Internet QoS
the underlying economics

Bob Briscoe
Chief Researcher, BT Group
Sep 2008

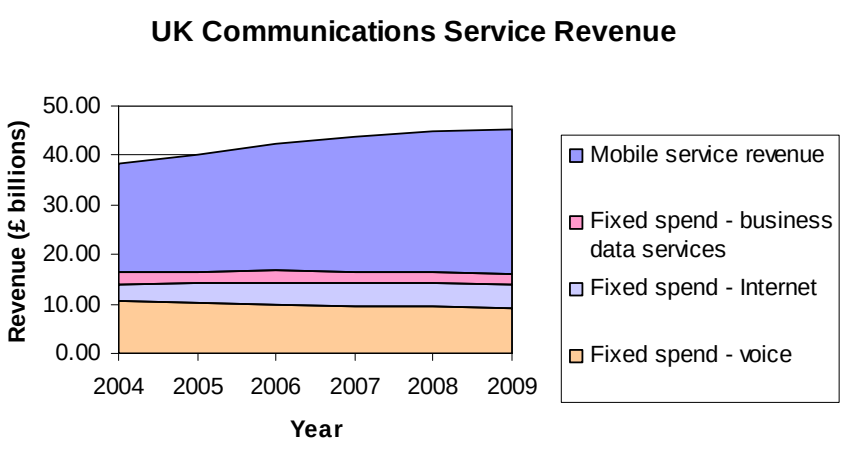
executive summary

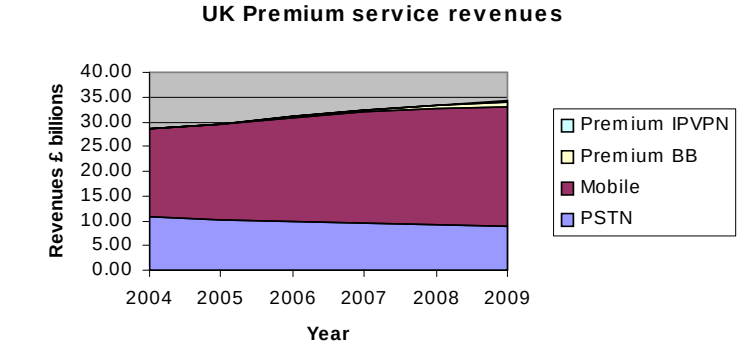
congestion accountability – the missing link

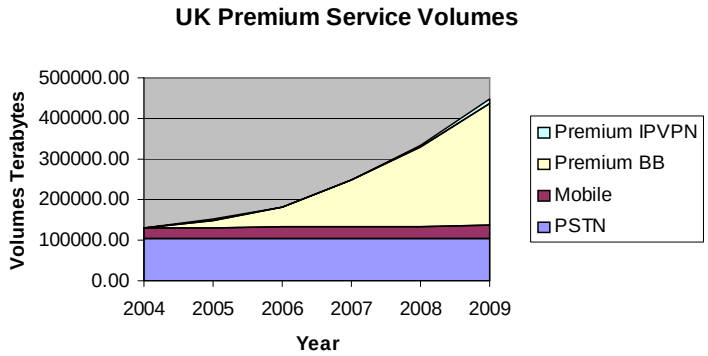
- unwise NGN obsession with per-session QoS guarantees
- scant attention to competition from 'cloud QoS'
  - rising general QoS expectation from the public Internet
  - cost-shifting between end-customers (including service providers)
  - questionable economic sustainability
- 'cloud' resource accountability is possible
  - principled way to heal the above ills
  - requires shift in economic thinking – from volume to congestion volume
- provides differentiated cloud QoS without further mechanism
- also the basis for a far simpler per-session QoS mechanism
  - having fixed the competitive environment to make per-session QoS viable

BT

QoS: value ≠ cost





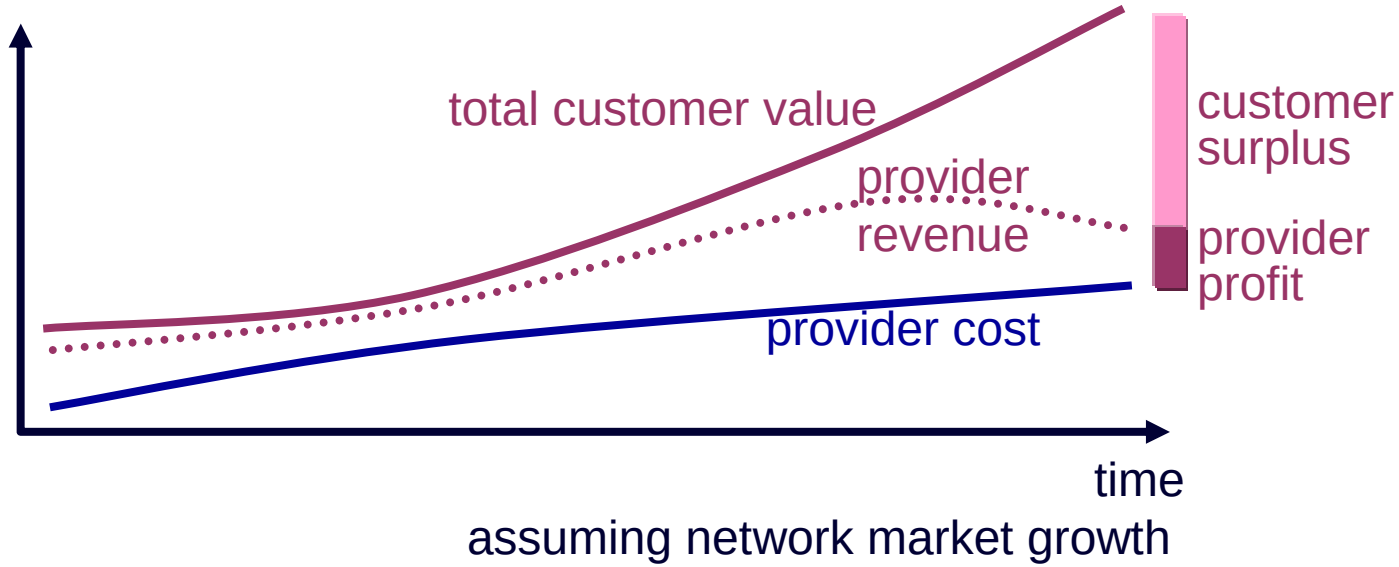


- definition of 'premium'
  - services *requiring* better than normal QoS
  - not necessarily *using* network QoS mechanisms (e.g. VoIP)

Sources: Analysys Research (2005) and S Rudkin, BT internal report (Oct 2005)

remember... competition

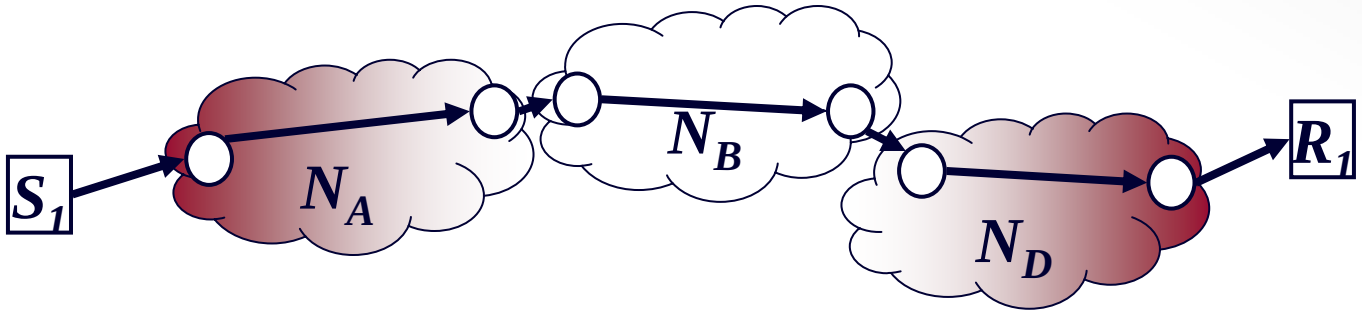
- drives revenue towards cost
- eventually ensures customers get the surplus value
  - not providers



Internet QoS
first, fix cost-based accountability

Bob Briscoe

capacity costs

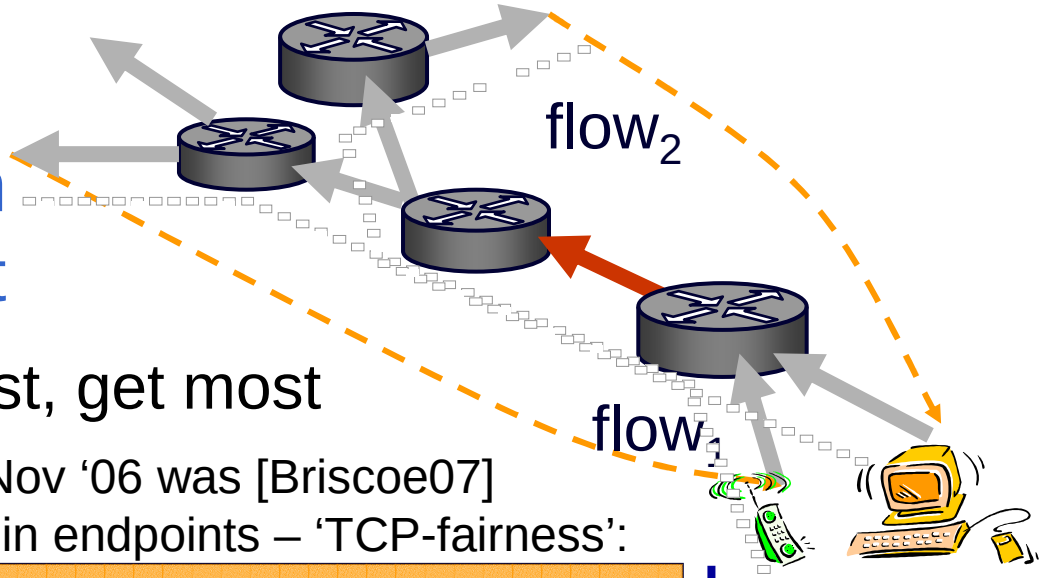


- selling QoS = managing risk of congestion
  - if no risk of congestion, can't sell QoS
  - congestion risk highest in access nets (cost economics of fan-out)
  - also *small* risk in cores/backbones (failures, anomalous demand)

BT

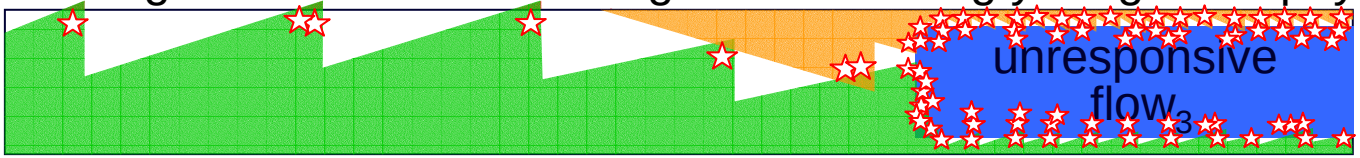
how Internet sharing 'works'

endemic congestion & voluntary restraint



- aka. those who take most, get most
  - technical consensus until Nov '06 was [Briscoe07] voluntarily polite algorithm in endpoints – 'TCP-fairness':

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



(VoIP, VoD Joost 700kbps)

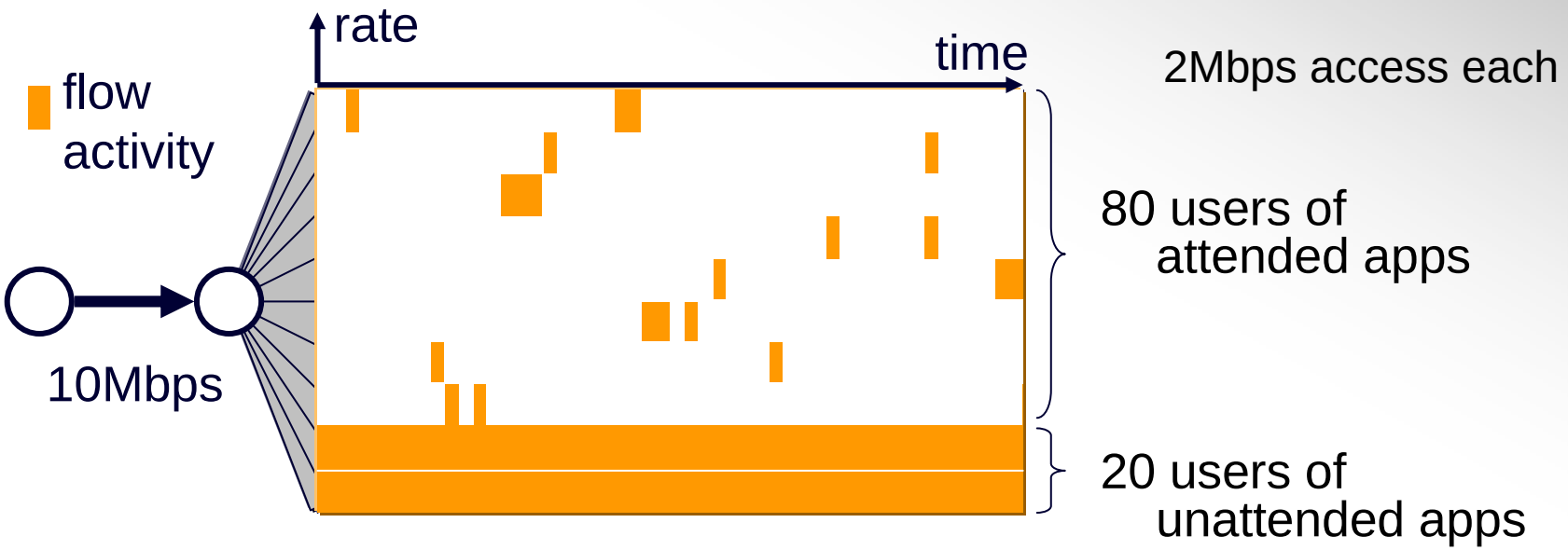
  - or starting more 'TCP-fair' flows than anyone else (Web: x2, p2p: x5-100)

  - or for much much longer than anyone else (p2p file-sharing x200)
  - net effect of both (p2p: x1,000-20,000 higher traffic intensity) [Briscoe08]

BT

TCP's broken resource sharing

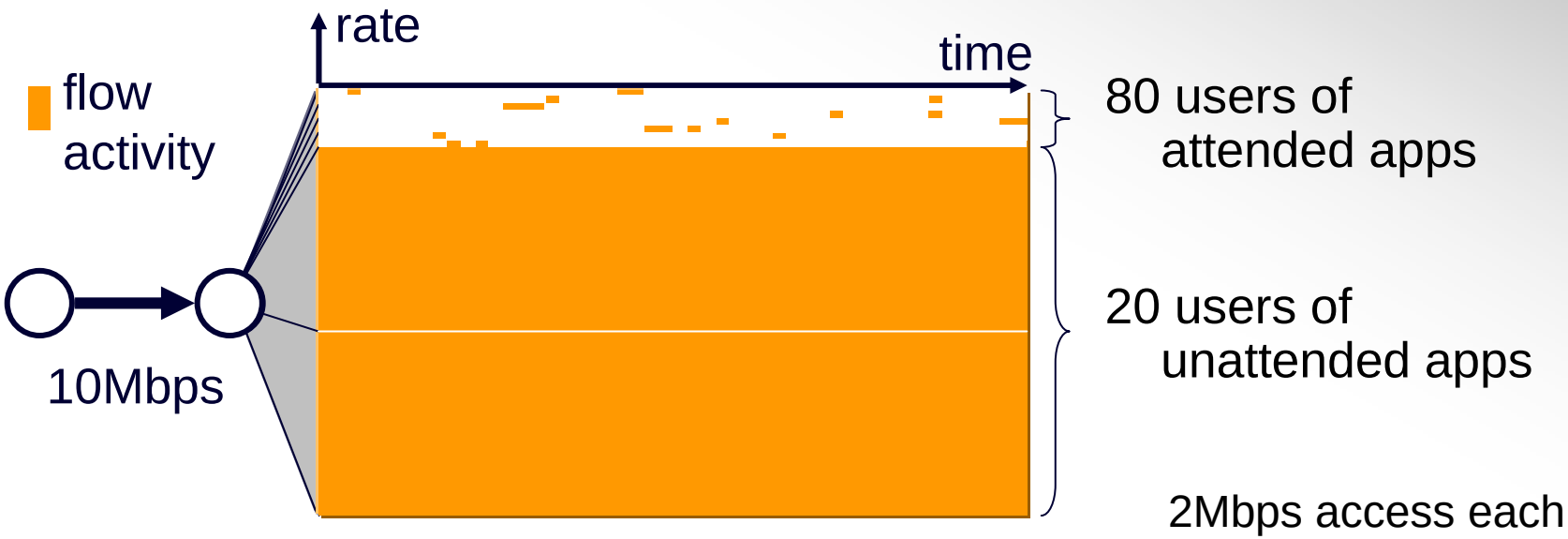
base example: different activity factors



usage type	no. of users	activity factor	ave.simul flows /user	TCP bit rate /user	vol/day (16hr) /user	traffic intensity /user
attended	80	5%	=	417kbps	150MB	21kbps
unattended	20	100%	=	417kbps	3000MB	417kbps
				x1	x20	x20

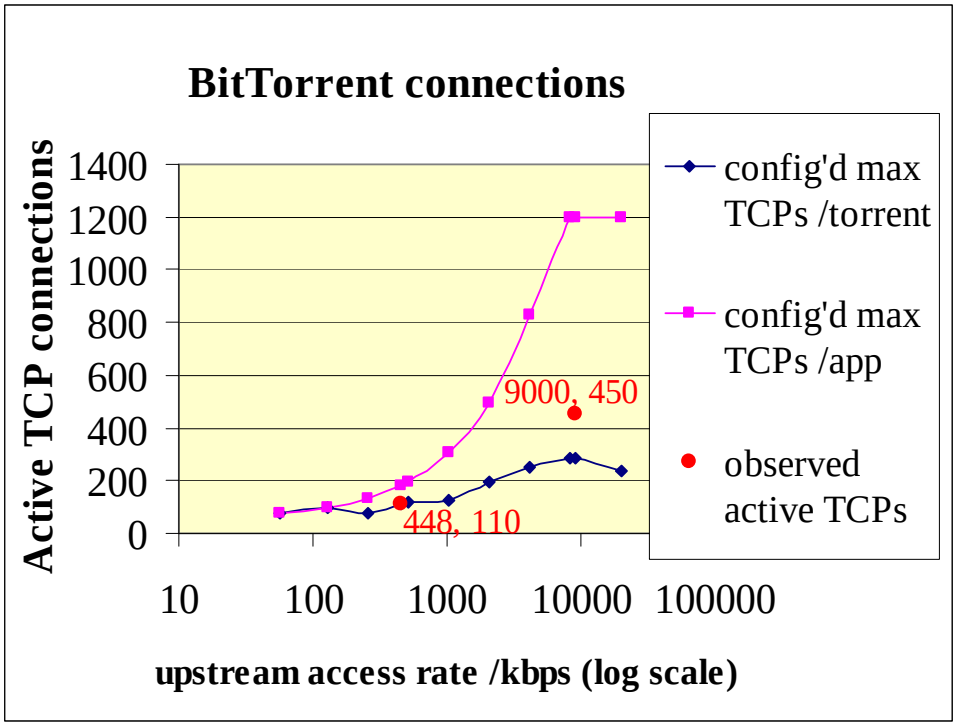
TCP's broken resource sharing

compounding activity factor & multiple flows



usage type	no. of users	activity factor	ave.simul flows /user	TCP bit rate /user	vol/day (16hr) /user	traffic intensity /user
attended	80	5%	2	20kbps	7.1MB	1kbps
unattended	20	100%	50	500kbps	3.6GB	500kbps
				x25	x500	x500

realistic numbers?
there are elephants in the room



- number of TCP connections
  - Web1.1: **2**
  - BitTorrent: **5-100** observed active
    - varies widely depending on
      - no. of torrents per user
      - maturity of swarm
      - config'd parameters

details suppressed:

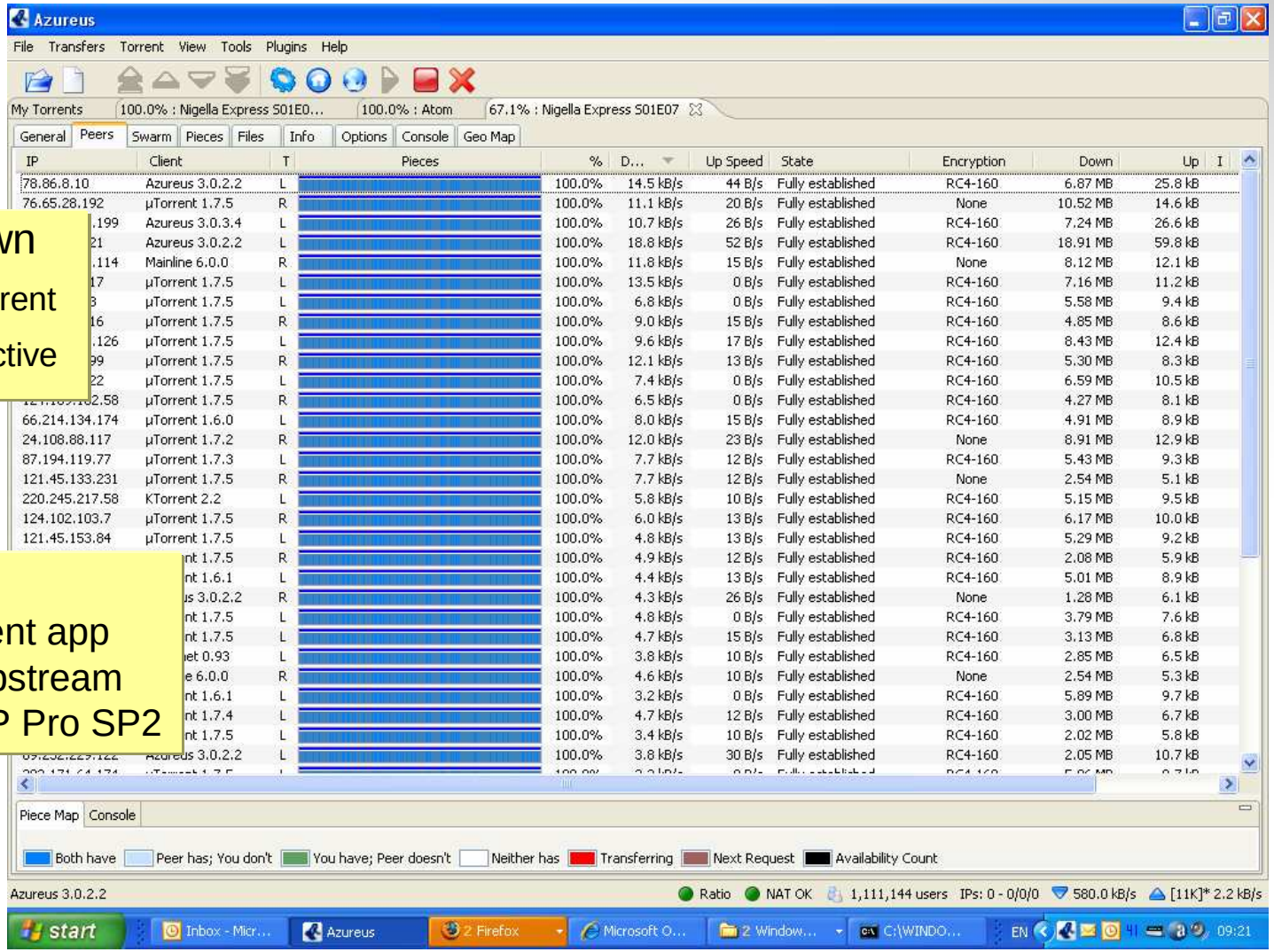
- utilisation never 100%
  - but near enough during peak period
- on DSL, upstream constrains most p2p apps
  - other access (fixed & wireless) more symmetric

