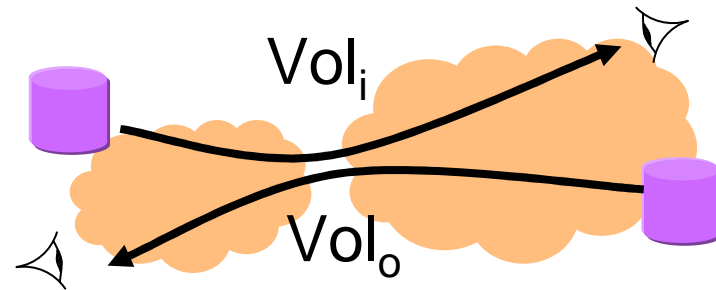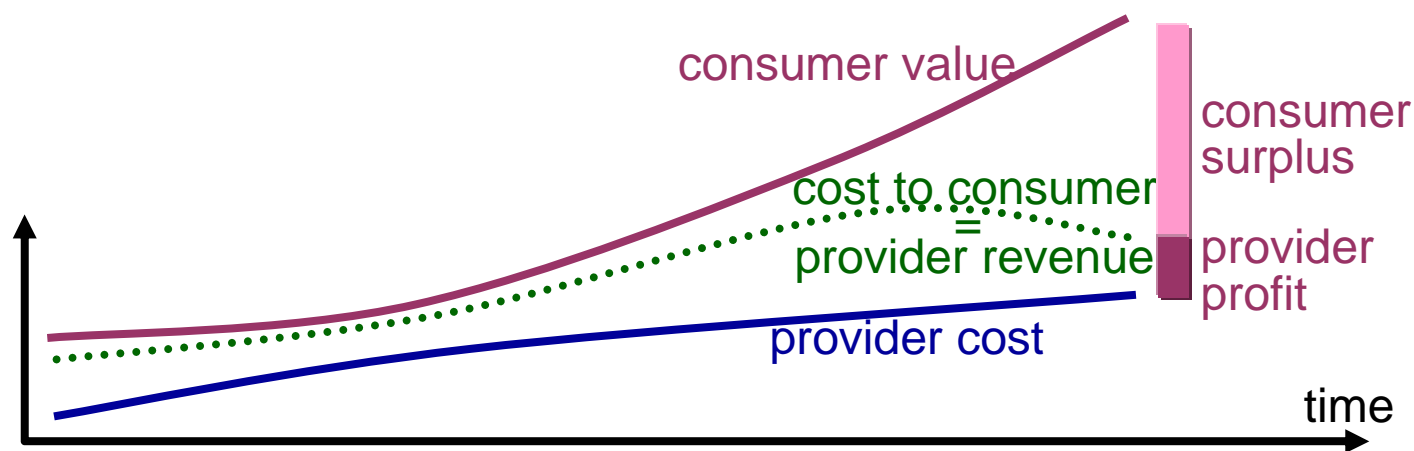
QoS interconnect

best without effort

coming soon…?

both value *and* cost

Vol_i

Vol_o

- industry contractual metrics are largely value-based
  - e.g. volume ratio, session instances
  - 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

consumer value

cost to consumer
=
provider revenue

provider cost

consumer surplus

provider profit

time

BT

© British Telecommunications plc

the big idea

- **pool all traffic together**
  - real-time, p2p file-sharing, Web, streaming, …
- **keep heavy traffic from harming QoS sensitive apps**
  - through economic incentives backed by enforcement

- **once cost-based mechanisms correct for *all* traffic**
  - serves as cost-based floor for value-based services

