IP QoS interconnect

business impact of new IETF simplification

Bob Briscoe

Chief Researcher, BT Group

Aug 2007

acks: Steve Rudkin, BT Retail
Andy Reid BT Group
PCN team, BT Group

scope of talk

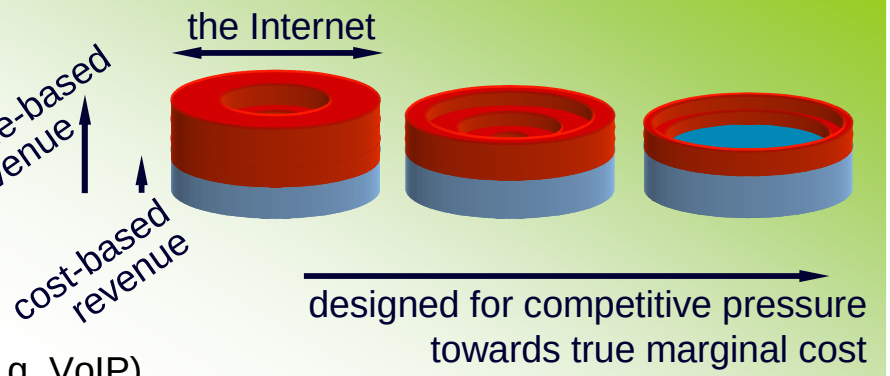
IP quality of service for inelastic apps

- inelastic applications
  - streamed media needing min bandwidth / latency / jitter
    - primarily interactive voice & video (fixed & mobile)
- new approach to QoS in the data plane
  - charging for session signalling ≠ charging for session QoS
- in UK alone, prediction (in 2005) for 2009
  - 39% of UK comms services revenue will depend on IP QoS interconnect

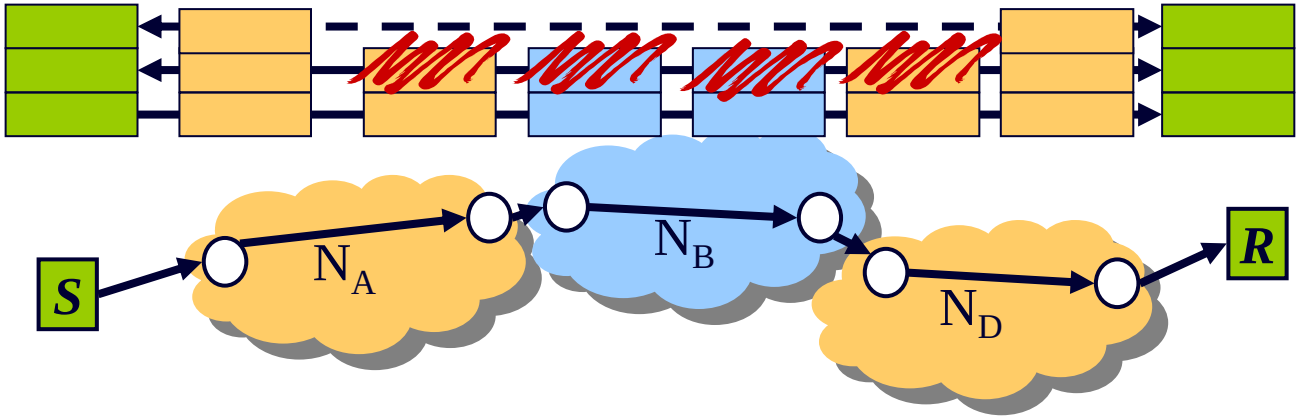
    = 79% from apps that depend on QoS
    x 49% that depend on interconnect
    and will have shifted to IP interconnect by 2009