typical p2p file-sharing apps

- 105-114 active TCP connections altogether

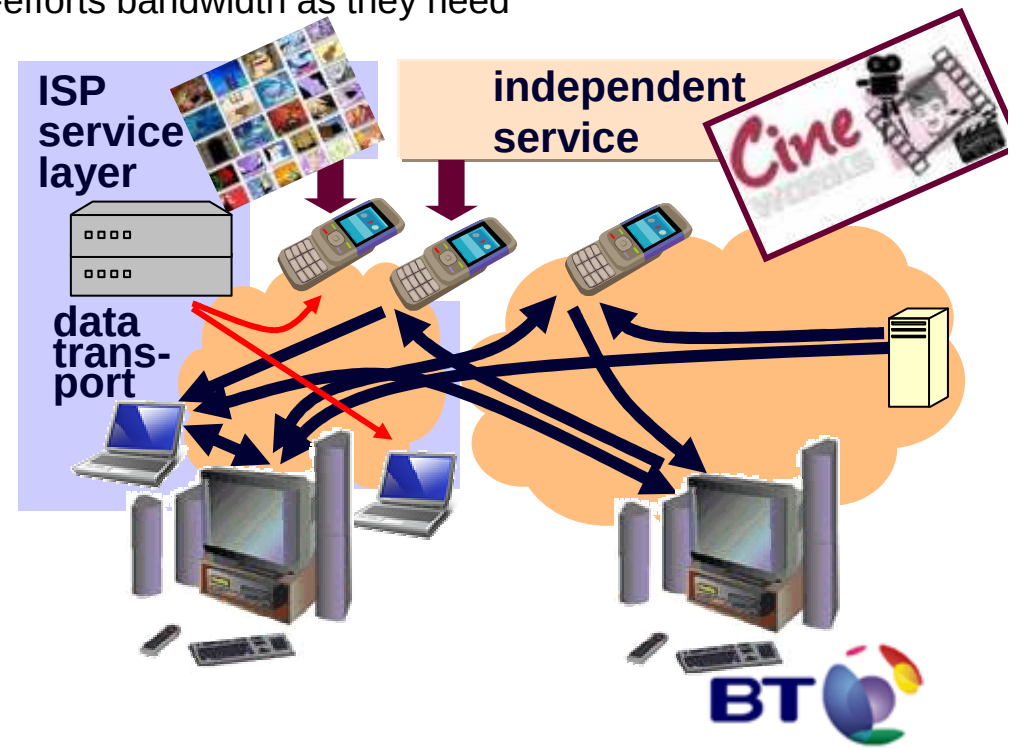
1 of 3 torrents shown
- ~45 TCPs per torrent
- but ~40/torrent active

environment
Azureus BitTorrent app
ADSL+ 448kb upstream
OS: Windows XP Pro SP2

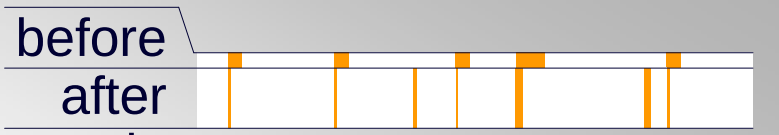


cost-shifting between services

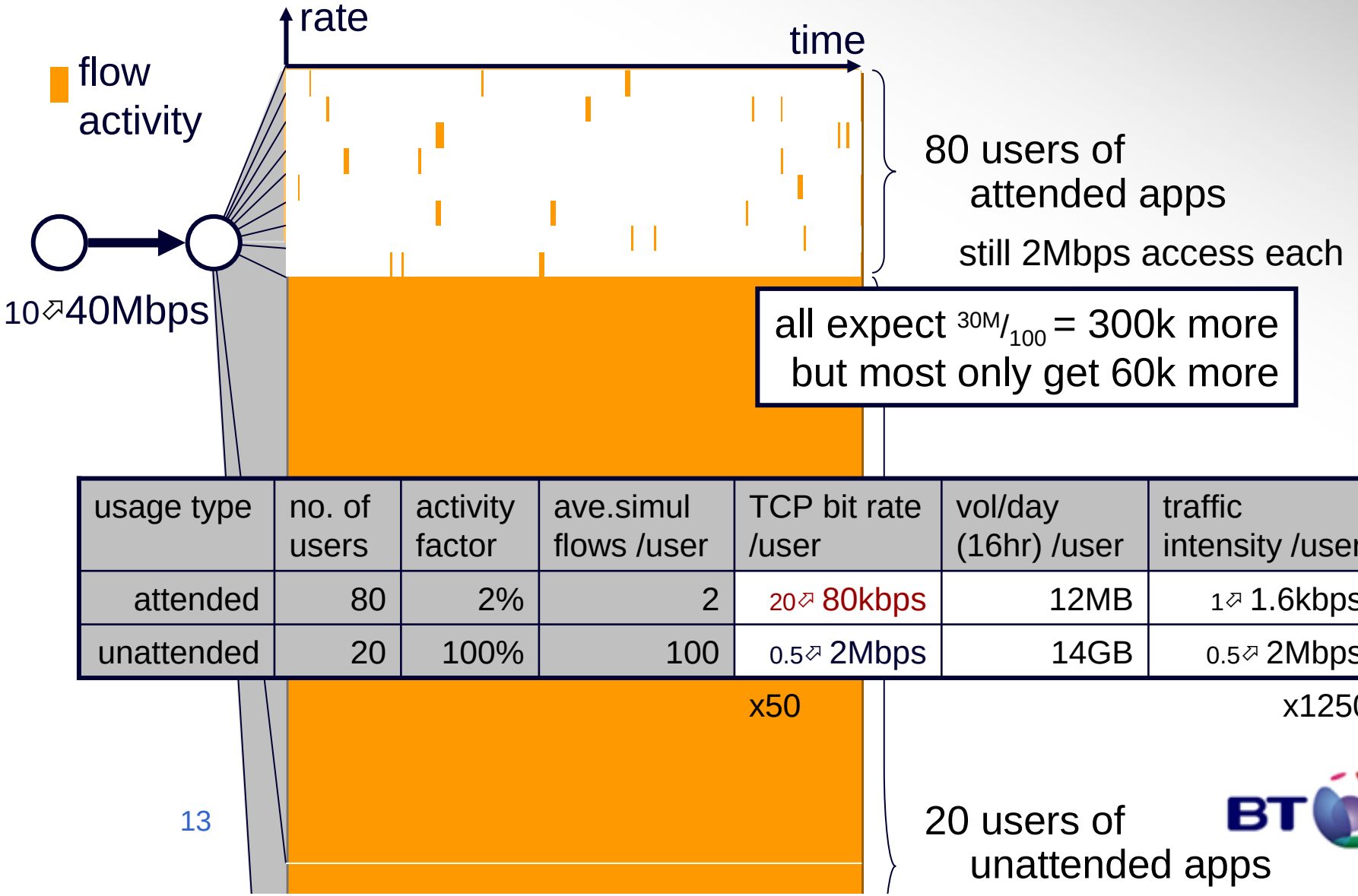
- scenario
  - ISP also a higher level service provider (TV, video phone, etc)
  - competing with independent service providers (Skype, YouTube, etc)
- capacity & QoS costs for high value services
  - ISP buys capacity & QoS internally
  - independent SP just takes as much best-efforts bandwidth as they need
  - because of how Internet sharing 'works'
- cost of heavy usage service subsidised by ISP's lighter users



most users hardly benefit from bottleneck upgrade



data limited flows
want rate more than volume

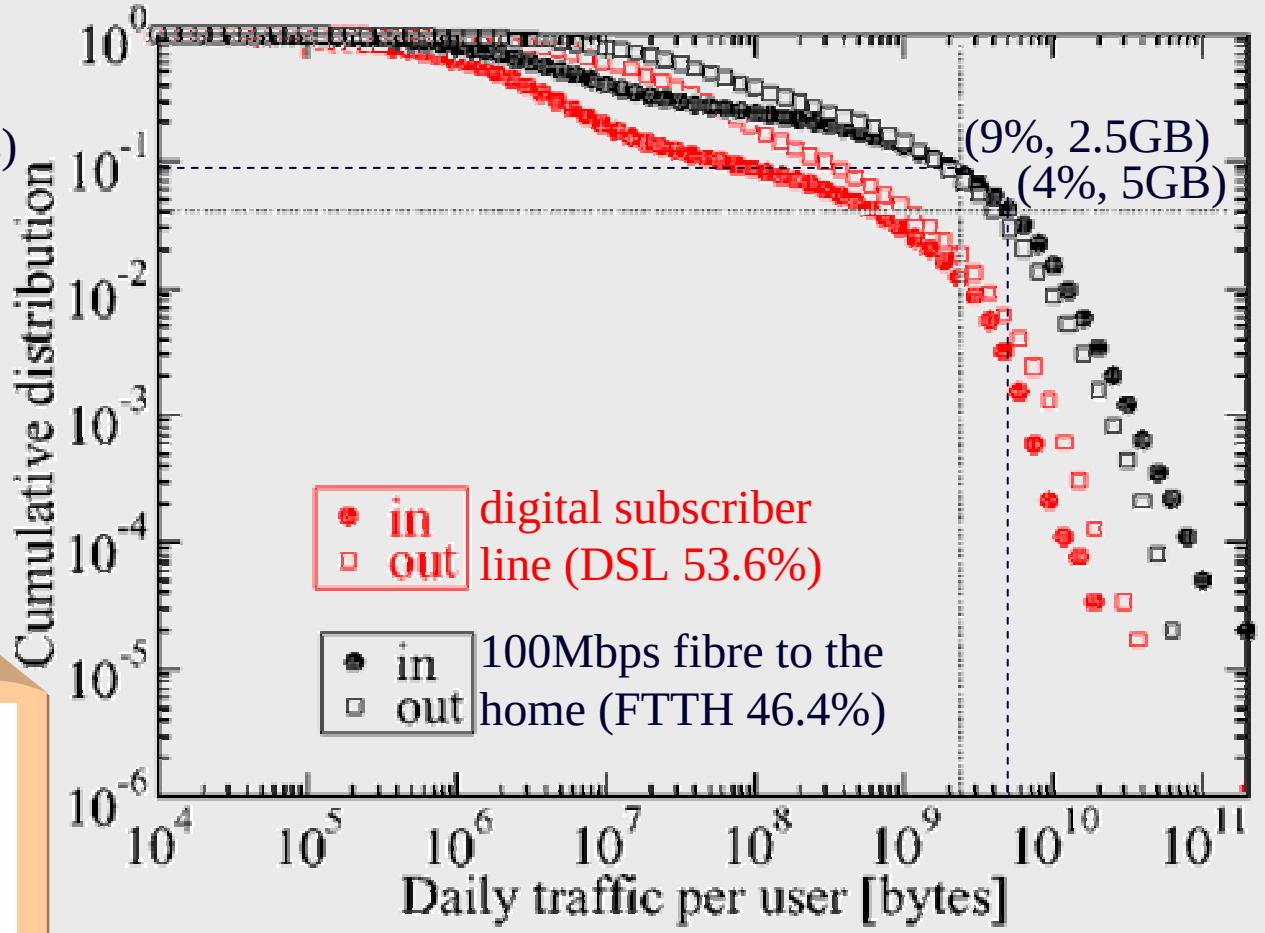


usage type	no. of users	activity factor	ave.simul flows /user	TCP bit rate /user	vol/day (16hr) /user	traffic intensity /user
attended	80	2%	2	20⇗ 80kbps	12MB	1⇗ 1.6kbps
unattended	20	100%	100	0.5⇗ 2Mbps	14GB	0.5⇗ 2Mbps

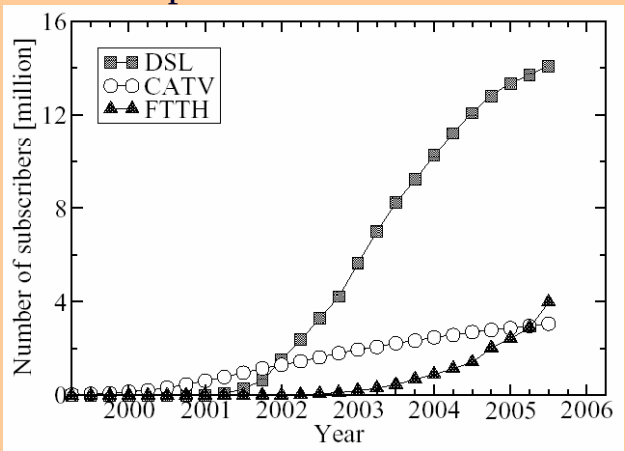
p2p quickly fills up fibre to the home

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

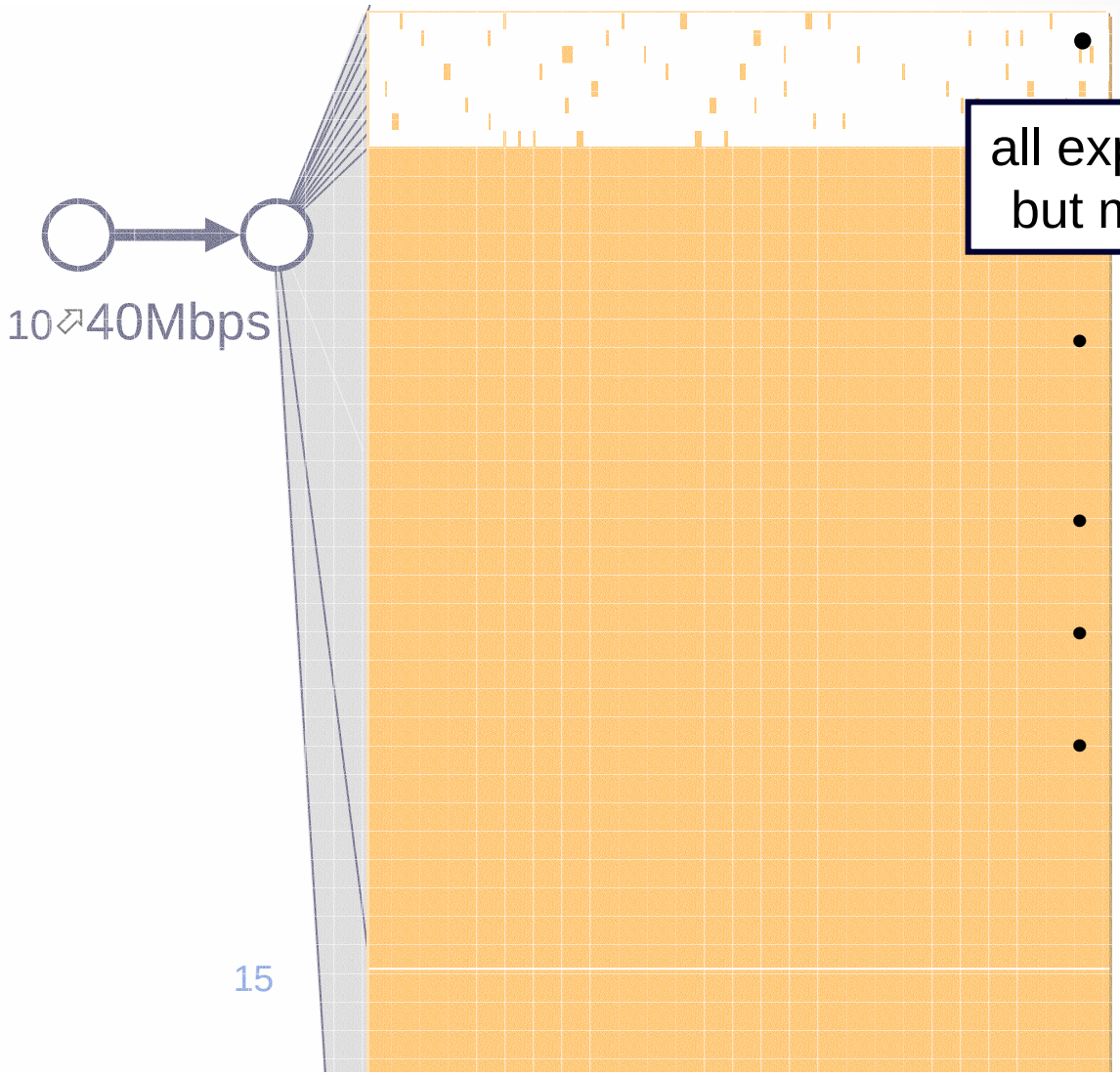


Courtesy of Kenjiro Cho et al [Cho06]
The Impact and Implications of the Growth in Residential User-to-User Traffic, SIGCOMM (Oct '06)

BT

consequence #1

higher investment risk



- recall

all expect $^{30M}/_{100}$ = 300k more
but most only get 60k more

- but ISP needs everyone to pay for 300k more
- if most users unhappy with ISP A's upgrade
- they will drift to ISP B who doesn't invest
- competitive ISPs will stop investing...

BT

consequence #2

trend towards bulk enforcement

- as access rates increase
  - attended apps leave access unused more of the time
  - anyone might as well fill the rest of their own access capacity
- operator choices:
  a) either continue to provision sufficiently excessive shared capacity
  b) or enforce tiered volume limits

---

see joint industry/academia (MIT) white paper “Broadband Incentives” [BBincent06]

BT

consequence #3

networks making choices for users

- characterisation as two user communities over-simplistic
  - heavy *users* mix heavy and light *usage*
- two enforcement choices

  a) bulk: network throttles all a heavy user's traffic indiscriminately
  - encourages the user to self-throttle least valued traffic
  - but many users have neither the software nor the expertise

  b) selective: network *infers* what the user would do
  - using deep packet inspection (DPI) and/or addresses to identify apps
- even if DPI intentions honourable
  - confusable with attempts to discriminate against certain apps
  - user's priorities are task-specific, not app-specific
  - customers understandably get upset when ISP guesses wrongly

BT

DPI: de facto standard QoS mechanism

- for many ISPs ‘network processing’ boxes are central to QoS
- but DPI fights the IP architecture, with predictably poor results

	the Internet way (TCP)	operators (& users)
degree of freedom	‘flow rate equality’	‘volume accounting’
multiple flows	✗	✓
activity factor	✗	✓
application control	✓	✗
congestion variation	✓	✗

- DPI can only work if it can infer customer priorities from the app
- QoS with no API and only a ‘busy-period’ notion of congestion

BT

underlying problems
blame our choices, not p2p

- commercial
  - Q. what is cost of network usage?
  - A. volume? NO; rate? NO
  - A. 'congestion volume'
  - *our own* unforgivable sloppiness over what our costs are
- technical
  - lack of cost accountability in the Internet protocol (IP)
  - p2p file-sharers exploiting loopholes in technology *we've* chosen
- *we* haven't designed *our* contracts & technology for machine-powered customers

BT

costs

- infrastructure costs: sunk
- operational costs: usage independent
- usage and congestion: cost operator nothing
- congestion: costs those sharing each resource

- approximations to congestion metrics
  1. by time: time-of-day volume pricing
  2. by route: on/off-net, domain hops, distance
  3. by class of service: flat fee for each class, volume price for each class
- accurate congestion metrics (in all 3 dimensions)
  - loss rate
  - explicit congestion notification…

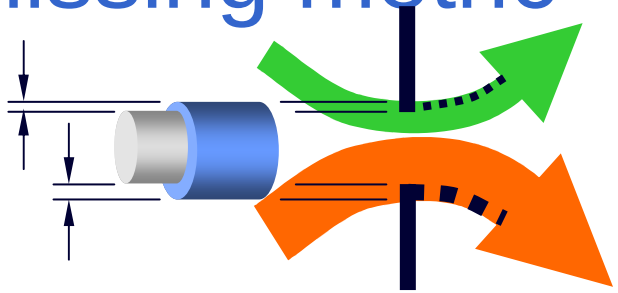
BT

not volume, but
congestion volume: the missing metric

- not ‘what you got’
  but ‘what you unsuccessfully tried to get’
  - proportional to what you got
  - *but also* to congestion at the time

1. congestion volume: cost to other users
2. the marginal cost of upgrading equipment
   - so it wouldn’t have been congested
   - so your behaviour wouldn’t have affected others

- 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

NOTE: congestion volume isn’t an extra cost

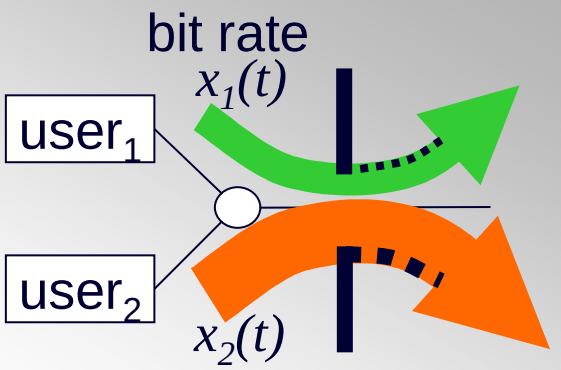
- part of the flat charge we already pay
- it's just the wrong people are paying it
- if we could measure who to blame for it we ***might*** see pricing like this...

access link	congestion volume allow’ce	charge
100Mbps	50MB/month	€15/month
100Mbps	100MB/month	€20/month

BT

core of solution

congestion harm (cost) metric



loss (marking) fraction $p(t)$

- bit rate *weighted* by each flow's congestion, over time

  congestion volume, $v \equiv \int p(t)x_i(t)\,dt$

  summed over all a sender's flows
- result is easy to measure per flow or per user
  - volume of bytes discarded (or ECN marked)
- a precise instantaneous measure of harm, counted over time
  - a measure for fairness over any timescale
  - and a precise measure of harm during dynamics

$$p(t) \equiv \frac{excess\_load(t)^+}{offered\_load(t)}$$

- intuition: volume is bit rate over time

  volume, $V \equiv \int x_i(t)\,dt$

  summed over all a sender's flows
- network operators often count volume only over peak period
  - as if $p(t)=1$ during peak and $p(t)=0$ otherwise

BT

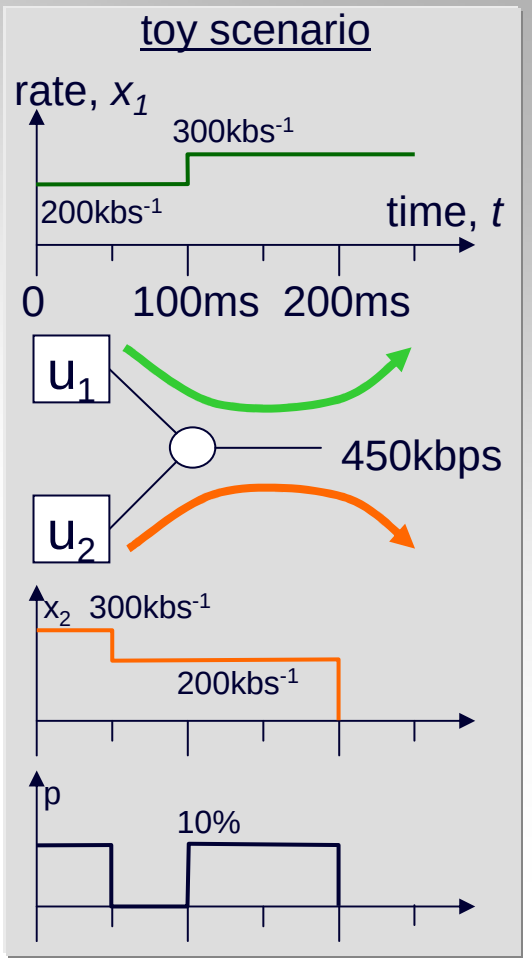
congestion volume metric
toy example

- cost of one user's behaviour on other users
  - congestion volume ≡ instantaneous congestion p...
  - ...shared proportionately over each user's bit rate, x_i
  - ...over (any) time

$$v_i \equiv \int p(t) x_i(t)\, dt$$

$$p(t) \equiv \frac{excess\_load(t)^+}{offered\_load(t)}$$

$$p_{\substack{t=0-50ms,\\ 100-200ms}} = \frac{(200+300)-450}{(200+300)} = 10\%$$



- example

v_1	= 10% x 200kbs^{-1} x 50ms		+ 10% x 300kbs^{-1} x 100ms			
	=	1kb	+	3kb	=	4kb
v_2	= 10% x 300kbs^{-1} x 50ms		+ 10% x 200kbs^{-1} x 100ms			
	=	1.5kb	+	2kb	=	3.5kb

toy scenario for illustration only; strictly...
- a super-linear marking algorithms to determine p is preferable for control stability
- the scenario assumes we're starting with full buffers