BT

congestion is not evil

congestion *signals* are healthy

- **no congestion across whole path is evil**
  - – for data transfer to complete ASAP, must fill bottlenecks
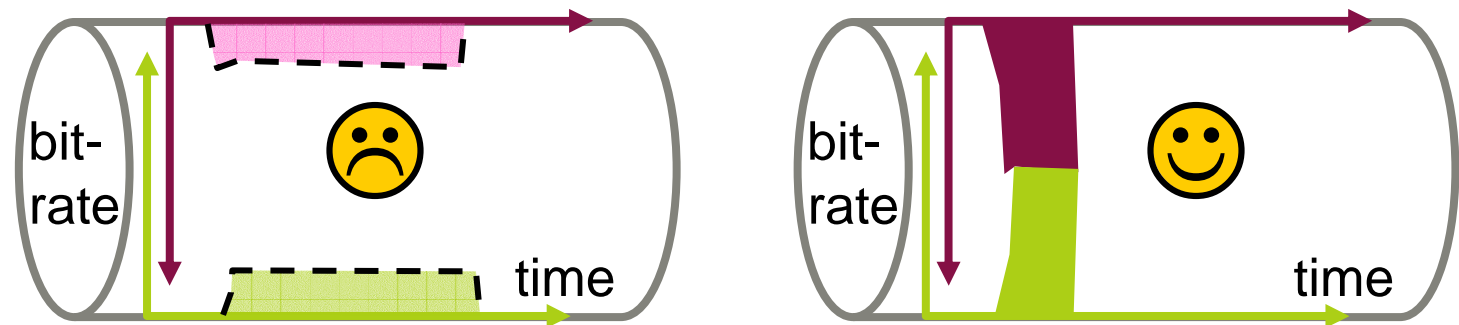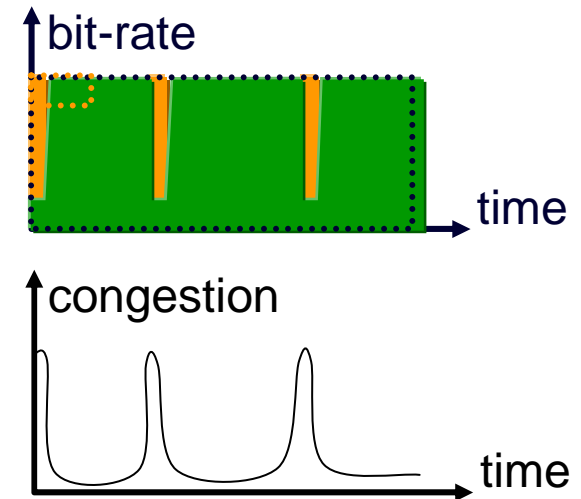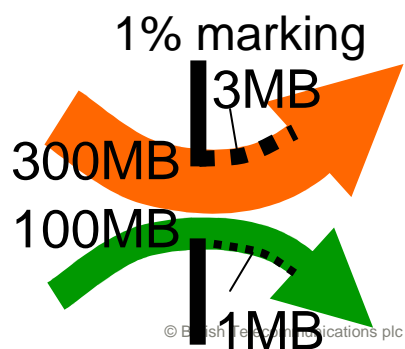

the trick

signal congestion *just before* impairment

- – explicit congestion notification (ECN)
  - • 2001 update to IP: as a queue builds mark more packets
- – then tiny queuing delay and tiny tiny loss for all traffic

marginal cost of network usage?

bit-rate

time

congestion

time

- volume is NOT a good measure

- heavy user yields at early congestion onset
  - very high volume but very low cost to others
  - e.g. LEDBAT (BitTorrent's or Microsoft's low extra delay background transport) or weighted TCP
- by counting volume, operators kill nice behaviour

- correct measure: congestion-volume
  - your contribution to congestion
  - = bytes marked

1% marking
3MB
300MB
100MB
1MB

0.01% marking
10GB
1MB

BT

Axis labels: **'Cost':** Congestion-Volume: Total TCP Loss [Byte] (y-axis), **Volume:** Total TCP Traffic Volume [Byte] (x-axis)