BT

summary



- 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**

BT

PCN

the wider it is deployed
the more cost it saves

legend

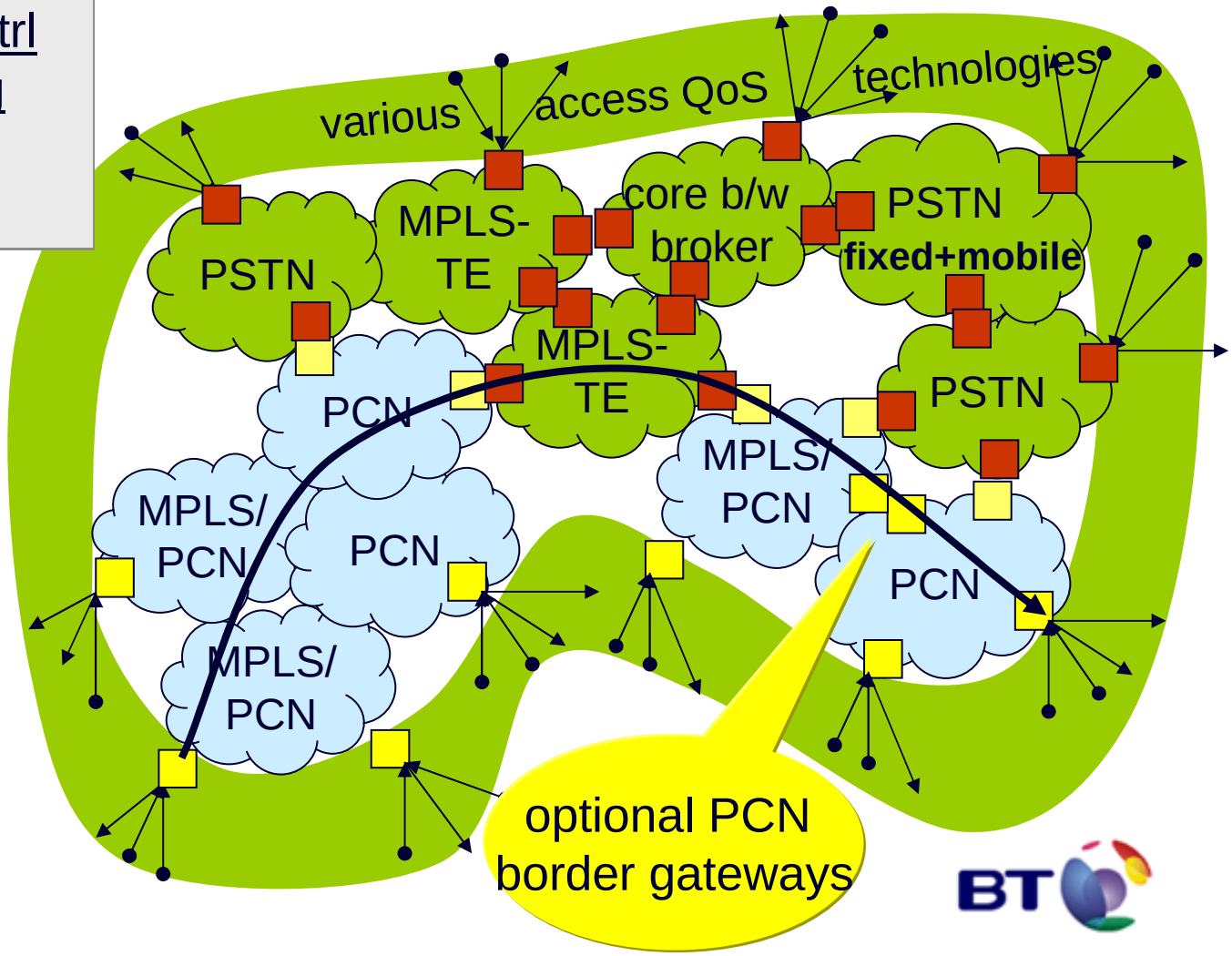
- connection-oriented (CO) QoS
- PCN QoS