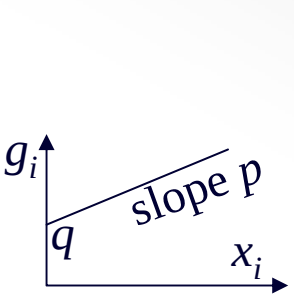
BT

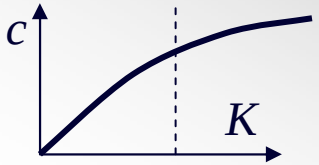
usage vs subscription prices

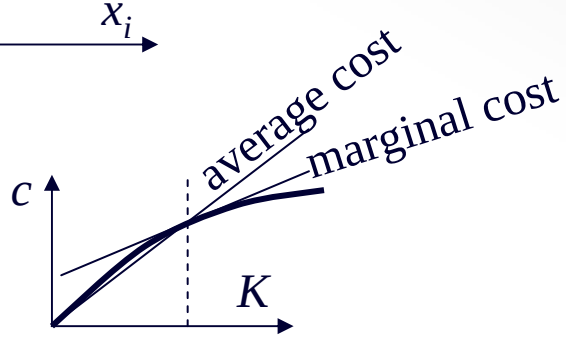
Pricing Congestible Network Resources [MacKieVarian95]

- assume competitive providers buy capacity K [b/s] at cost rate [€/s] of $c(K)$
- assume they offer a dual tariff to customer *i*
  - subscription price q [€/s]
  - usage price p [€/b] for usage x_i [b/s], then

  charge rate [€/s], $g_i = q + px_i$
- what's the most competitive choice of p & q?
- $\frac{\text{usage revenue}}{\text{capacity cost}} = \frac{1}{e}$ where e is elasticity of scale
  - if charge less for usage and more for subscription, quality will be worse than competitors
  - if charge more for usage and less for subscription, utilisation will be poorer than competitors



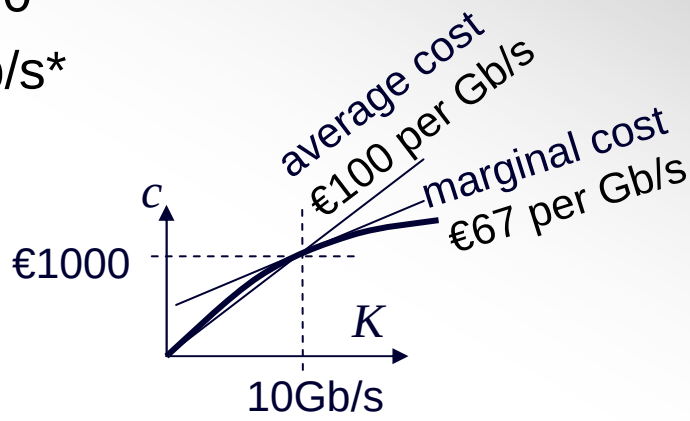




$$e = \frac{\text{average cost}}{\text{marginal cost}} = \frac{c(K)}{K} \cdot \frac{1}{c'(K)}$$

for example

- if a 10Gb/s link interface costs €1000
- and it costs €67 to upgrade to 11Gb/s*
  - average cost = €100 per Gb/s
  - marginal cost ~ €67 per Gb/s



$$e = \frac{\text{average cost}}{\text{marginal cost}} = \frac{3}{2}$$

$$\therefore \frac{\text{usage revenue}}{\text{capacity cost}} = \frac{1}{e} = \frac{2}{3} \qquad \frac{\text{subscription revenue}}{\text{capacity cost}} = \frac{1}{3}$$

- ie usage revenue covers marginal cost
  subscription revenue covers the rest

* obviously not practical to physically upgrade in such small steps

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	☹	☹	☹

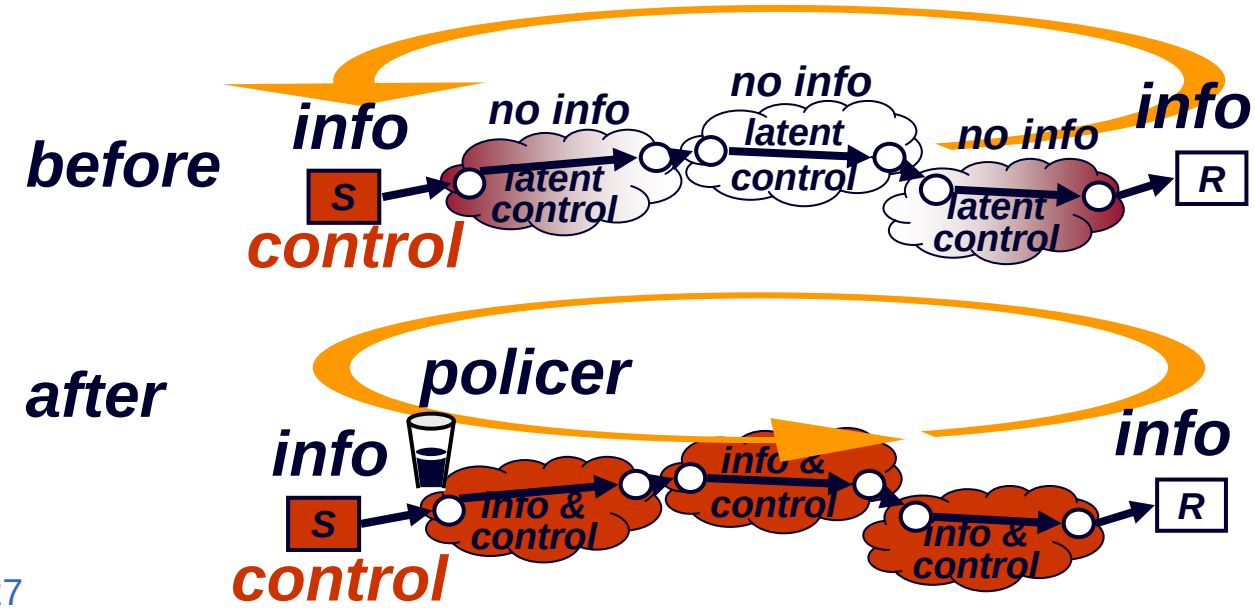
1. **loss:** used to signal congestion since the Internet's inception
   - computers detect congestion by detecting gaps in the sequence of packets
   - computers can hide these gaps from the network with encryption
2. **explicit congestion notification [ECN]:** standardised into TCP/IP in 2001
   - 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

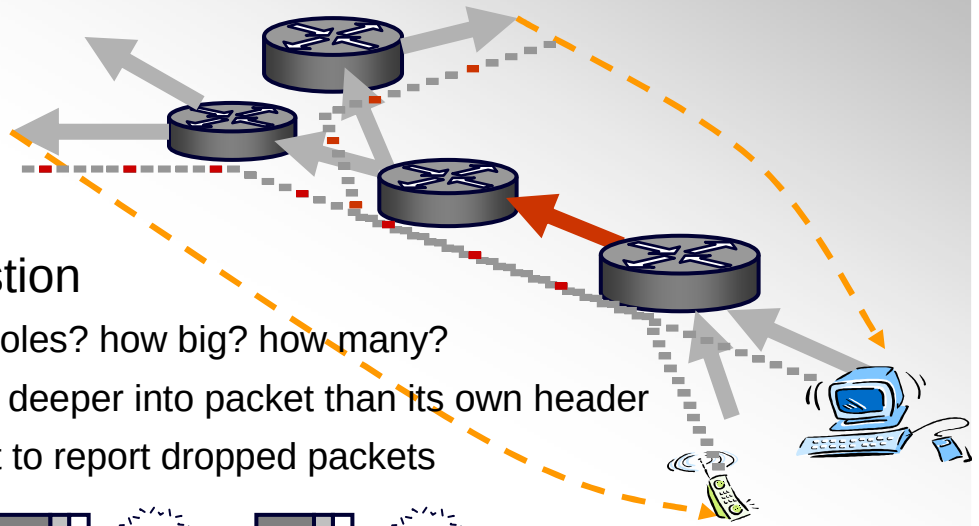
addition of re-feedback [re-ECN] – in brief

- *before:* congested nodes mark packets
  receiver feeds back marks to sender
- *after:* sender must pre-load expected congestion
  by re-inserting feedback
- if sender understates expected compared to actual congestion, network discards packets
- result: packets will carry prediction of downstream congestion
- policer can then limit congestion caused (or base penalties on it)



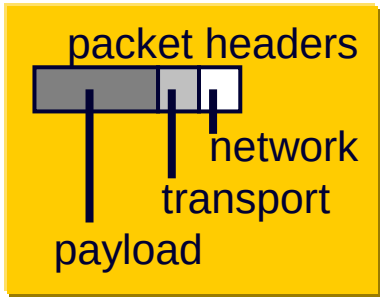
solution step #1: ECN

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

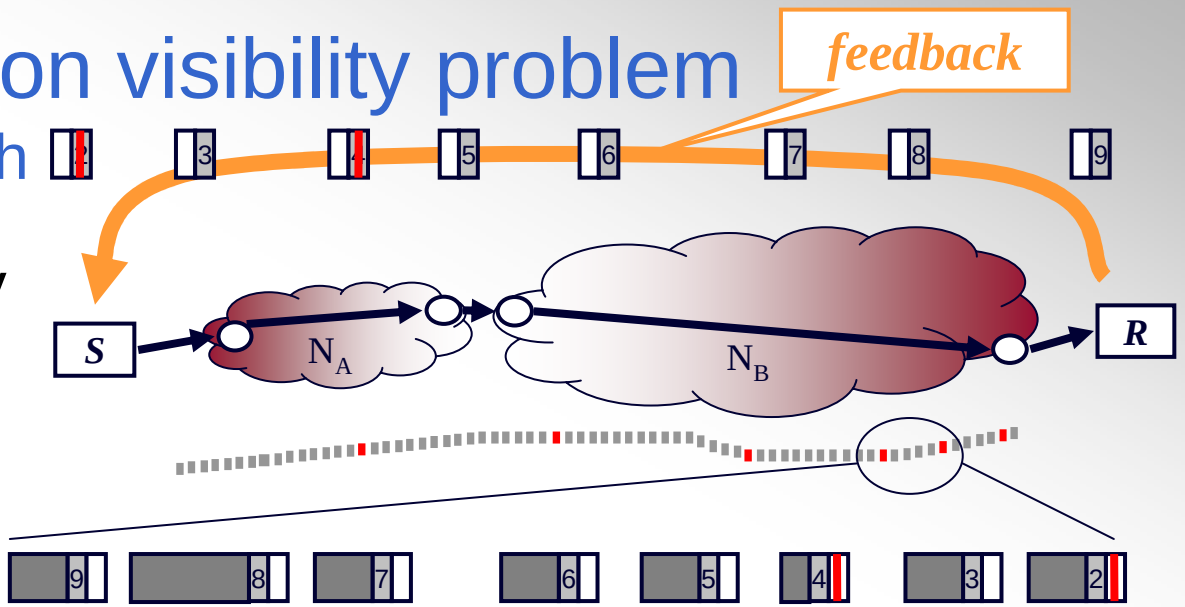


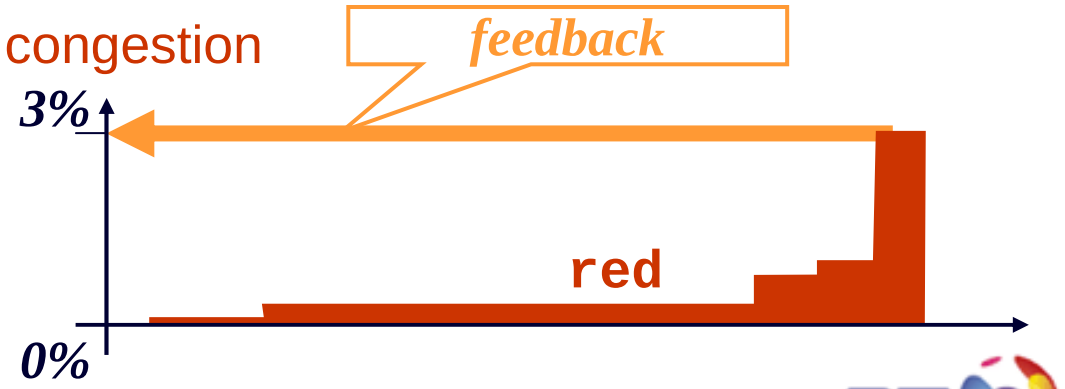
BT

new information visibility problem

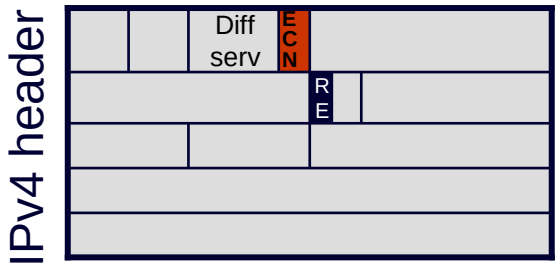
ECN is not enough

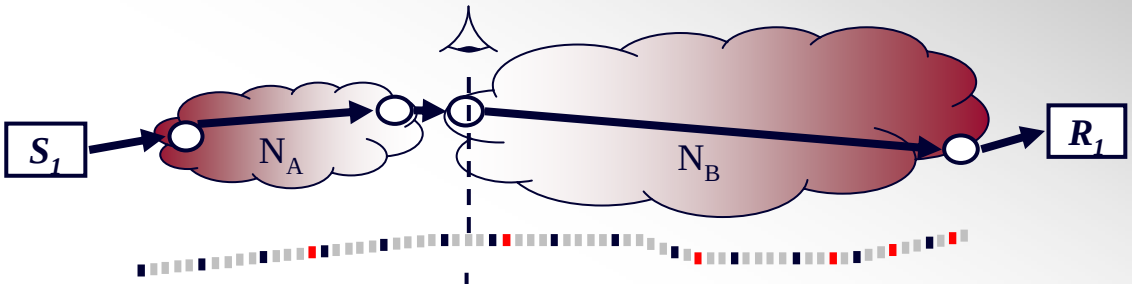
- path congestion only measurable at exit
- can't measure path congestion at entry
  - can't presume allowed deeper into feedback packets





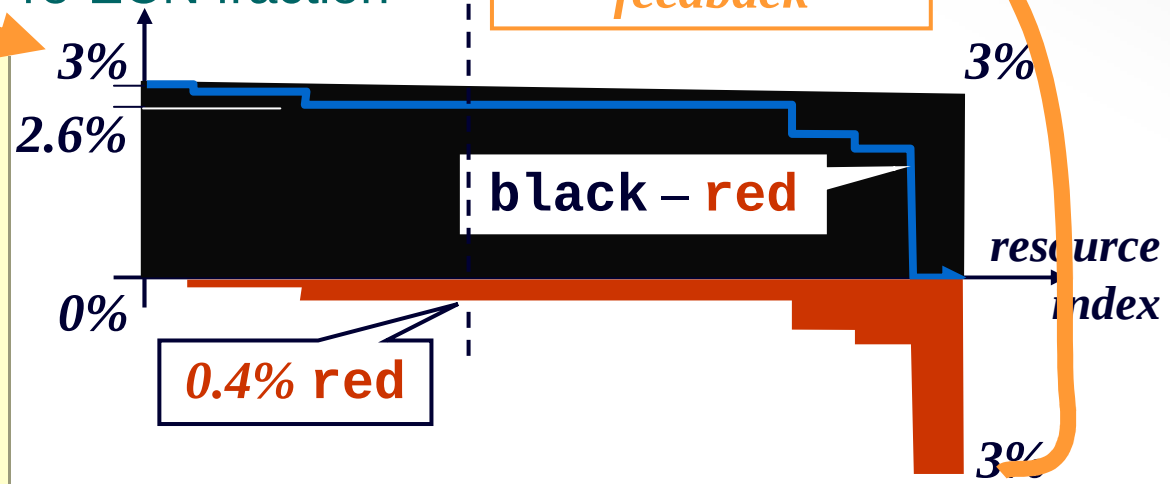
solution step #2: re-ECN
measurable downstream congestion





re-feedback

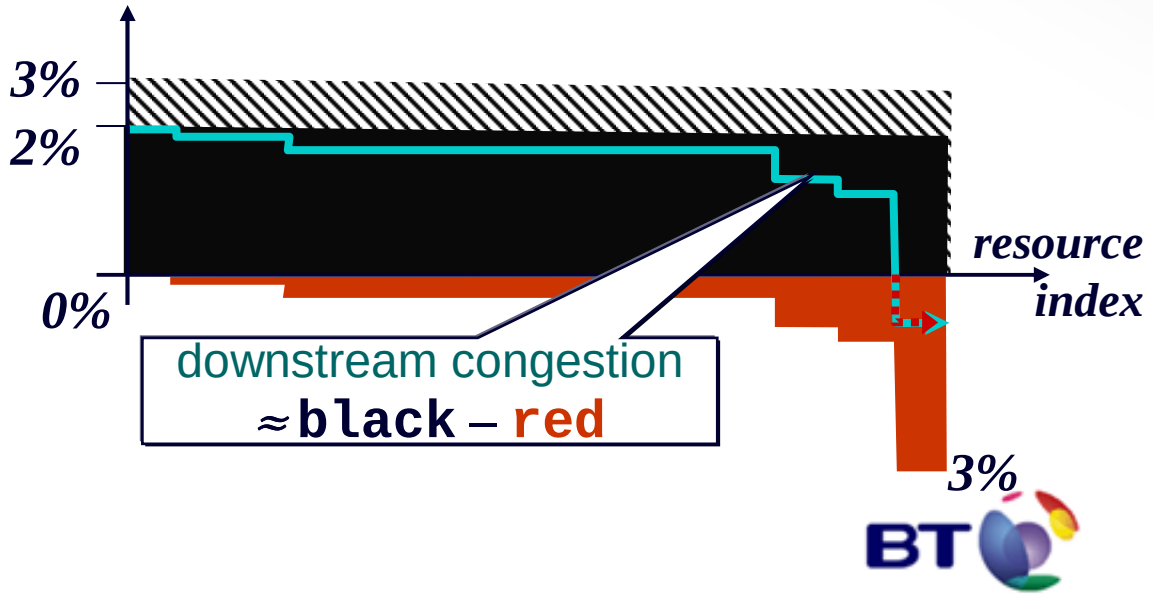
feedback



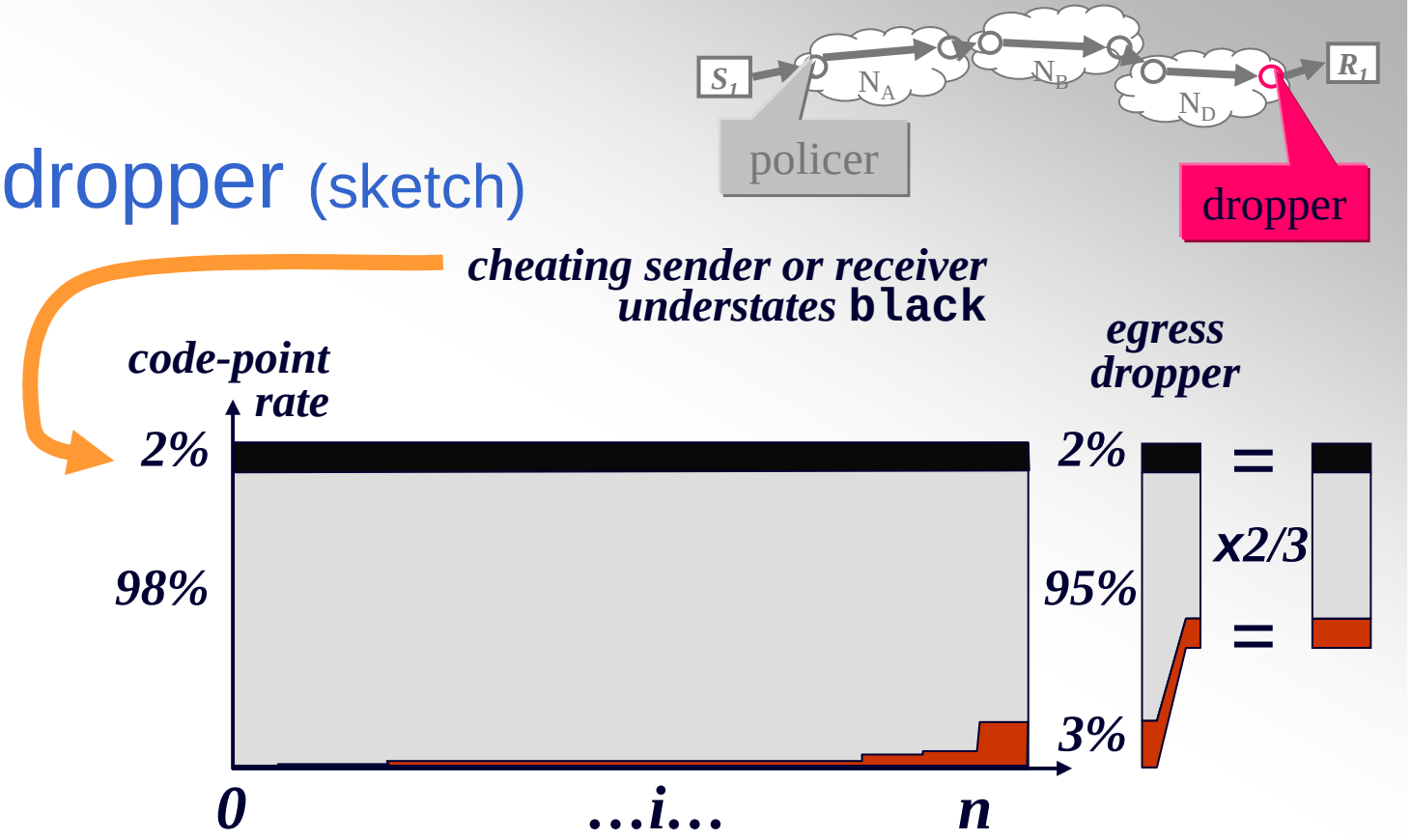
- sender re-inserts feedback by marking packets **black**
- at any point on path,diff betw fractions of **black** & **red** bytes is downstream congestion
- ECN routers unchanged
- **black** marking e2e but visible at net layer for accountability

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



egress dropper (sketch)



- drop enough traffic to make fraction of **red** = **black**
- goodput best if rcvr & sender honest about feedback & re-feedback

how to limit congestion with flat fee pricing

- only throttles traffic when contribution to congestion elsewhere exceeds allowance
- otherwise free to go at any bit-rate

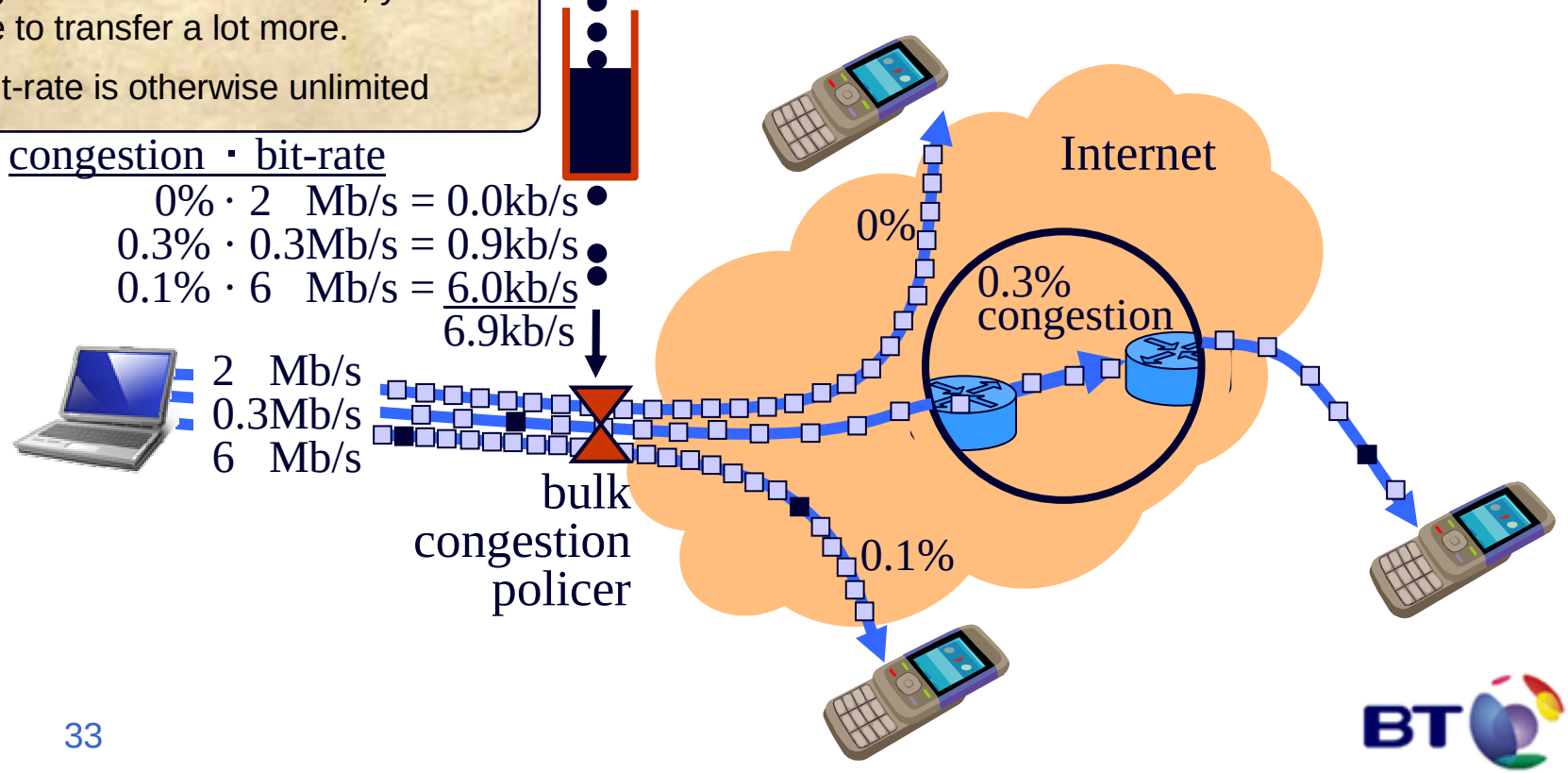
Acceptable Use Policy

Your 'congestion volume' allowance: 1GB/month (= 3kb/s continuous)
This only limits the traffic you can try to transfer above the maximum the Internet can take when it is congested.