☑ congestion-volume
user's contribution to congestion

WARNING:
Requires validation
with more sample data

100%

☒ volume

Initial results
measured on Naples Uni net
Each point is a user
correlation coefficient: 0.43
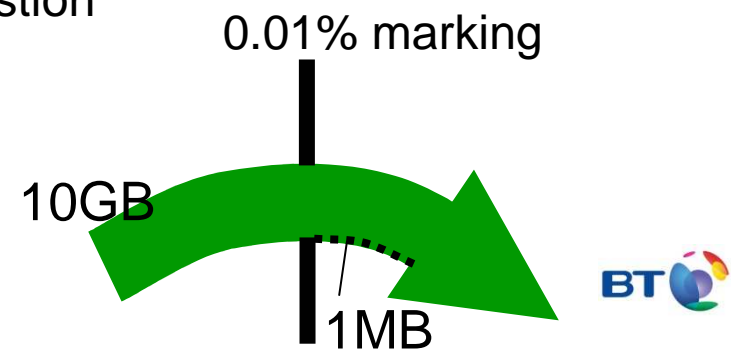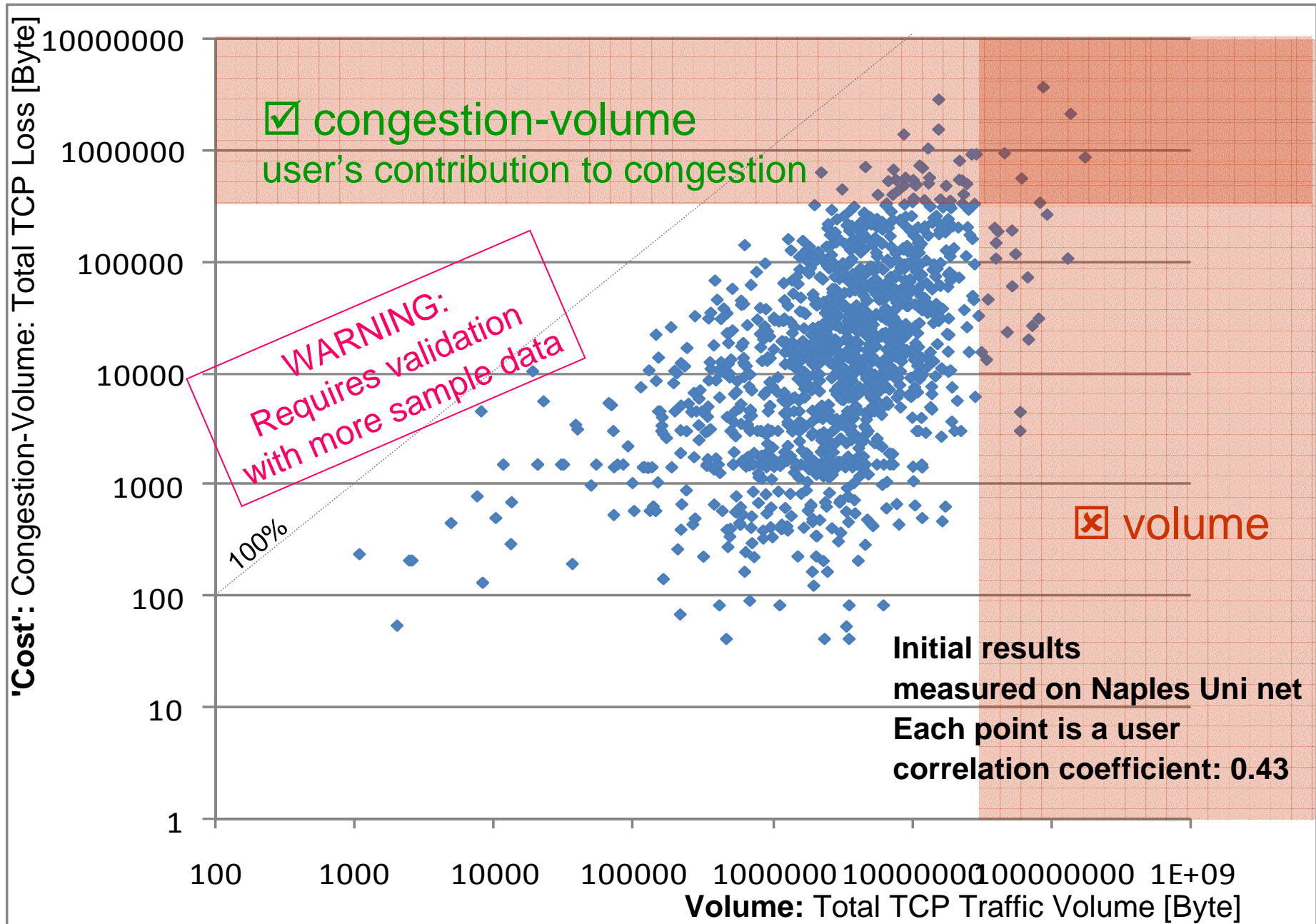
dual demand & supply role