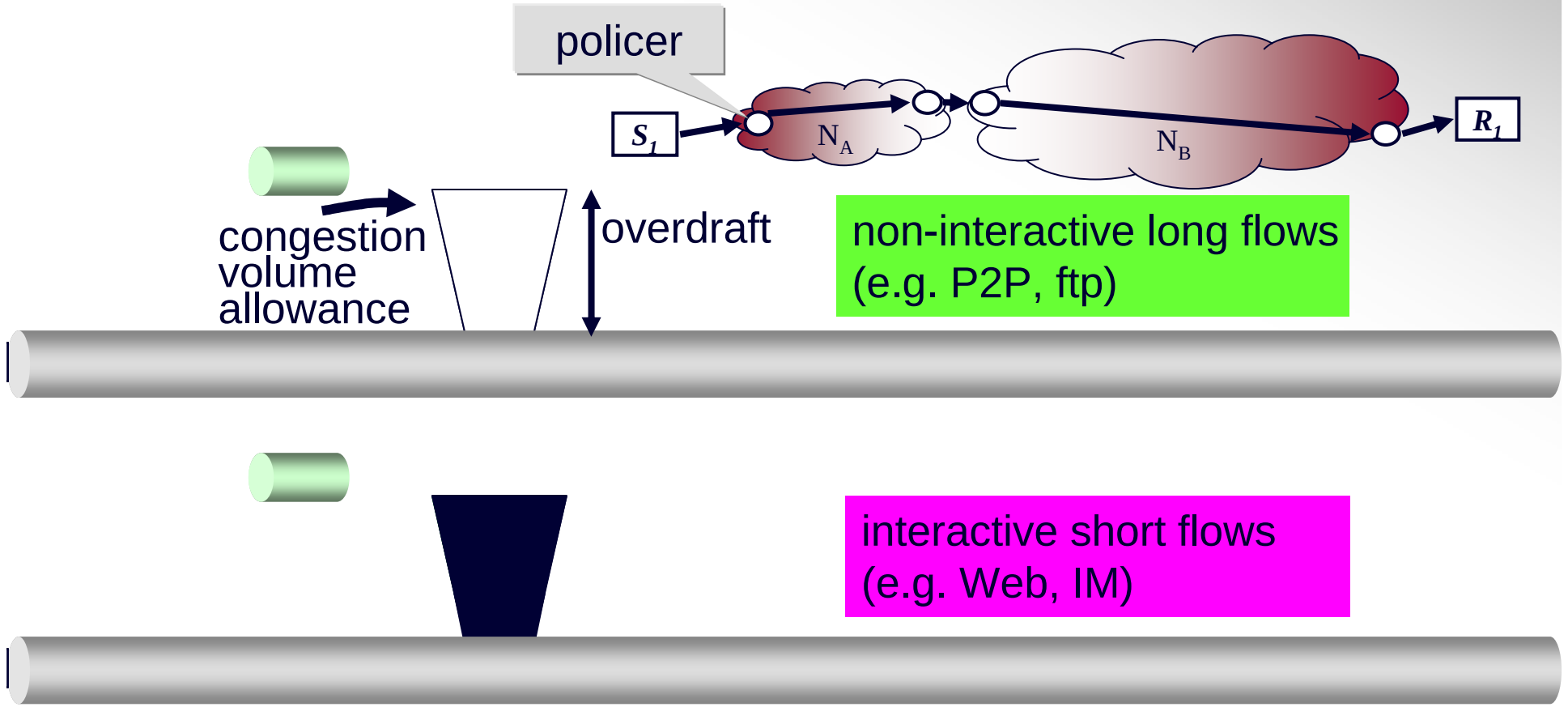
Under typical conditions this will allow you to transfer about **70GB per day**.

If you use software that seeks out uncongested times and routes, you will be able to transfer a lot more.

Your bit-rate is otherwise unlimited



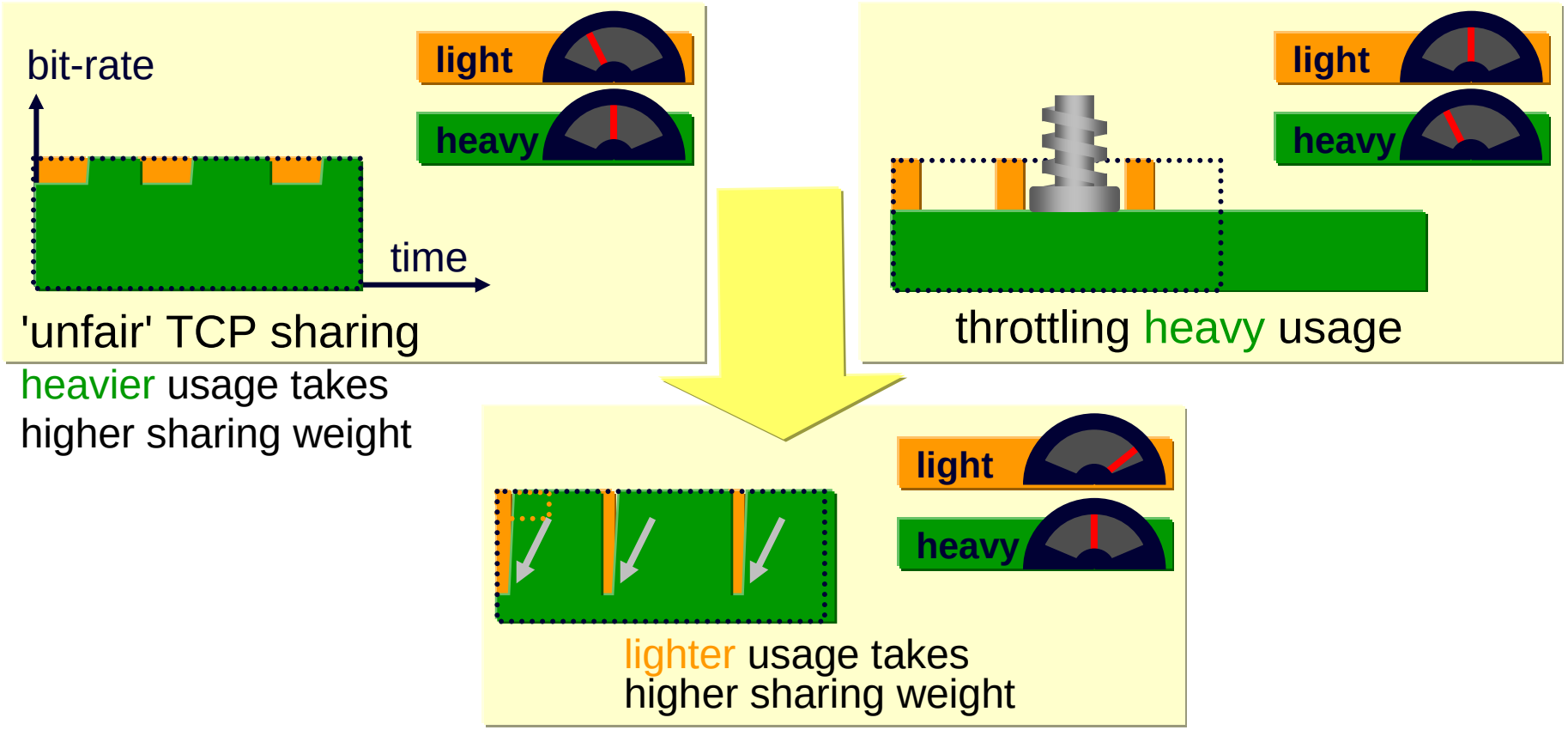
congestion policer – one example: per-user



two different customers, same deal

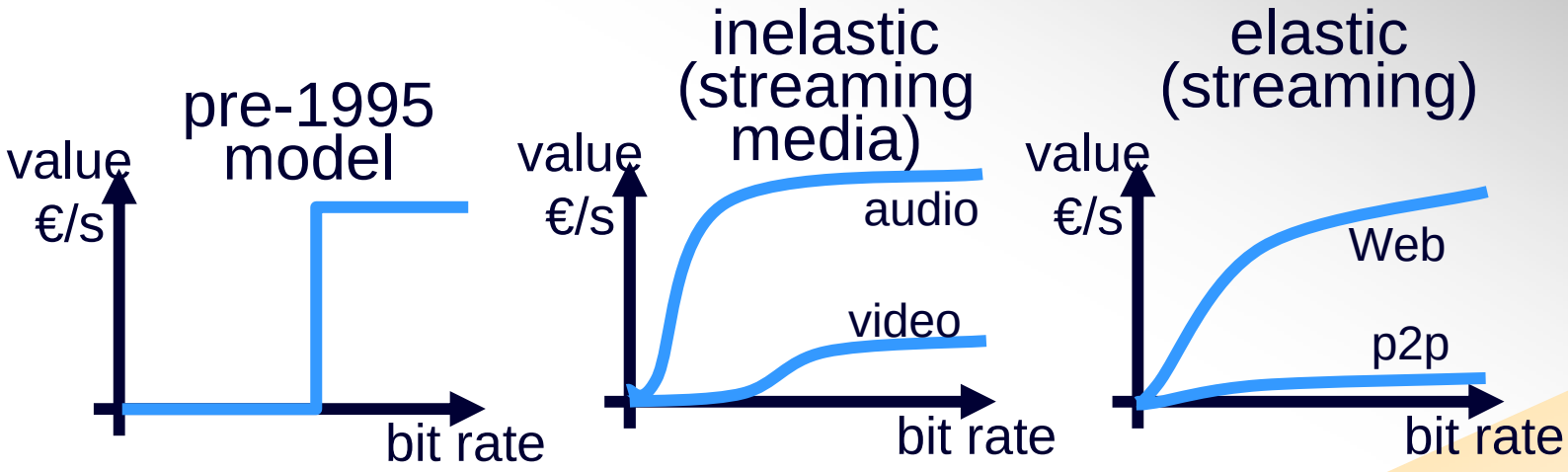
BT

fairer is faster – incentivise end host behaviour



- enabler: limit congestion, not volume
- then end system congestion control will quickly evolve (cf. BitTorrent DNA)
  - heavy usage will back away whenever light usage appears
  - so light usage can go much faster
  - hardly affecting completion times of heavy usage
- differentiated QoS *as if* in the network

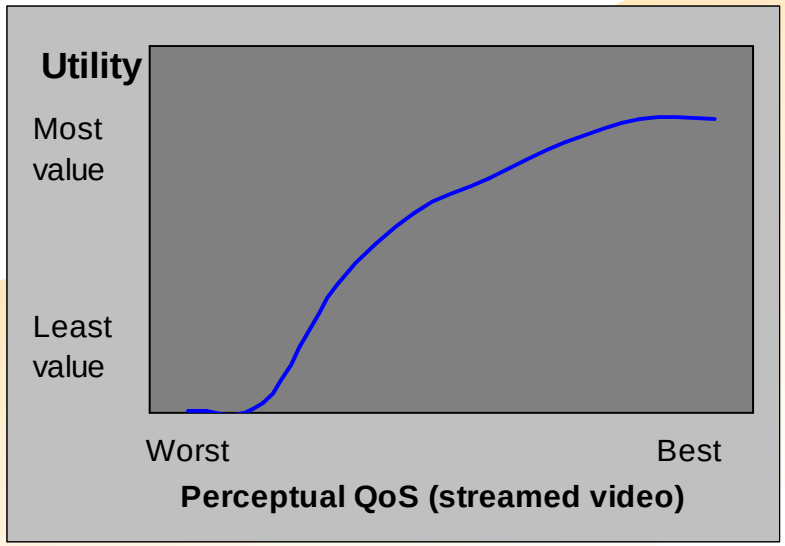
utility (value) wrt bit rate: curve families



theoretical
[Shenker95]

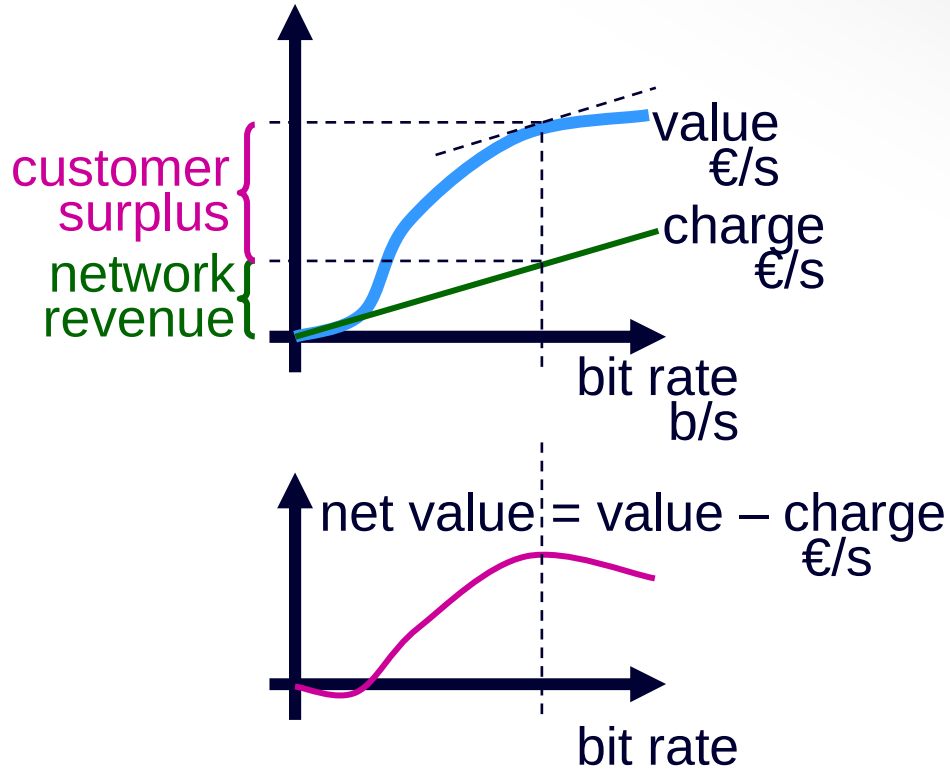
&

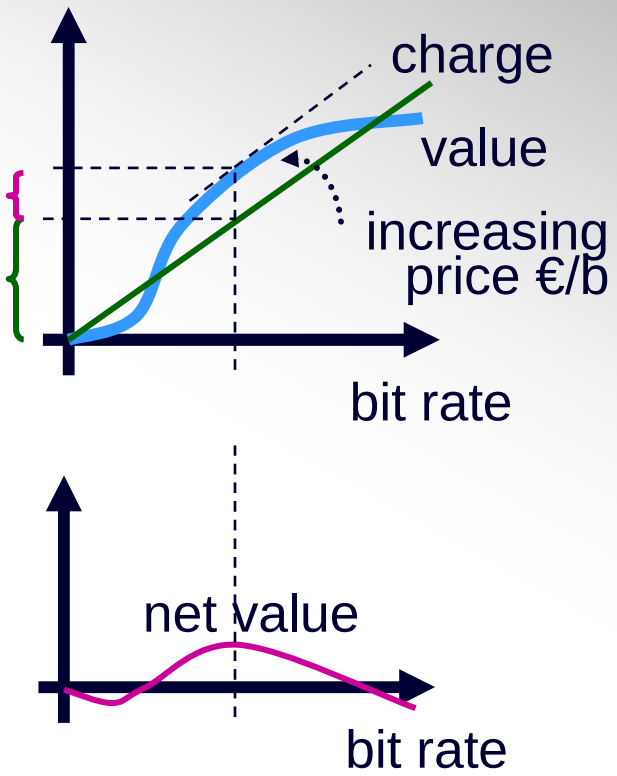
actual
value models



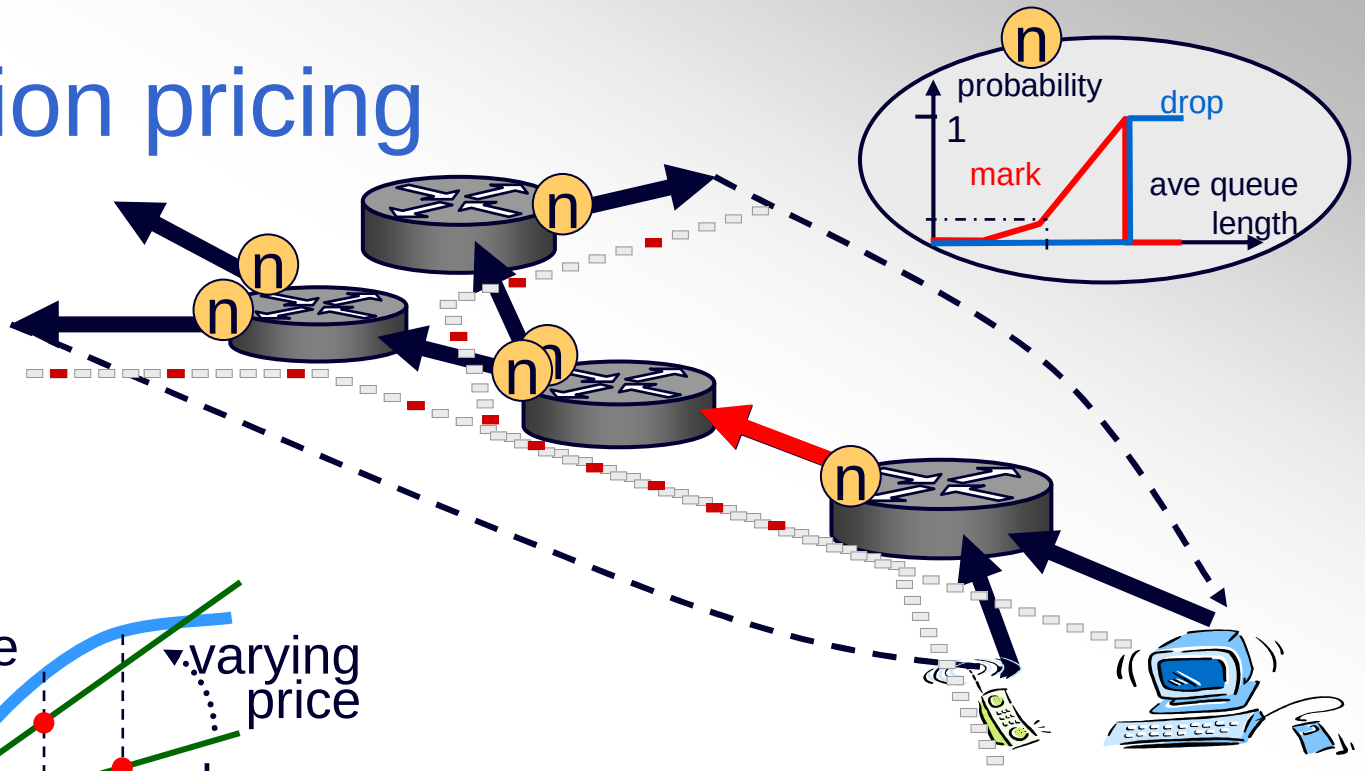
average of normalised curves from a set of experiments on paying customers [Hands02]

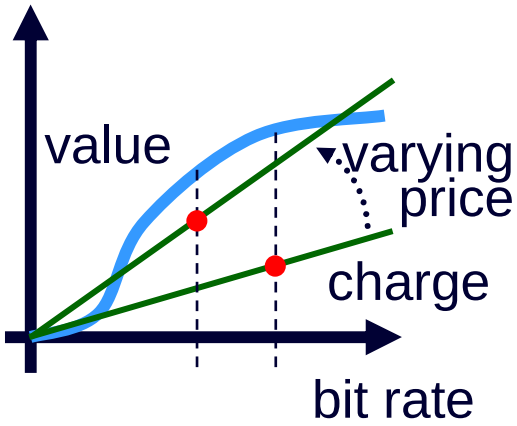
value – cost: customer's optimisation [Kelly98]

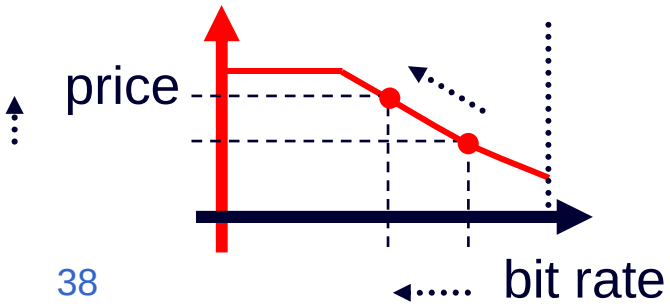




congestion pricing





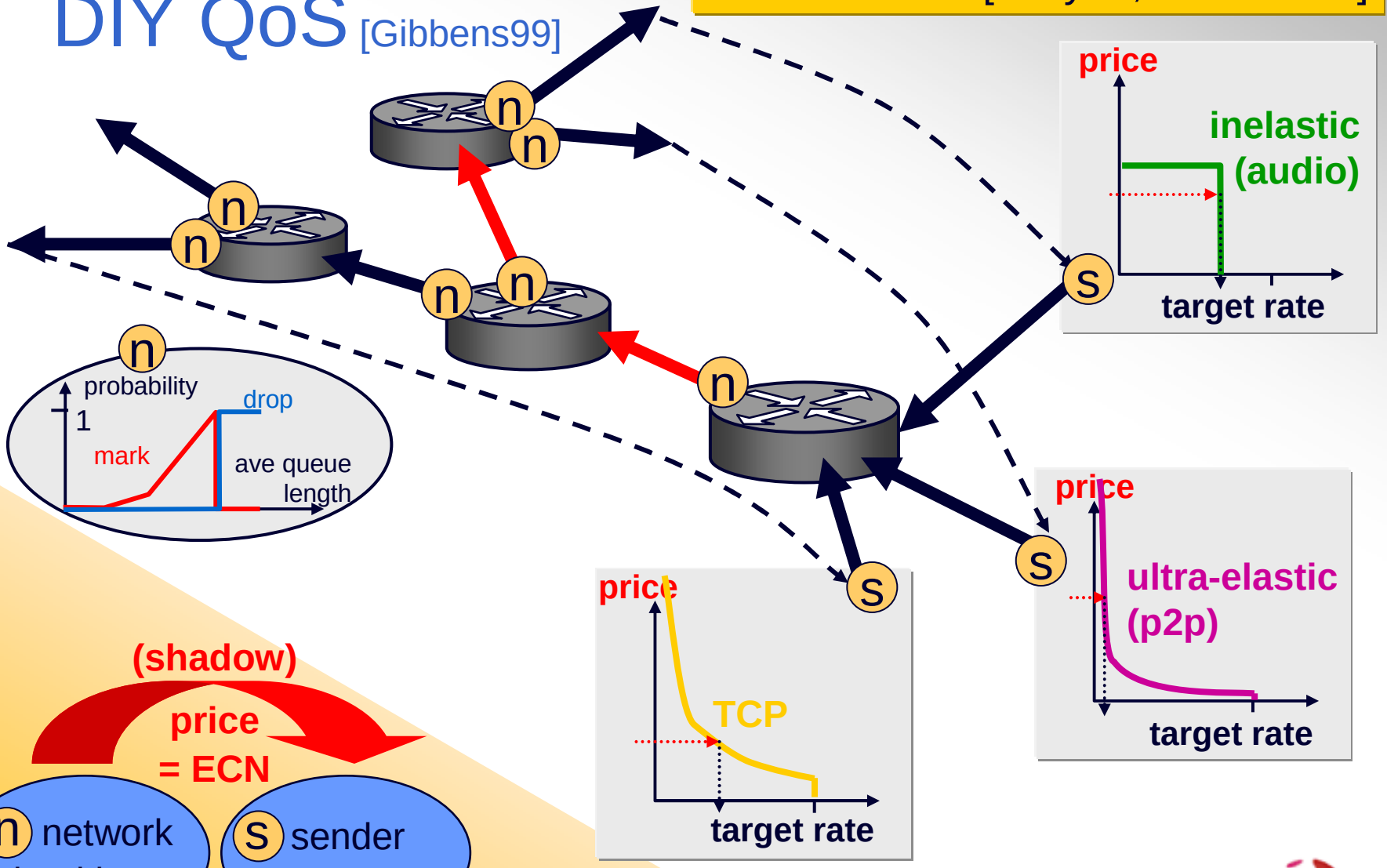


- volume charging
  - but only of marked packets

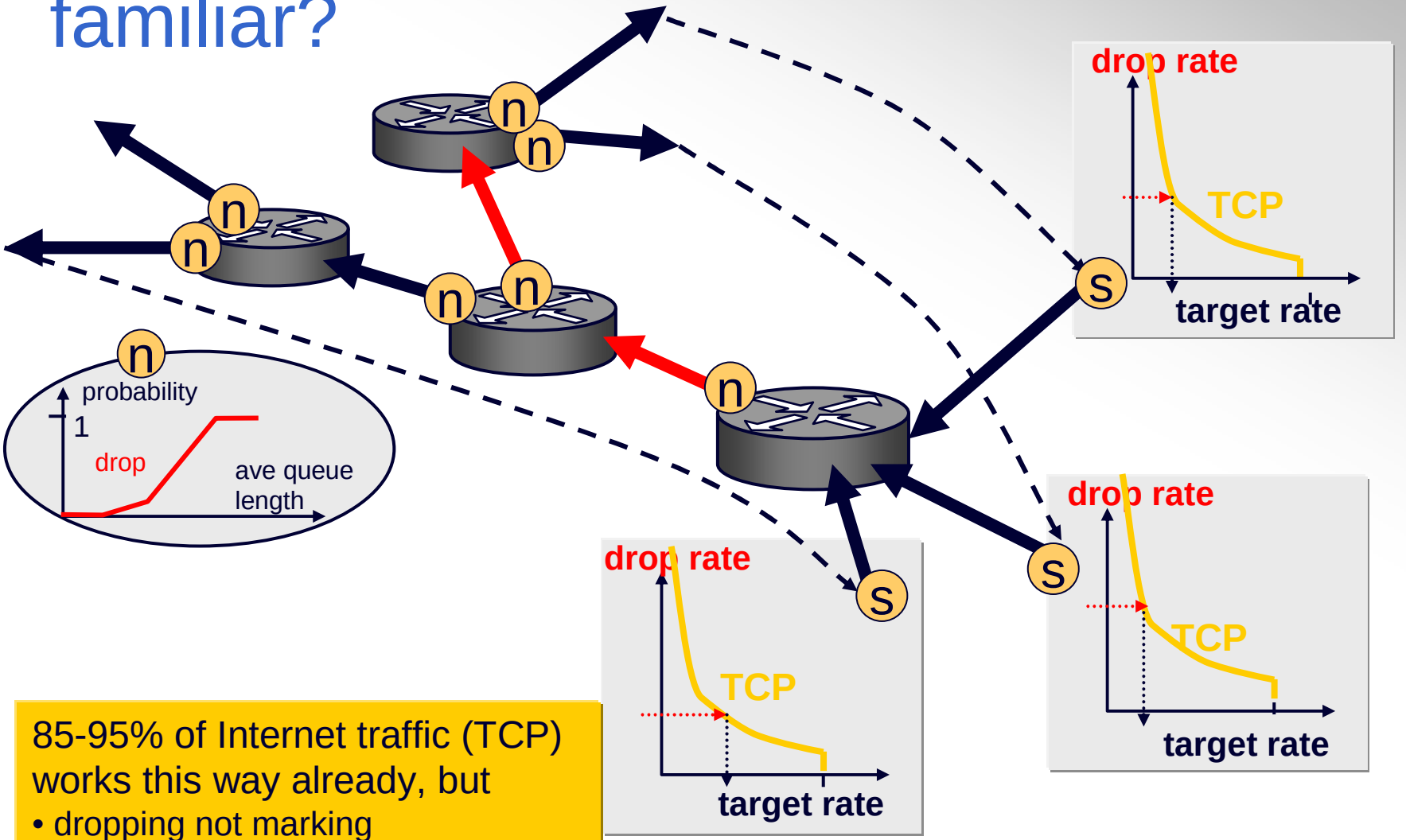
⇒ congestion charging

DIY QoS [Gibbens99]

maximises social welfare across whole Internet [Kelly98, Gibbens99]



familiar?

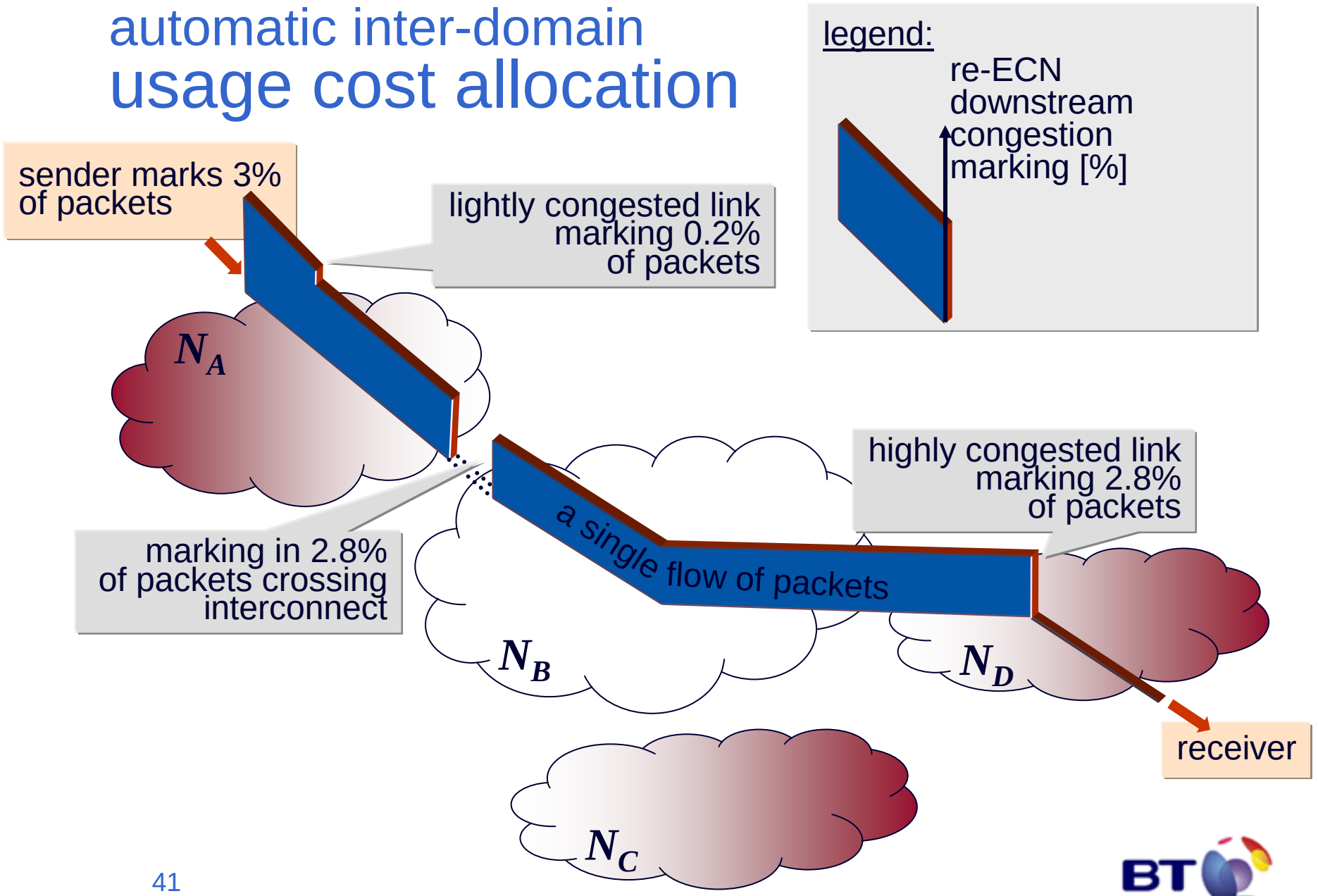


85-95% of Internet traffic (TCP) works this way already, but

- dropping not marking
- senders respond voluntarily

  ***as if* congestion charged**
- every sender responds identically

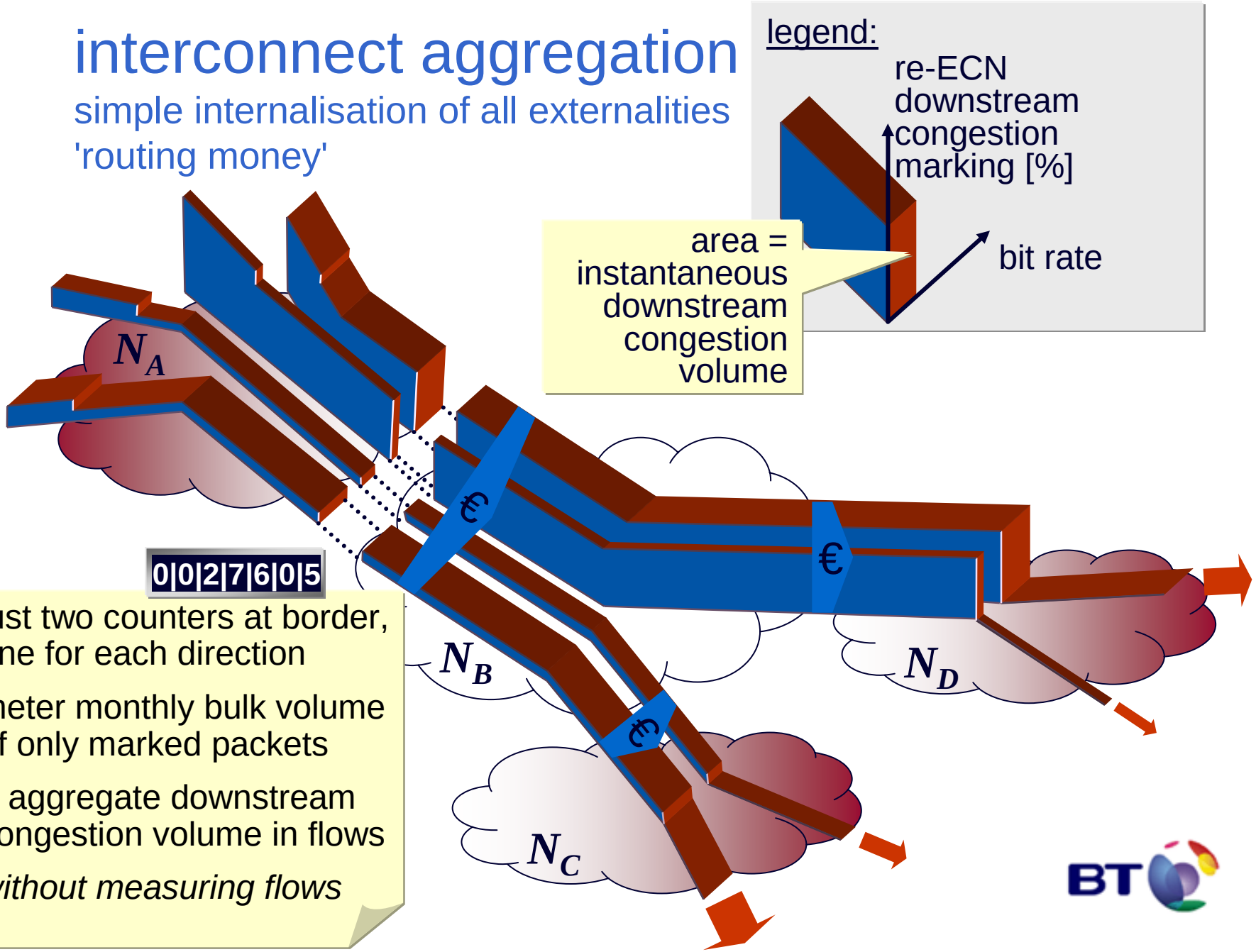
automatic inter-domain usage cost allocation



interconnect aggregation

simple internalisation of all externalities

'routing money'



just two counters at border, one for each direction

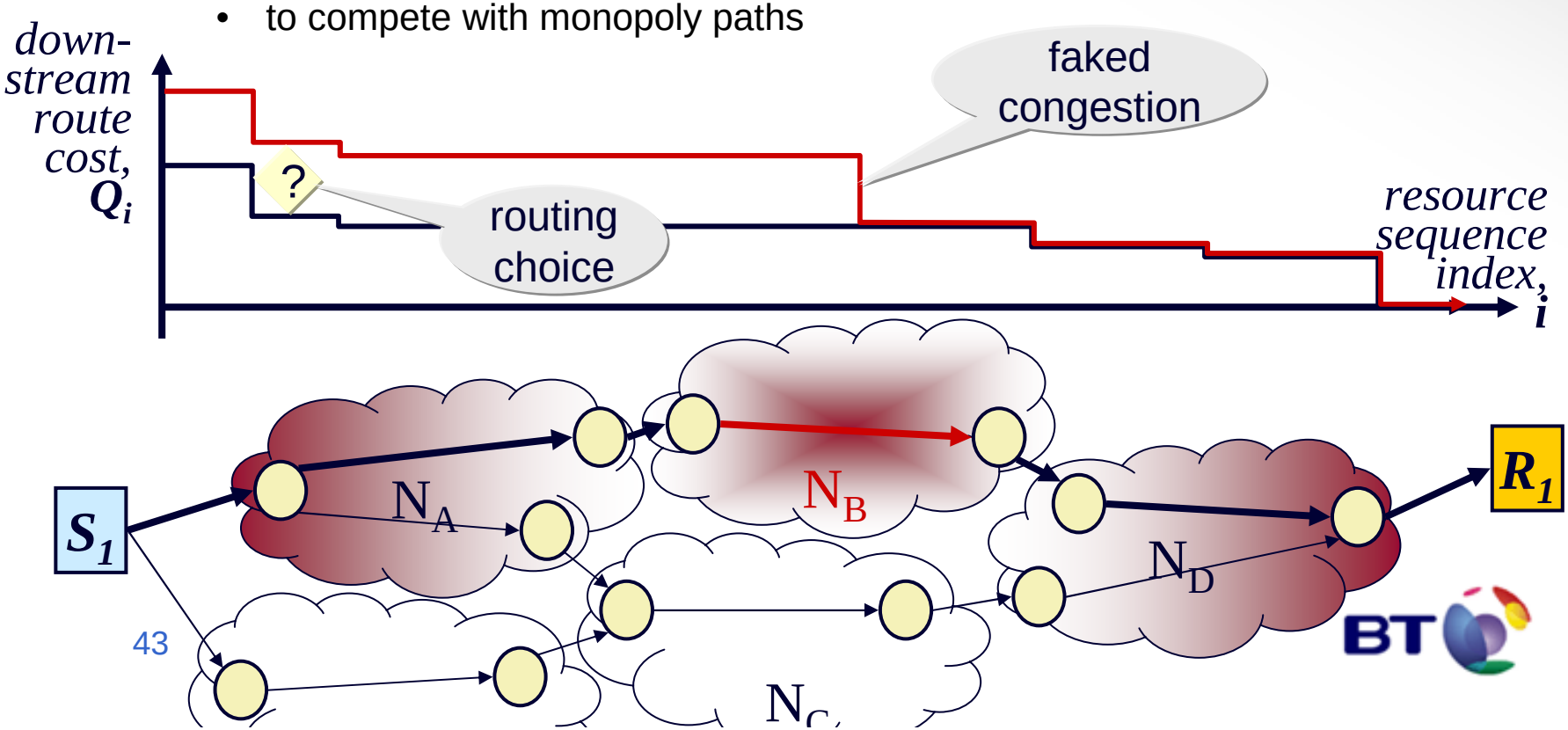
meter monthly bulk volume of only marked packets

= aggregate downstream congestion volume in flows

without measuring flows

congestion competition – inter-domain routing

- why won't a network overstate congestion?
  - 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
- also incentivises new provision
  - to compete with monopoly paths



minimal operational support system impact

- single bulk contractual mechanism
  - for end-customers and inter-network
  - also designed to simplify layered wholesale/retail market
- automated provisioning
  - driven by per-interface ECN stats – demand-driven supply
- automated inter-network monitoring & accounting
- QoS an attribute of customer contract not network
  - automatically adjusts to attachment point during mobility

BT

summary so far

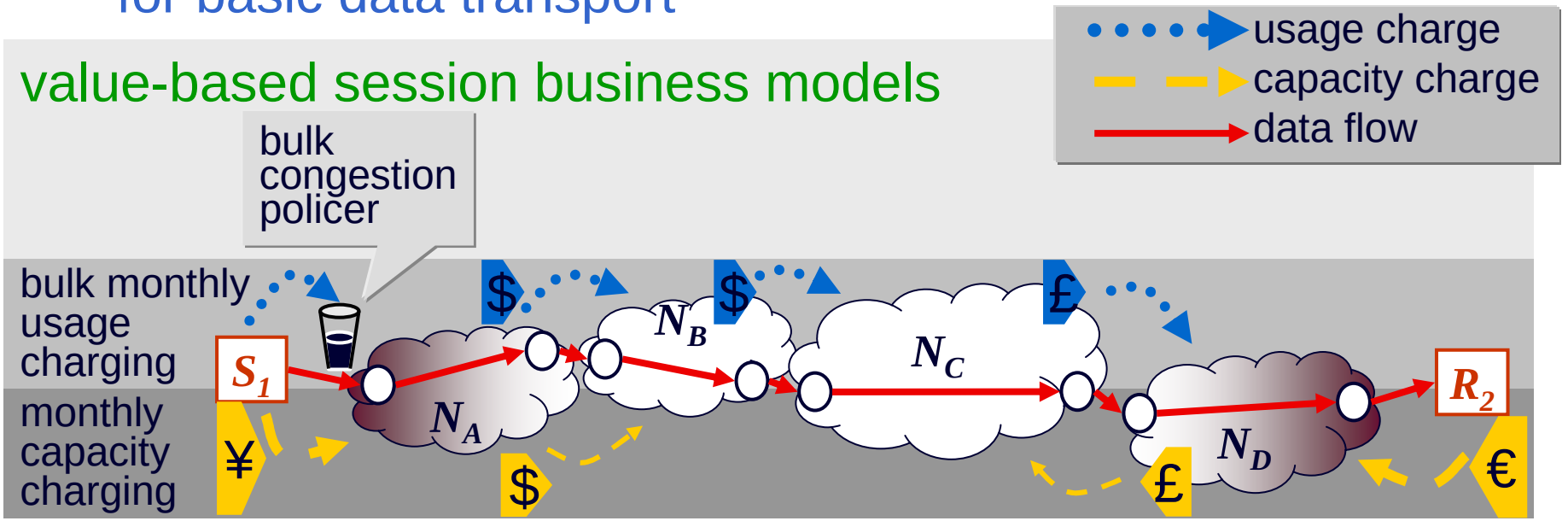
congestion accountability – the missing link

- unwise NGN obsession with per-session QoS guarantees
- scant attention to competition from 'cloud QoS'
  - rising general QoS expectation from the public Internet
  - cost-shifting between end-customers (including service providers)
  - questionable economic sustainability
- 'cloud' resource accountability is possible
  - principled way to heal the above ills
  - requires shift in economic thinking – from volume to congestion volume
- provides differentiated cloud QoS without further mechanism
- also the basis for a far simpler per-session QoS mechanism
  - having fixed the competitive environment to make per-session QoS viable