- 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
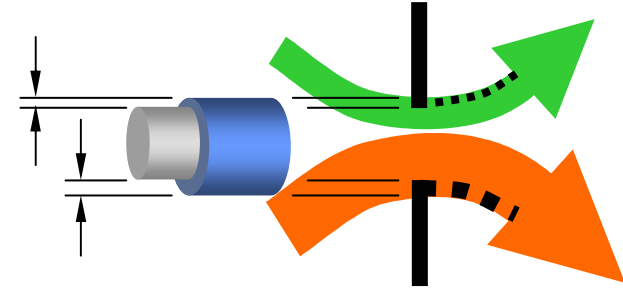
congestion exposure

- by Internet design, endpoints detect & handle congestion
  - v hard for networks to see losses (the marginal cost)
- explicit congestion notification (ECN) not enough
  - visible, but too late – as packets leave the network
- proposed IETF working group: "congestion exposure"
  - re-ECN: *sender* marks IP headers to expose congestion
  - to measure traffic cost as easily as we measure volume
  - just count volume of marked packets in aggregate
- >40 significant offers of help just in last fortnight
  - Microsoft, Nokia, Cisco, Huawei, Alcatel-Lucent, NEC, Ericsson, NSN, Sandvine, Comcast, DT, Verizon, …

example consumer use of exposed congestion
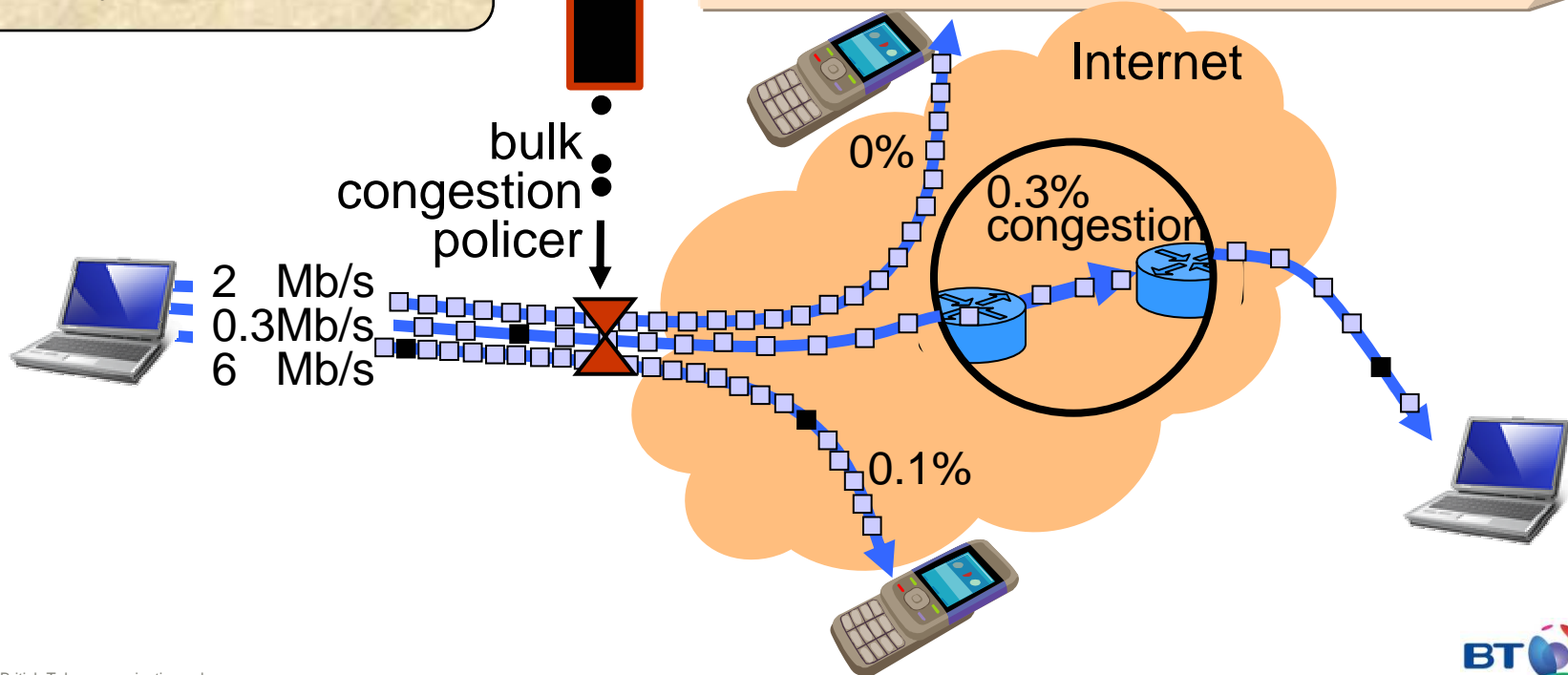## still with flat fee

Acceptable Use Policy

'congestion-volume'
allowance: 1GB/month

@ €15 / month

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

- incentive for apps to avoid congestion

- backed by enforcement

bulk
congestion
policer

2 Mb/s
0.3Mb/s
6 Mb/s

Internet

0%

0.3%
congestion

0.1%

BT

reveals bulk marginal cost 'routing money'

legend:

re-ECN downstream congestion marking [%]

bit rate
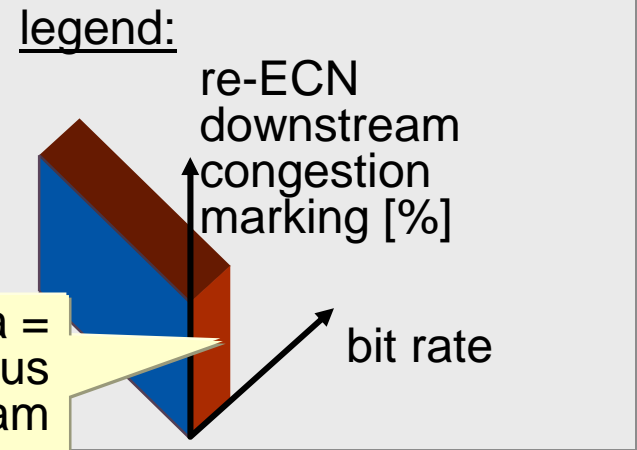
area = instantaneous downstream congestion volume

N_A

N_B

N_C

N_D

€

€

€

0|0|2|7|6|0|5

just two counters at border meter monthly bulk congestion-volume

= true marginal cost

without measuring flows

BT
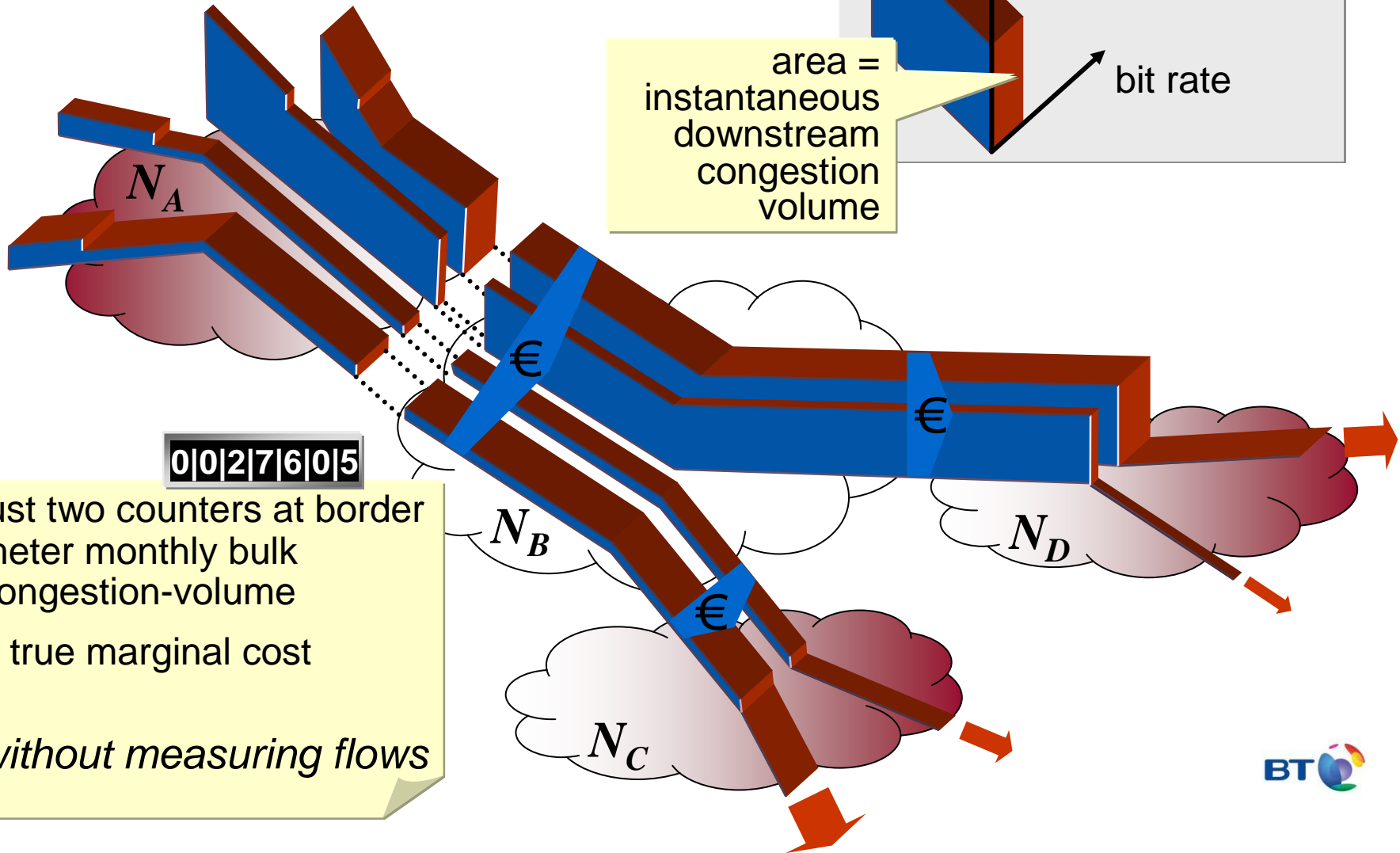
I'm a conservative, get me out of here!