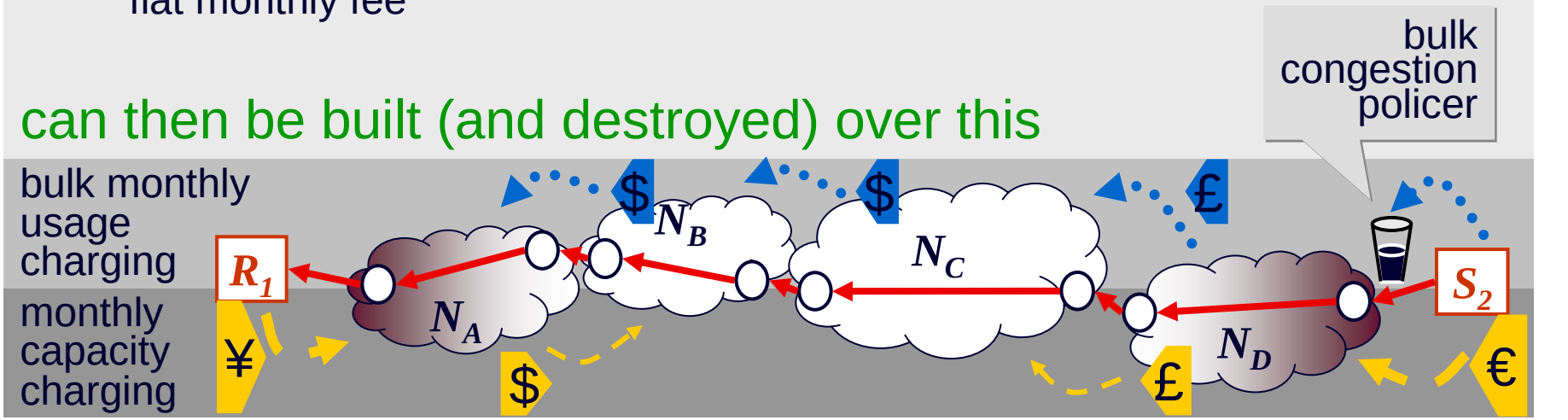
BT

sustainable business model

for basic data transport



usage flat fee
+ capacity flat fee
flat monthly fee

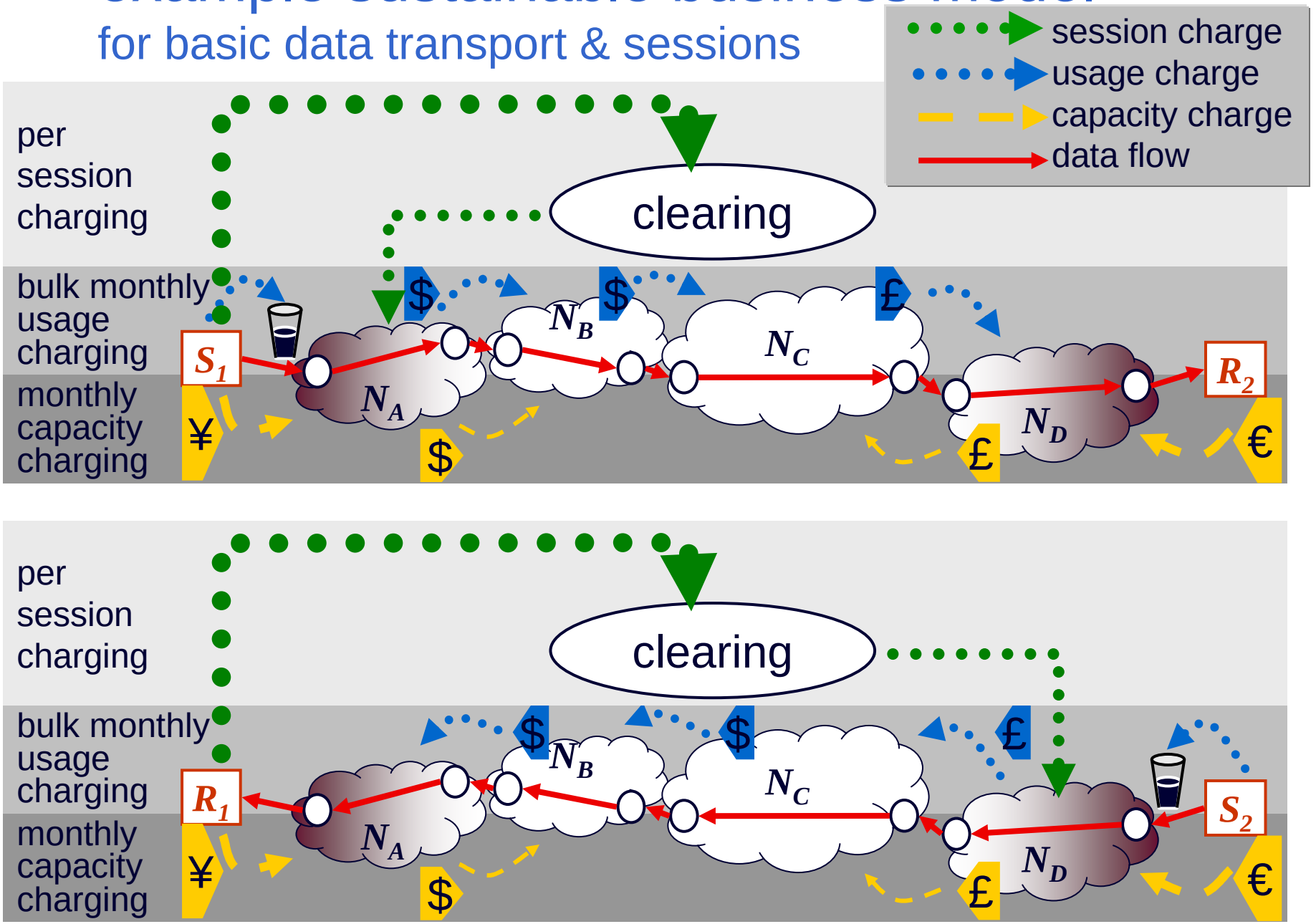


Internet QoS
value-based per-session charging

Bob Briscoe

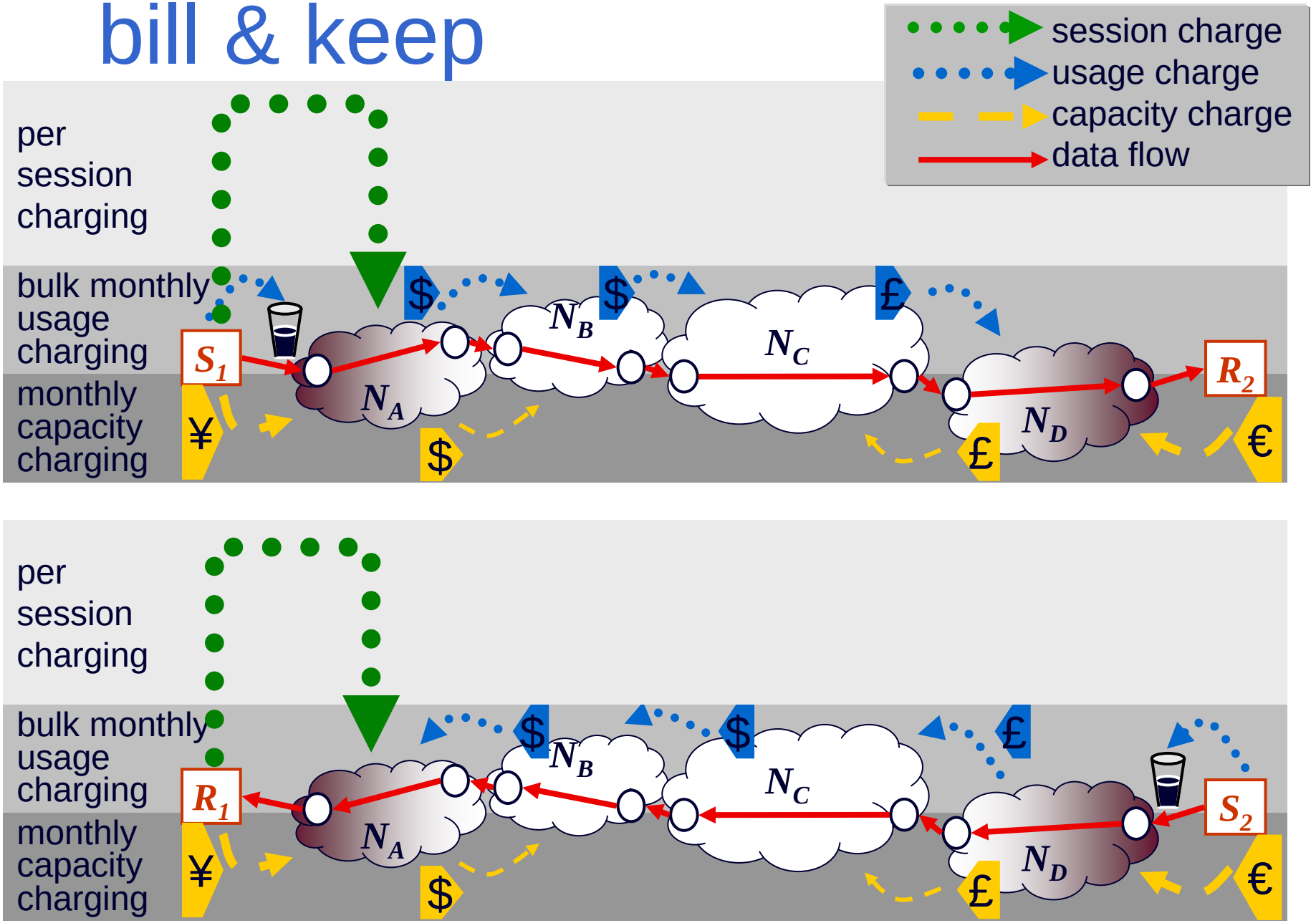
example sustainable business model

for basic data transport & sessions



more simply

bill & keep



what's the added value to sessions?

- insurance – risk brokerage
  - once admitted, a session will complete
  - at a fixed per session price (per service, per time, etc)
- low loss, low jitter
  - even for high & variable bandwidth
  - video, audio

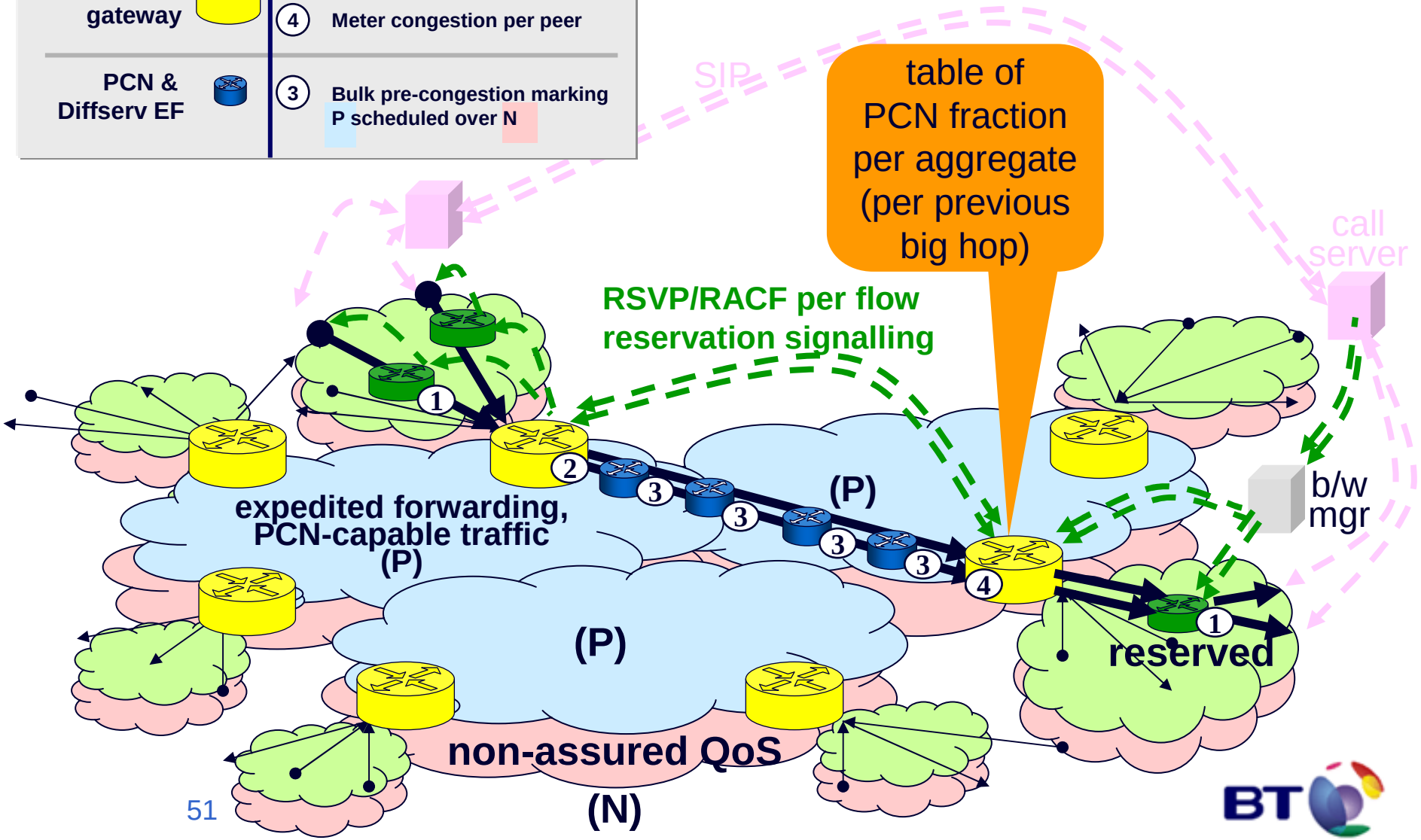
- re-ECN proposal is not 'carrier grade'
- but with two tweaks it is
  - pre-congestion notification [PCN]
  - admission control
- both are also built on similar simple economic principles...

BT

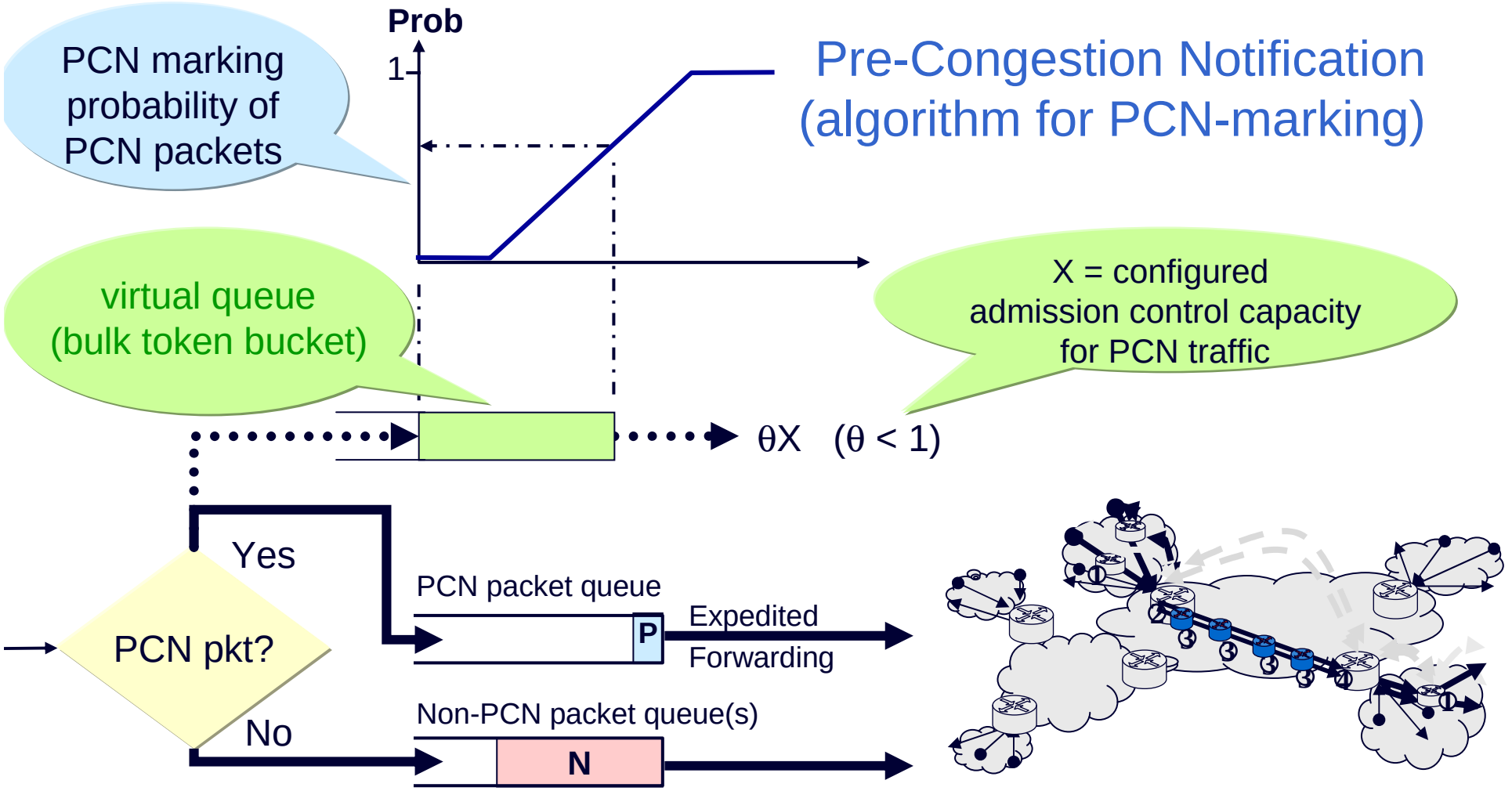
PCN system arrangement

highlighting 2 flows

IP routers	Data path processing
Reservation enabled	1 Reserved flow processing
RSVP/PCN gateway	2 Policing flow entry to P 4 Meter congestion per peer
PCN & Diffserv EF	3 Bulk pre-congestion marking P scheduled over N

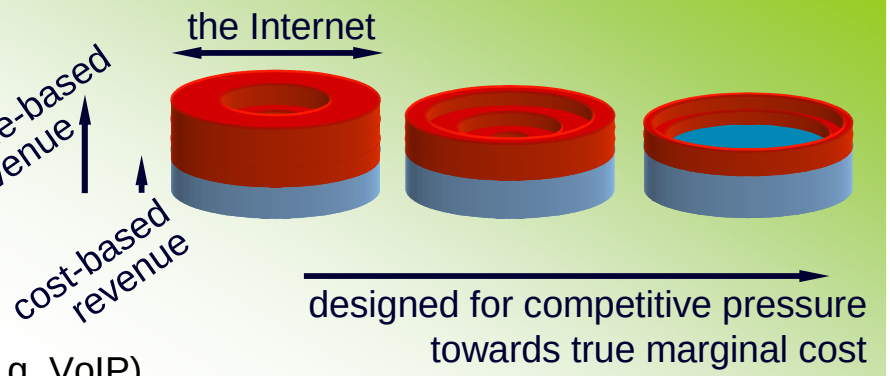


Pre-Congestion Notification (algorithm for PCN-marking)

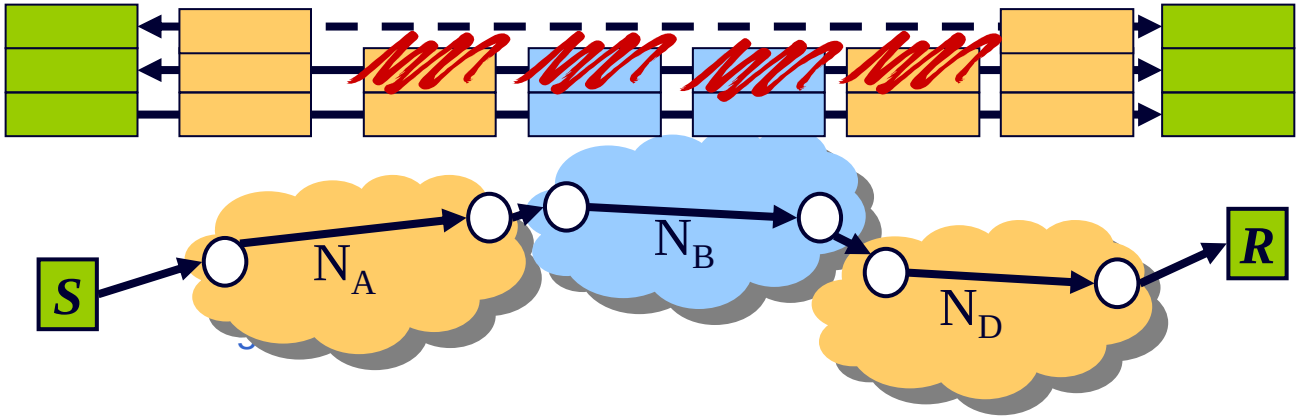


- virtual queue (a conceptual queue – actually a simple counter):
  - drained somewhat slower than the rate configured for adm ctrl of PCN traffic
  - therefore build up of virtual queue is ‘early warning’ that the amount of PCN traffic is getting close to the configured capacity
  - NB mean number of packets in real PCN queue is still very small

value-based charges over low cost floor



- over IP, currently choice between
  - A. “good enough” service with no QoS costs (e.g. VoIP)
    - but can brown-out during peak demand or anomalies
  - B. fairly costly QoS mechanisms – either admission control or generous sizing
- this talk: where the premium end of the market (B) is headed
  - a new IETF technology: pre-congestion notification (PCN)
  - service of ‘B’ but mechanism cost competes with ‘A’
    - assured bandwidth & latency + PSTN-equivalent call admission probability
    - fail-safe fast recovery from even multiple disasters
- core networks could soon fully guarantee sessions without touching sessions
  - some may forego falling session-value margins to compete on cost



app signal (SIP) QoS admission	**per session**
priority forwarding **& PCN**	**bulk data**

PCN

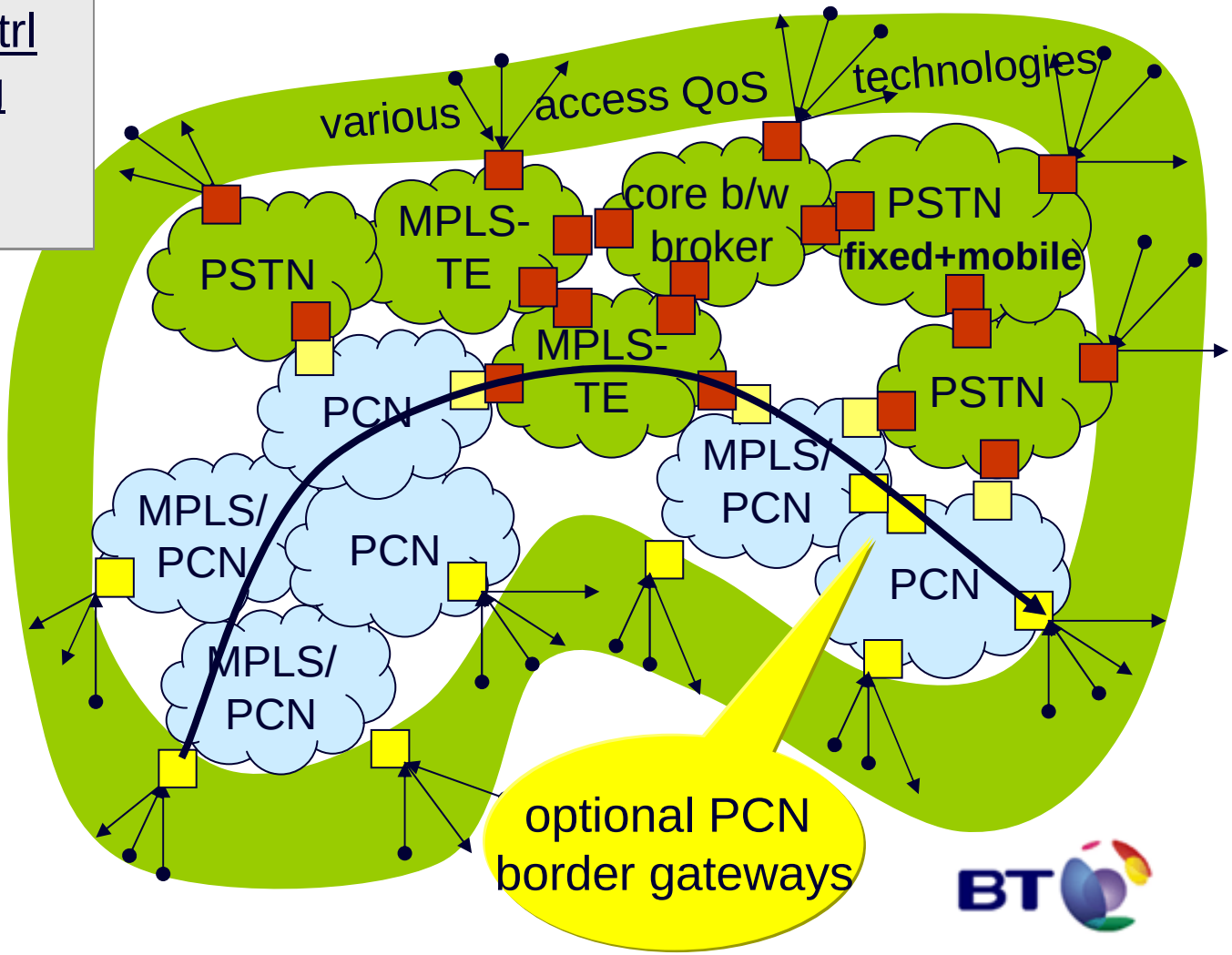
the wider it is deployed
the more cost it saves

legend

- connection-oriented (CO) QoS
- PCN QoS

flow admission ctrl & border policing

- PCN / CO
- CO / CO



Still initiated by end to end app layer signalling (SIP)

Figure focuses on layers below

PCN status

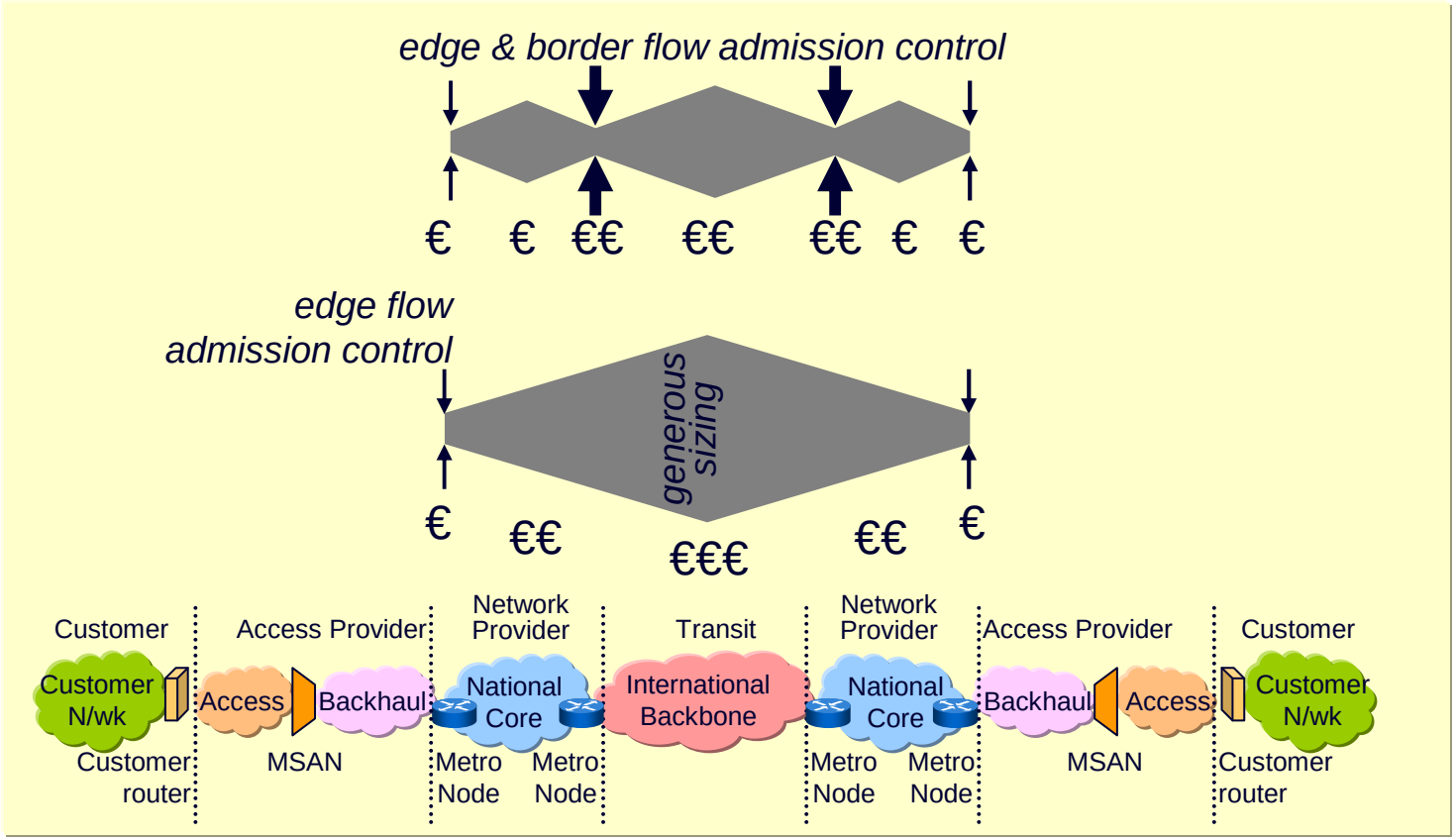
- main IETF PCN standards scheduled for Mar'09
  - main author team from companies on right (+Universities)
  - wide & active industry encouragement (no detractors)
- IETF initially focusing on *intra*-domain
  - but chartered to "keep *inter*-domain strongly in mind"
  - re-charter likely to shift focus to interconnect around Mar'09
- detailed extension for interconnect already tabled (BT)
  - holy grail of last 14yrs of IP QoS effort
  - fully guaranteed global internetwork QoS with economy of scale
- ITU integrating new IETF PCN standards
  - into NGN resource admission control framework (RACF)