- 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

guaranteed bit-rate?
or much faster 99.9% of the time?

harnessing flexibility

constant quality video encoding (qp=28)
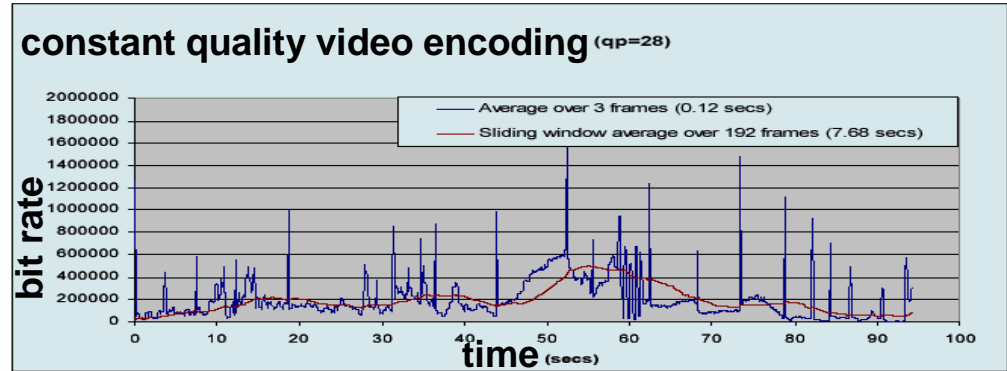
- the idea that humans want to buy a known fixed bit-rate
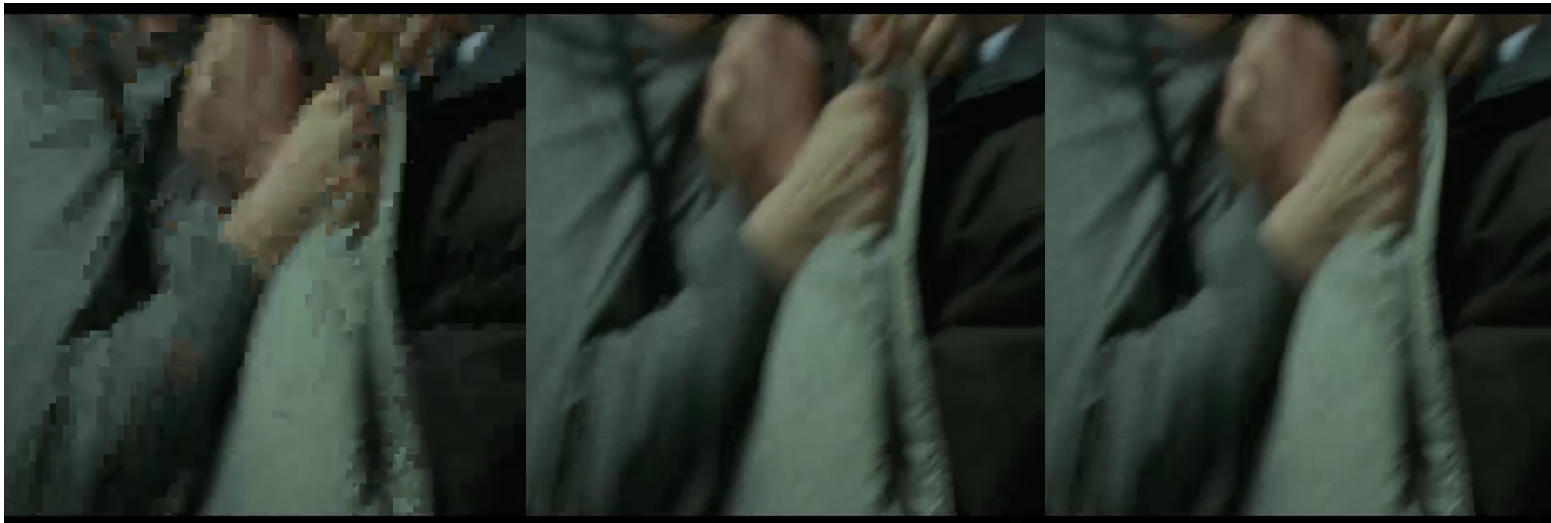 - comes from the needs of media delivery technology
  - hardly ever a human need or desire

- 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

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

Constant Bit Rate **100%**   Constant Quality **125%**   Equitable Quality **216%**
sequences encoded at same average of 500kb/s   [Crabtree09]

bringing cost 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

open

closed

Internet

telco /NGN

cable

cellular

satellite

1995

2009

Internet

BT
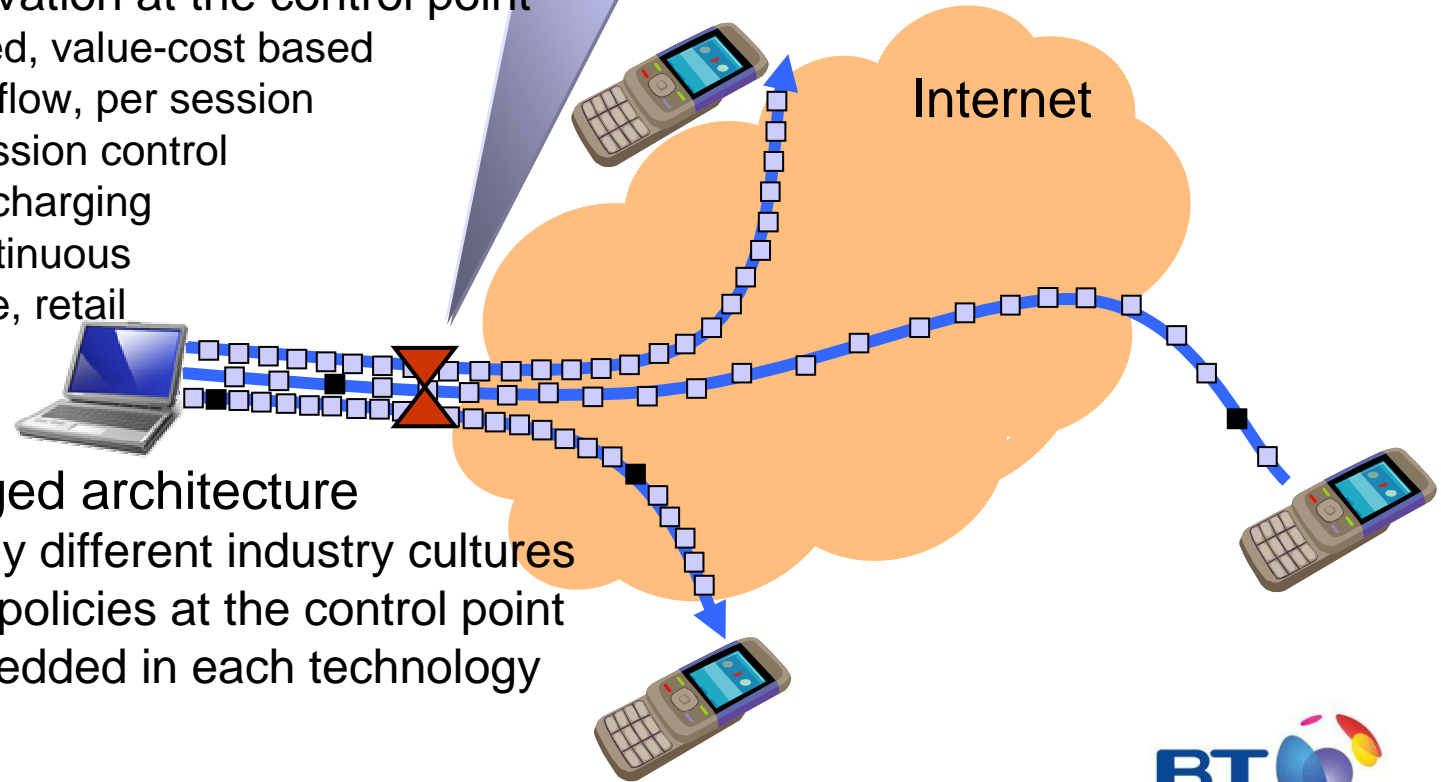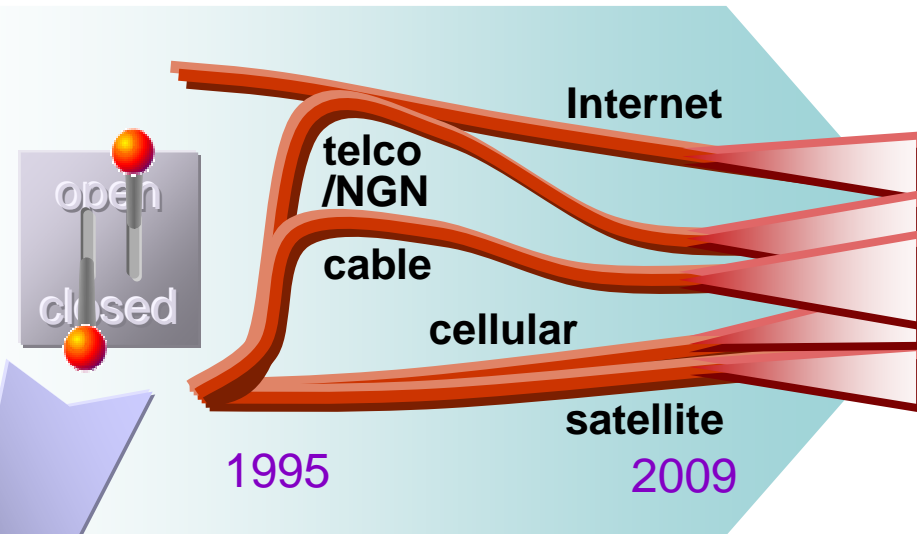
best without effort

- 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
    - visible 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
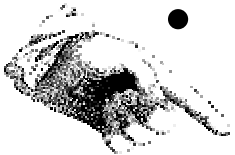
more info...

- **BT IP Exchange**
  - jens.specht@bt.com

- **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)
- **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

- **Congestion Exposure (re-ECN) at the IETF**
  - <trac.tools.ietf.org/area/tsv/trac/wiki/re-ECN>

BT

QoS interconnection

best without effort

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

marked ACK

ACKnowledgement packets

packet headers

network
transport

data

data

**probabilistic
packet marking algorithm
on all egress interfaces**

probability

1

drop

mark

**ave queue
length**

marked packet

		0	5 6 7
00:	**Not ECN Capable Transport (ECT)**		
01 or 10:	**ECN Capable Transport - no Congestion Experienced (sender initialises)**	DSCP	ECN
11:	**ECN Capable Transport - and Congestion Experienced (CE)**		

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

congestion exposure in one bit

Diff serv **ECN**

R E

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

Feedback path

Networks

Routers

+1   +1   +1   -1   +1   -1

Sender

Data packet flow

Receiver

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

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