- BT's leading role: extreme persistence
  - 1999: identified value of original idea (from Cambridge Uni)
  - 2000-02: BT-led EU project: extensive economic analysis & engineering
  - 2003-06: extensive further simulations, prototyping, analysis
  - 2004: invented globally scalable interconnect solution
  - 2004: convened vendor design team (2 bringing similar ideas)
  - 2005-: introduced to IETF & continually pushing standards onward
  - 2006-08: extended to MPLS (& Ethernet next) with vendors

classic trade-off with diseconomy of scale either way

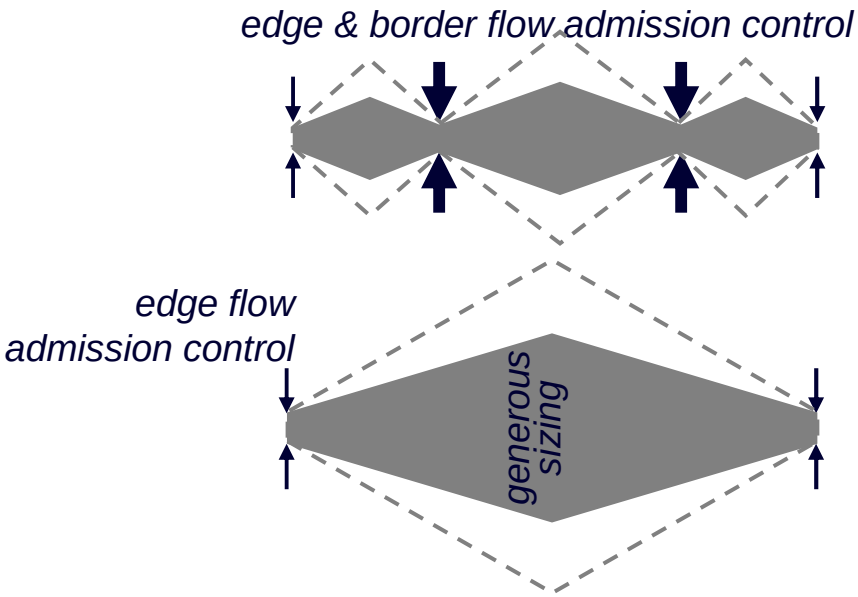
seen in all QoS schemes before PCN

- flow admission ctrl (smarts) vs. generous sizing (capacity)
  - the more hops away from admission control smarts
  - the more generous sizing is needed for the voice/video class



current Diffserv interior link provisioning
for voice/video expedited forwarding (EF) class

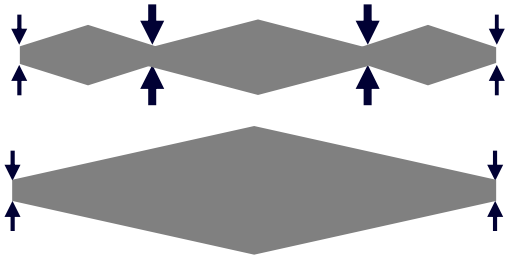
- admission control at network edge but not in interior
  - use typical calling patterns for base size of interior links, then...
  - add normal, PSTN-like over-provisioning to keep call blocking probability low
  - add extra Diffserv generous provisioning in case admitted calls are unusually focused



- residual risk of overload
  - reduces as oversizing increases
- stakes
  - brown out of *all* calls in progress

new IETF simplification

pre-congestion notification [PCN]



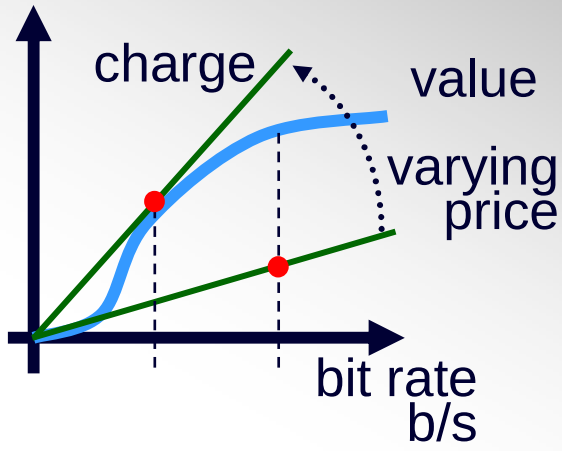
- PCN: radical cost reduction
  - compared here against simplest alternative – against 6 alternatives on spare slide
  - no need for any Diffserv generous provisioning between admission control points
    - 81% less b/w for BT's UK PSTN-replacement
    - ~89% less b/w for BT Global's premium IP QoS
    - still provisioned for low (PSTN-equivalent) call blocking ratios as well as carrying re-routed traffic after any dual failure
  - no need for interior flow admission control smarts, just one big hop between edges
- PCN involves a simple change to Diffserv

  - interior nodes randomly *mark* packets as the class nears its provisioned rate
  - pairs of edge nodes use level of marking between them to control flow admissions
  - much cheaper and more certain way to handle very unlikely possibilities
- interior nodes can be IP, MPLS or Ethernet
  - can use existing hardware, tho not all is ideal

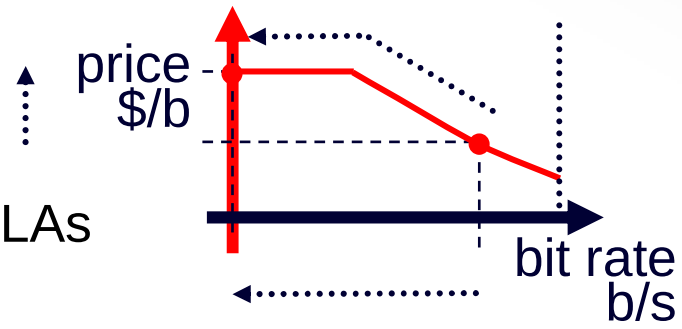
BT

congestion notification also underlies…

- scalable flow admission control
  - for S-shaped value curves (inelastic streaming media)
  - See [PCN]



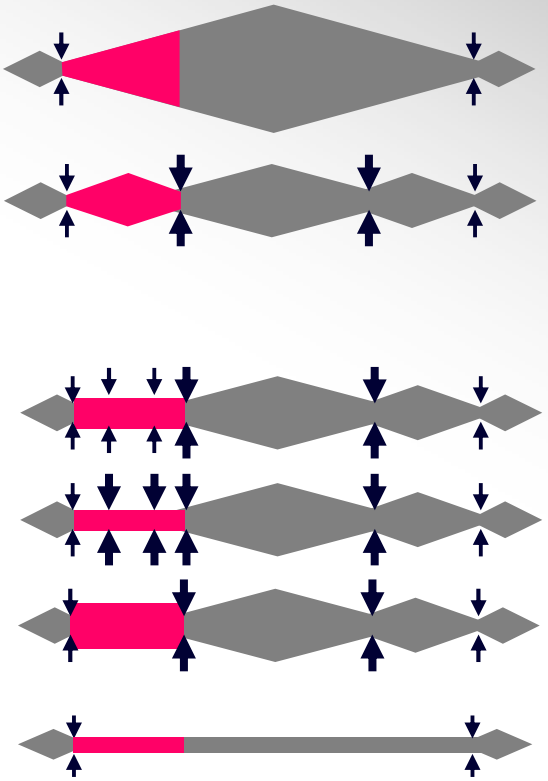
- class of service pricing
- verifying impairment budgets in SLAs
- resource allocation for VPNs
- …



BT

core & interconnect QoS
comparative evaluation

	inter-connect	brown-out risk	opex	capex	
				capacity	flow smarts
Diffserv with edge AC but no border AC	bulk rate	finite	££	£££	£
Diffserv with edge and border AC	flow AC	finite	££	££	££
core bandwidth broker	vapour-ware?	finite?	££	£	£££
MPLS-TE hard LSPs and border AC	flow AC	~0	£	££	££
MPLS-TE soft LSPs and border AC	flow AC	~0	£	£	£££
non-blocking core and border AC	flow AC	~0	£	££	££
PCN	bulk congestion	~0	£	£	£

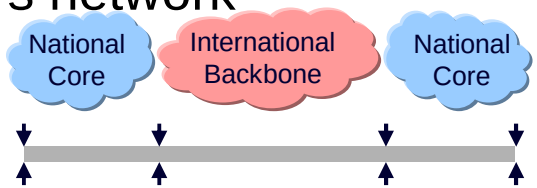


downside to PCN: not available quite yet!

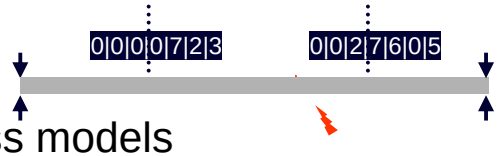
BT

PCN best with new interconnect business model
bulk border QoS

- can deploy independently within each operator's network
  - with session border controllers & flow rate policing
  - preserves traditional interconnect business model



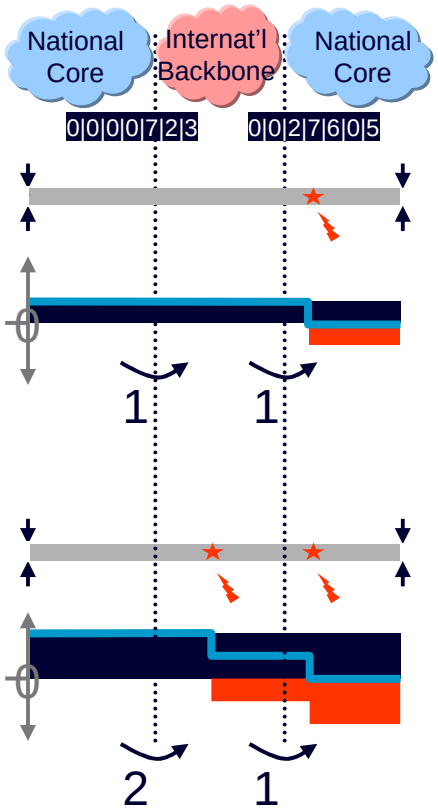
- but most benefit from removing all per-flow border controls
  - instead, simple bulk count of bytes in PCN marked packets crossing border
    - out of band (also helps future move to all-optical borders)
  - each flow needs just one per-flow admission control hop edge to edge
- new business model only at interconnect
  - no change needed to edge / customer-facing business models
  - not selling same things across interconnects as is sold to end-customer
  - but bulk interconnect SLAs with penalties for causing pre-congestion can create the same guaranteed retail service



BT

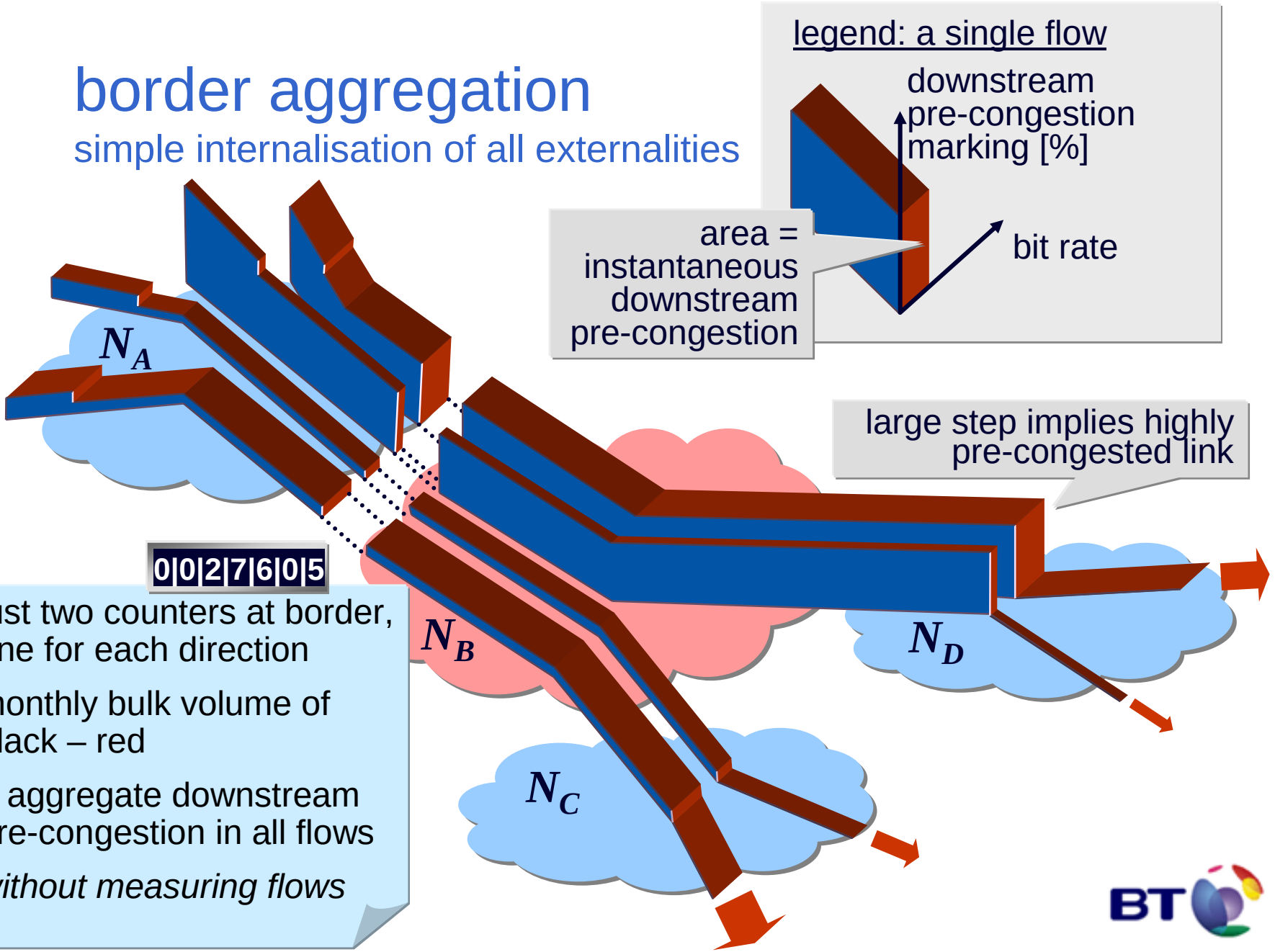
accountability of sending networks

- in connectionless layers (IP, MPLS, Ethernet)
  - marks only meterable downstream of network being congested
  - but sending network directly controls traffic
- trick: introduce another colour marking (black) [re-PCN]
  - contractual obligation for flows to carry as much black as red
    - sending net must insert enough black
  - black minus red = pre-congestion being caused downstream
  - still measured at borders in bulk, not within flows
- apportionment of penalties
  - for most metrics, hard to work out how to apportion them
  - as local border measurements decrement along the path they naturally apportion any penalties



border aggregation

simple internalisation of all externalities



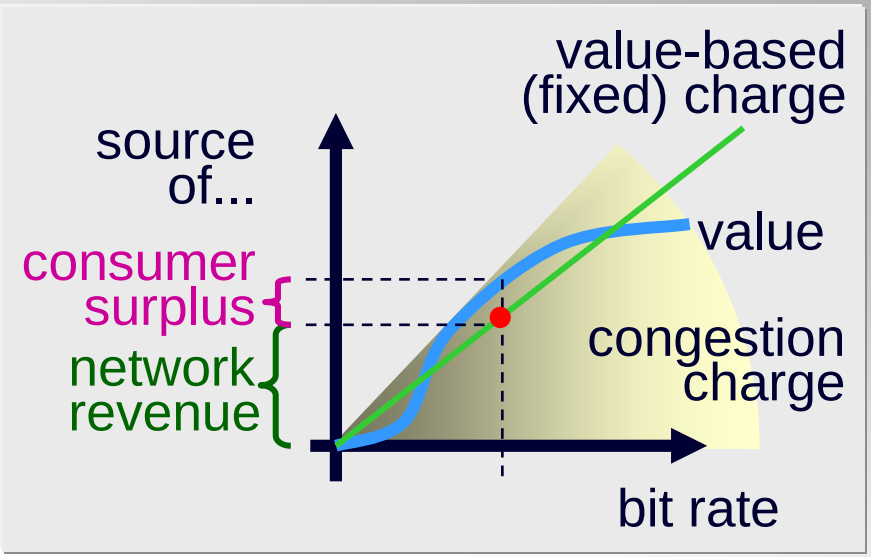
just two counters at border, one for each direction

monthly bulk volume of black – red

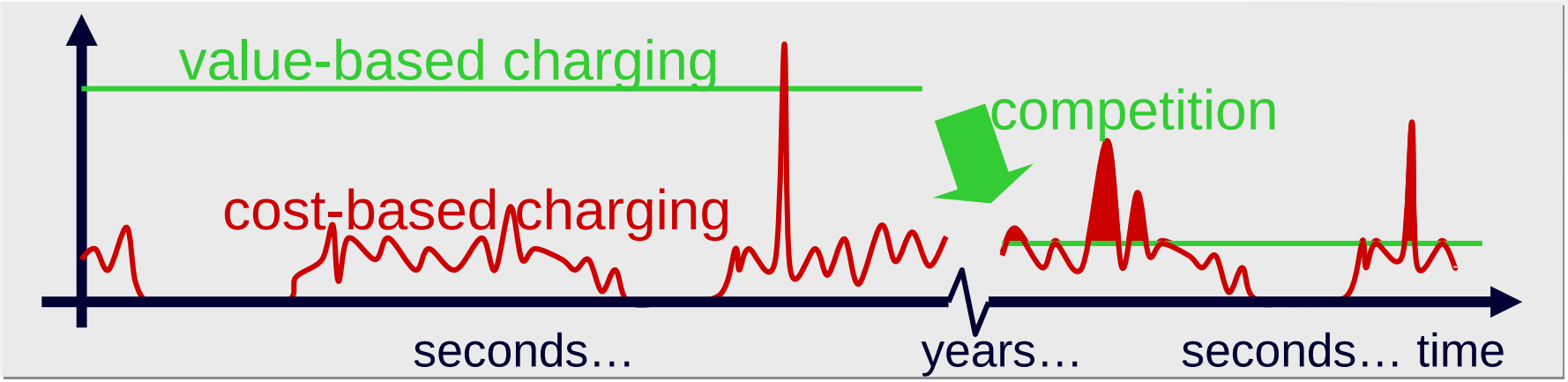
= aggregate downstream pre-congestion in all flows

without measuring flows

value-based charging
& competitive pressure



- instead of flapping around
  - why not just fix the price high?
- fine if you can get away with it



- if charge more than "cost plus normal profit"
  - competitors undercut
- demand exceeds supply
  - nearly half the time

BT

Internet QoS summary

Bob Briscoe

executive summary

congestion accountability – the missing link

- unwise NGN obsession with per-session QoS guarantees
- scant attention to competition from 'cloud QoS'
  - rising general QoS expectation from the public Internet
  - cost-shifting between end-customers (including service providers)
  - questionable economic sustainability
- 'cloud' resource accountability is possible
  - principled way to heal the above ills
  - requires shift in economic thinking – from volume to congestion volume
- provides differentiated cloud QoS without further mechanism
- also the basis for a far simpler per-session QoS mechanism
  - having fixed the competitive environment to make per-session QoS viable

BT

more info...

- Inevitability of policing
  - [BBincent06] 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
  - [Cho06] 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
  - [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
  - [Briscoe08] 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
  - [Briscoe05] Bob Briscoe and Steve Rudkin "Commercial Models for IP Quality of Service Interconnect" BT Technology Journal 23 (2) pp. 171--195 (April, 2005)
- Growth in value of a network with size
  - [Briscoe06] Bob Briscoe, Andrew Odlyzko & Ben Tilly, "Metcalfe's Law is Wrong", IEEE Spectrum, Jul 2006
- Re-architecting the Future Internet:
  - The Trilogy project
- Re-ECN & re-feedback project page:
  [re-ECN] http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/
- These slides
  <www.cs.ucl.ac.uk/staff/B.Briscoe/present.html>

BT

more info on pre-congestion notification (PCN)

- Diffserv's scaling problem

  [Reid05] Andy B. Reid, *Economics and scalability of QoS solutions*, BT Technology Journal, 23(2) 97–117 (Apr'05)

- PCN interconnection for commercial and technical audiences:

  [Briscoe05] Bob Briscoe and Steve Rudkin, *Commercial Models for IP Quality of Service Interconnect*, in BTTJ Special Edition on IP Quality of Service, 23(2) 171–195 (Apr'05) <www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#ixqos>

- IETF PCN working group documents <tools.ietf.org/wg/pcn/> in particular:

  [PCN] Phil Eardley (Ed), *Pre-Congestion Notification Architecture*, Internet Draft <www.ietf.org/internet-drafts/draft-ietf-pcn-architecture-06.txt> (Sep'08)

  [re-PCN] Bob Briscoe, *Emulating Border Flow Policing using Re-PCN on Bulk Data,* Internet Draft <www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#repcn> (Sep'08)

- These slides
  <www.cs.ucl.ac.uk/staff/B.Briscoe/present.html>

BT

further references

- [Clark05] David D Clark, John Wroclawski, Karen Sollins and Bob Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet," IEEE/ACM Transactions on Networking (ToN) 13(3) 462–475 (June 2005) <portal.acm.org/citation.cfm?id=1074049>
- [MacKieVarian95] MacKie-Mason, J. and H. Varian, “Pricing Congestible Network Resources,” IEEE Journal on Selected Areas in Communications, `Advances in the Fundamentals of Networking' 13(7)1141--1149, 1995 http://www.sims.berkeley.edu/~hal/Papers/pricing-congestible.pdf
- [Shenker95] Scott Shenker. Fundamental design issues for the future Internet. IEEE Journal on Selected Areas in Communications, 13(7):1176–1188, 1995
- [Hands02] David Hands (Ed.). M3I user experiment results. Deliverable 15 Pt2, M3I Eu Vth Framework Project IST-1999-11429, URL: http://www.m3i.org/private/, February 2002. (Partner access only)
- [Kelly98] Frank P. Kelly, Aman K. Maulloo, and David K. H. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. Journal of the Operational Research Society, 49(3):237–252, 1998
- [Gibbens99] Richard J. Gibbens and Frank P. Kelly, Resource pricing and the evolution of congestion control, Automatica 35 (12) pp. 1969—1985, December 1999 (lighter version of [Kelly98])
- [ECN] KK Ramakrishnan, Sally Floyd and David Black "The Addition of Explicit Congestion Notification (ECN) to IP" IETF RFC3168 (Sep 2001)
- [Key04] Key, P., Massoulié, L., Bain, A., and F. Kelly, “Fair Internet traffic integration: network flow models and analysis,” Annales des Télécommunications 59 pp1338--1352, 2004 http://citeseer.ist.psu.edu/641158.html
- [Briscoe05] Bob Briscoe, Arnaud Jacquet, Carla Di-Cairano Gilfedder, Andrea Soppera and Martin Koyabe, "Policing Congestion Response in an Inter-Network Using Re-Feedback“ In: Proc. ACM SIGCOMM'05, Computer Communication Review 35 (4) (September, 2005)
- [Siris] Future Wireless Network Architecture <www.ics.forth.gr/netlab/wireless.html>
- Market Managed Multi-service Internet consortium <www.m3i_project.org/>

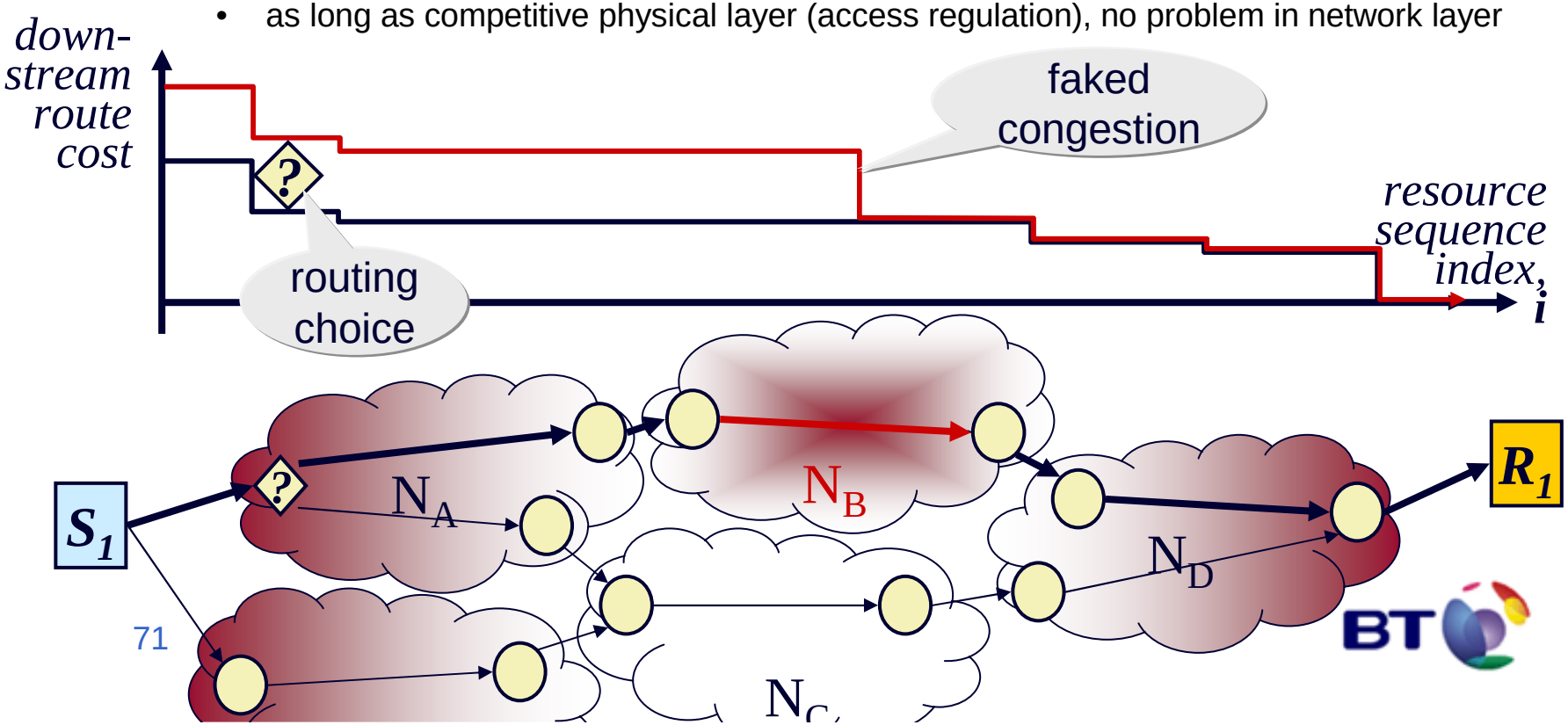
BT

Internet QoS
the underlying economics

Q&A

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

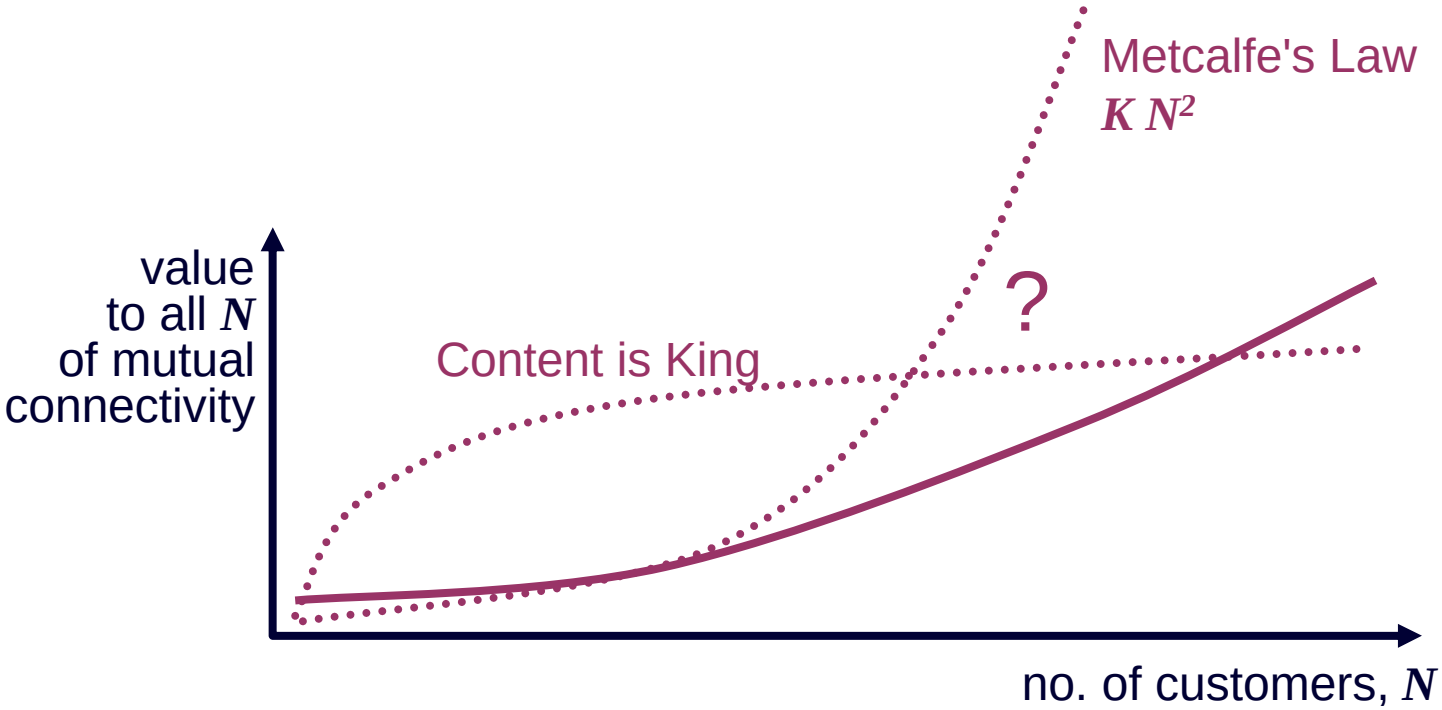
- network
  - turn on explicit congestion notification in routers (already available)
  - deploy simple active policing functions at customer interfaces around participating networks
  - passive metering functions at inter-domain borders
- terminal devices
  - (minor) addition to TCP/IP stack of sending 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
  - IETF recognises it has no process to change its own architecture
  - Apr'07: IETF supporting re-ECN with (unofficial) mailing list & co-located meetings

Internet QoS
the value of connectivity

Bob Briscoe

Content is King or The Long Tail?

community & social networking, interest groups



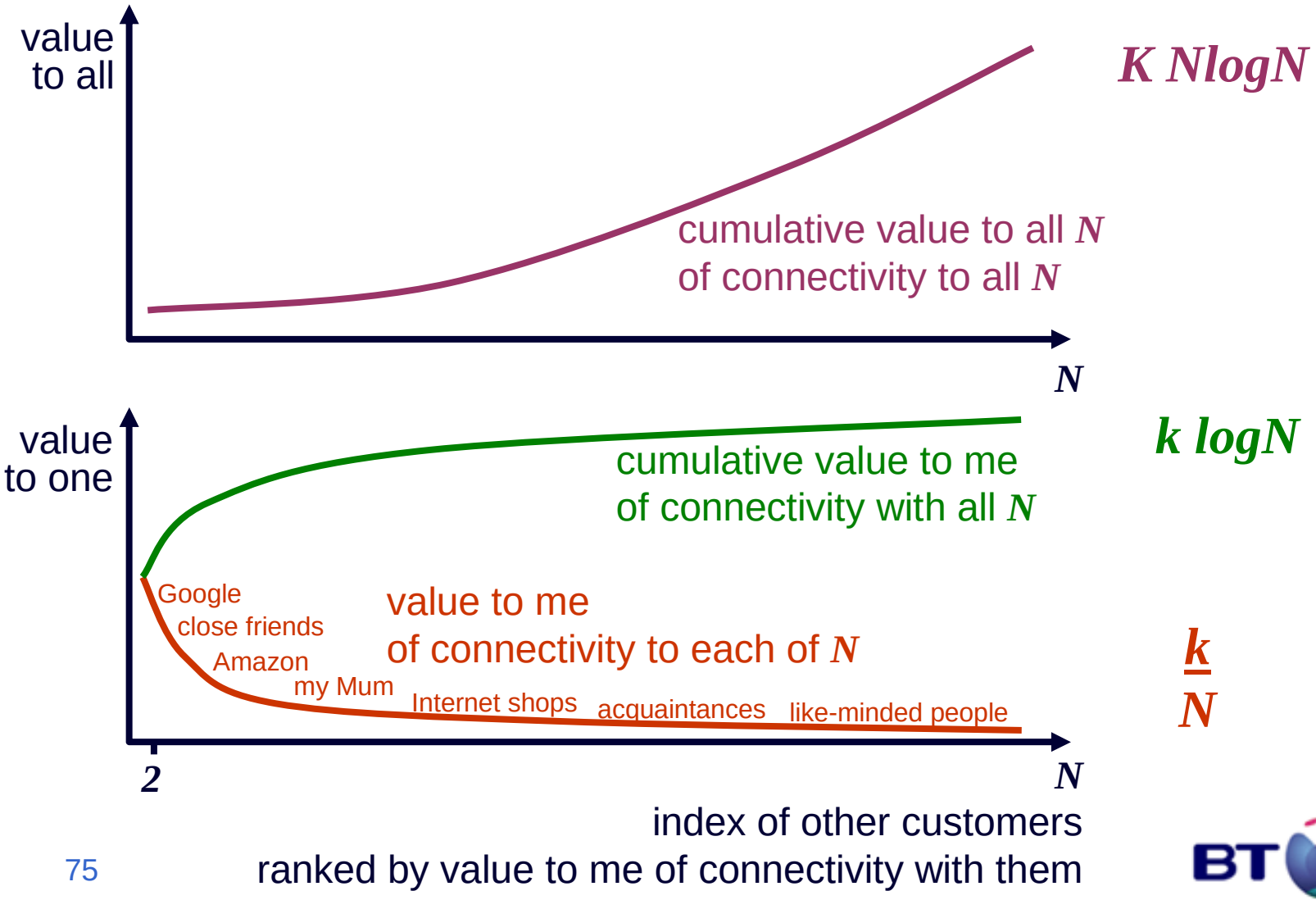
- the long tail effect eventually predominates
- but not as strongly as Metcalfe's Law predicted

Odlyzko, "Content is Not King"

Briscoe, Odlyzko & Tilly, "Metcalfe's Law is Wrong"

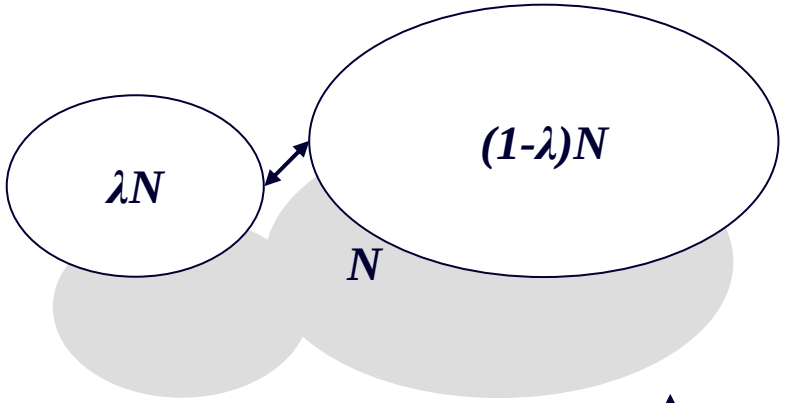
BT

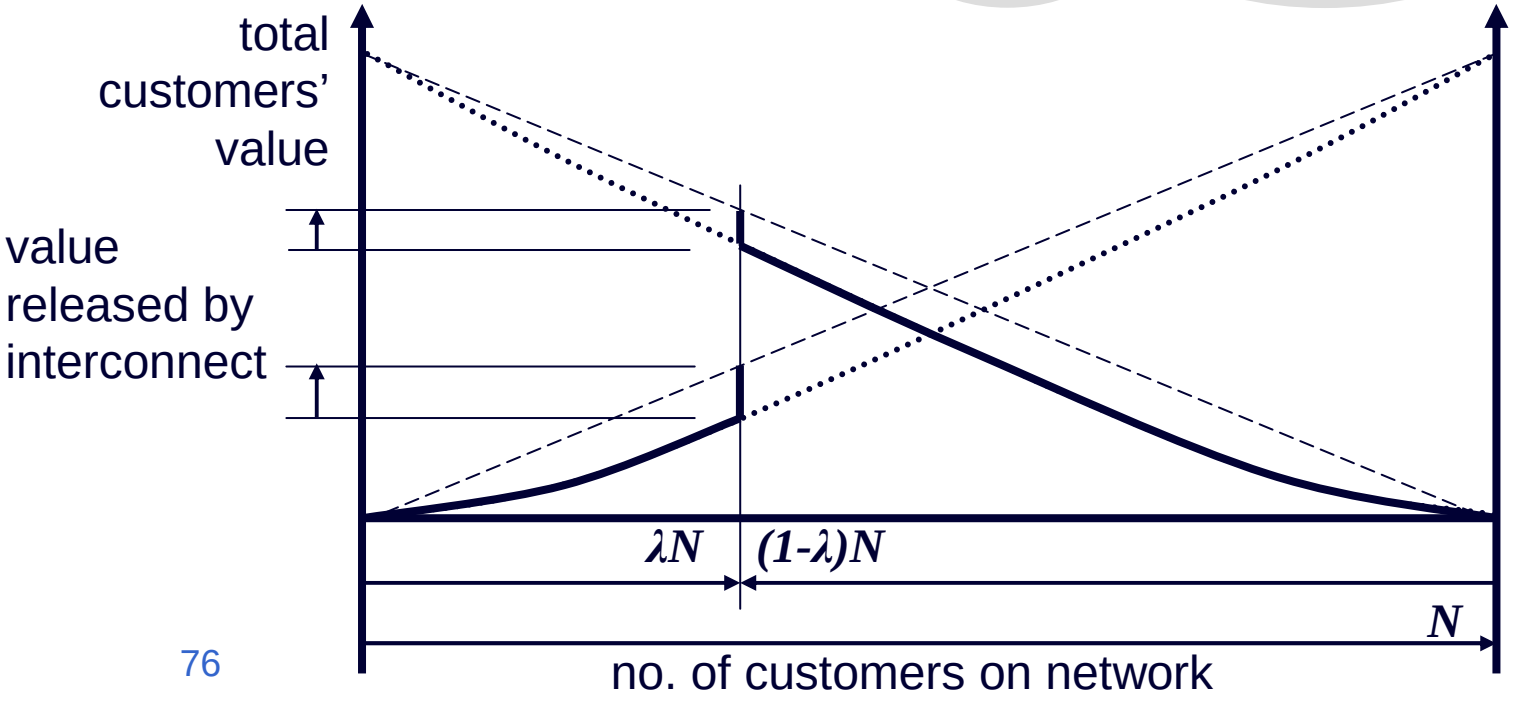
potential peers: value in numbers



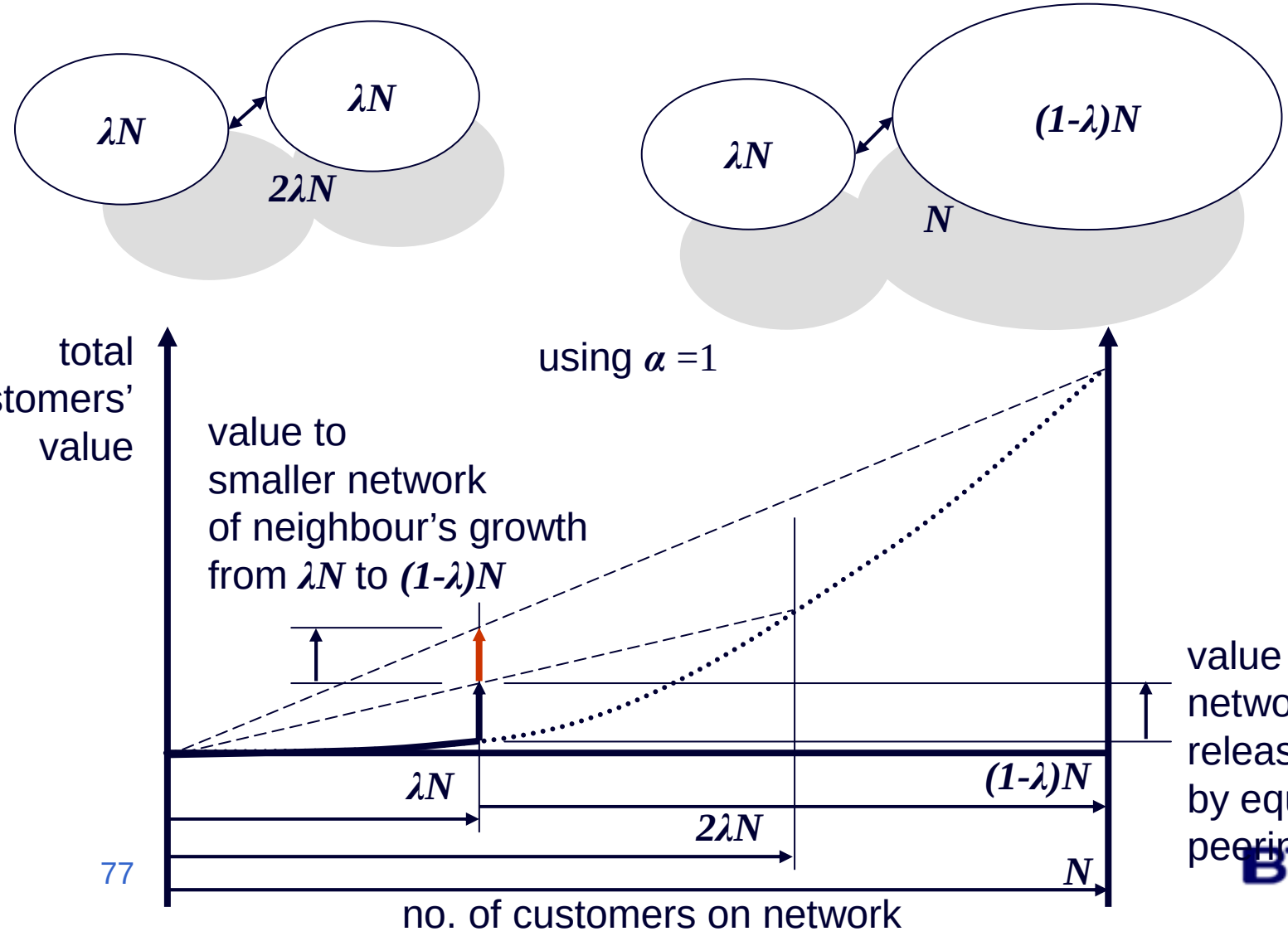
growth in potential network value

by scaling & interconnect



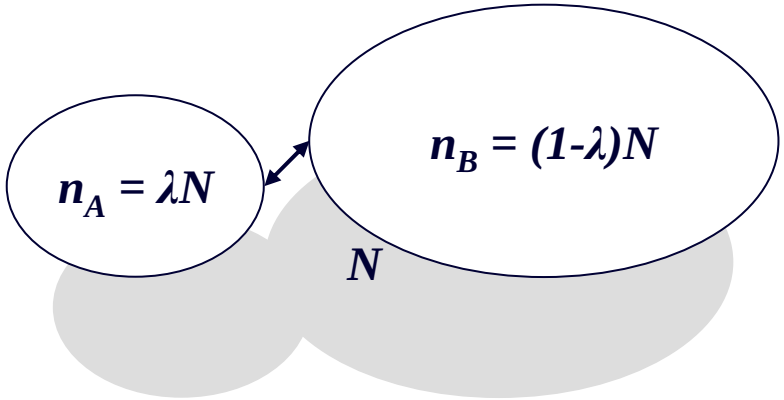


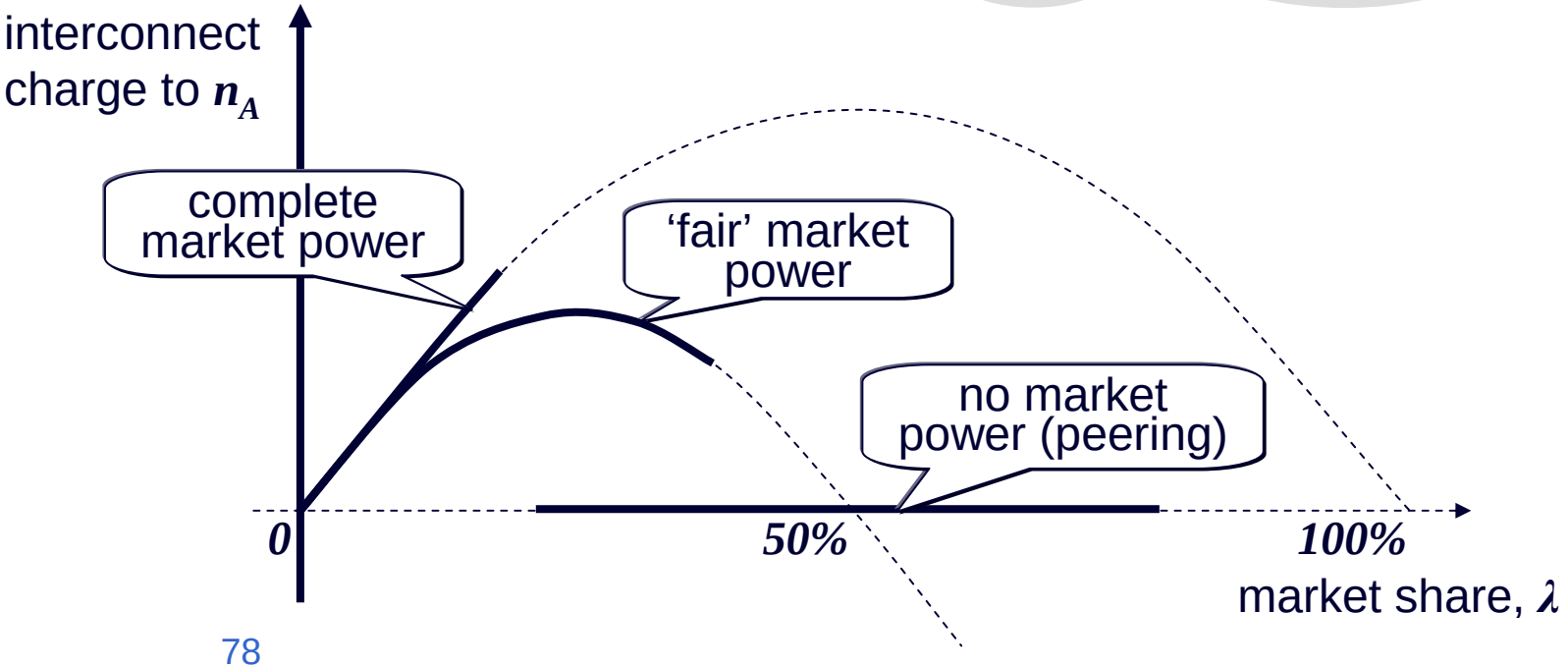
interconnect settlement



charging for interconnect
within the same market

legend
assumptions
no longer hold





charging for interconnect

within the same market

legend
assumptions
no longer hold