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

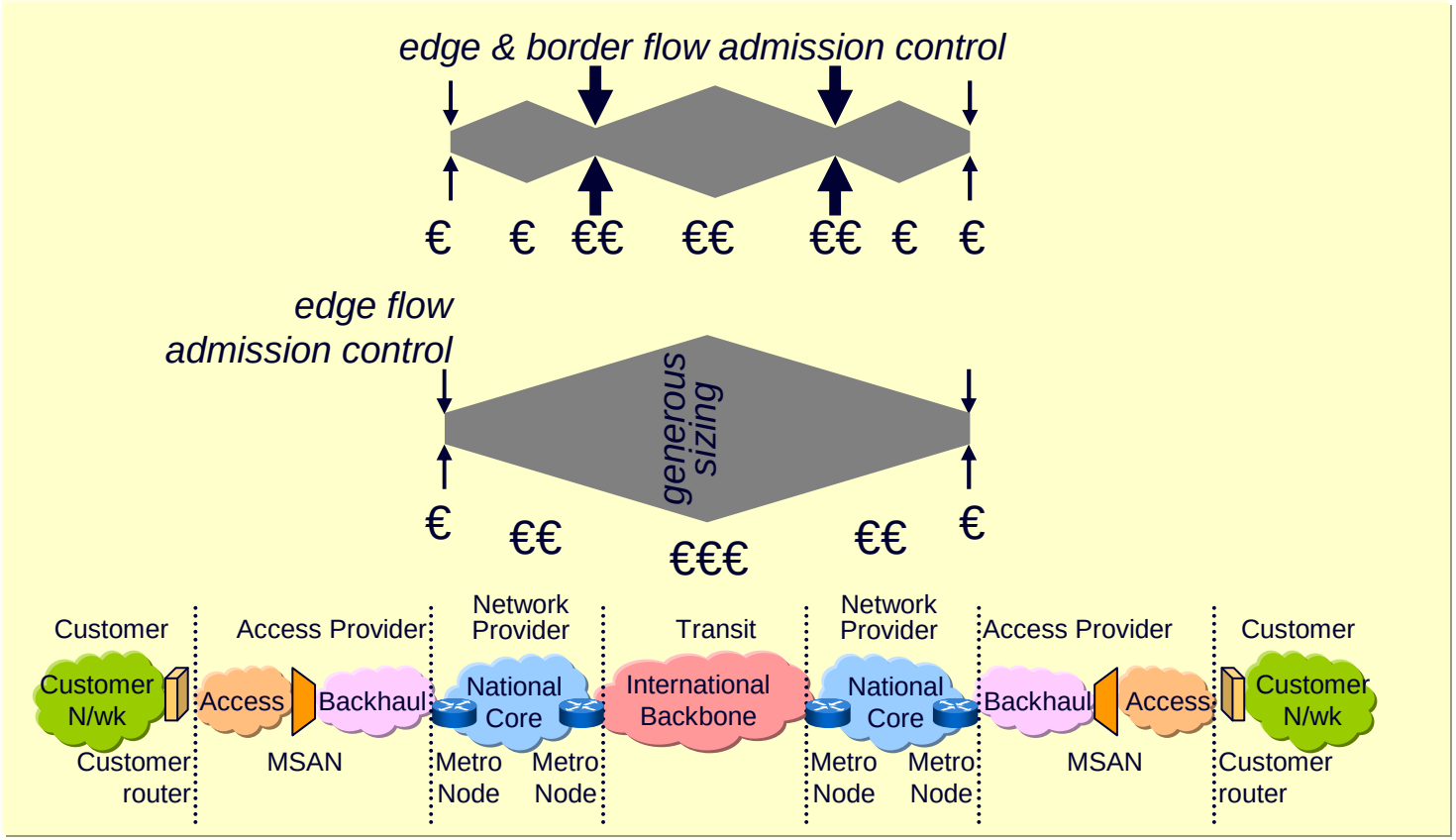
- main IETF PCN standards scheduled for Mar'08
  - main author team from companies on right (+Universities)
  - wide & active industry encouragement (no detractors)
- IETF initially focusing on *intra*-domain
  - but chartered to "keep *inter*-domain strongly in mind"
  - re-charter likely to shift focus to interconnect around Mar'08
- 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

- 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)
  - 2005-07: introduced to IETF & continually pushing standards onward
  - 2006-07: extending to MPLS & Ethernet with vendors

classic trade-off with diseconomy of scale either way

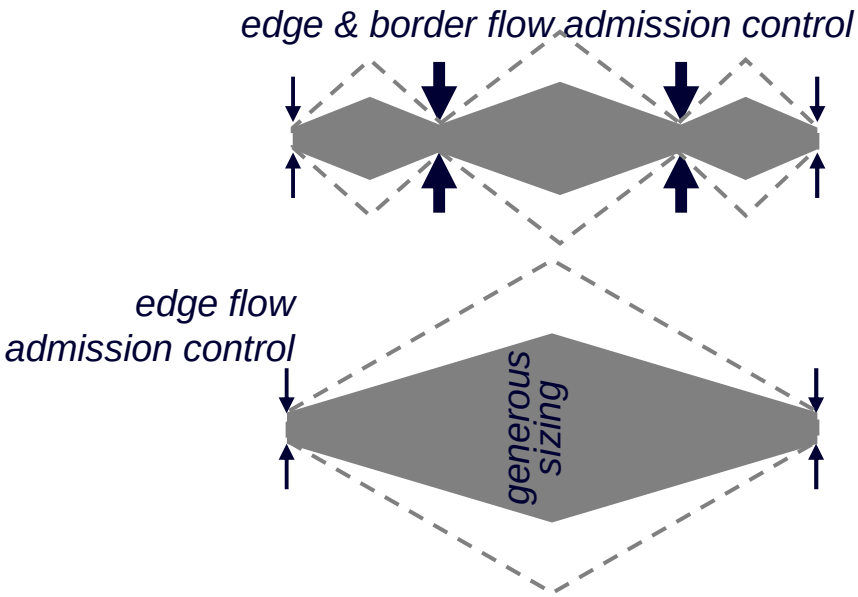
seen in all QoS schemes before PCN

- 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



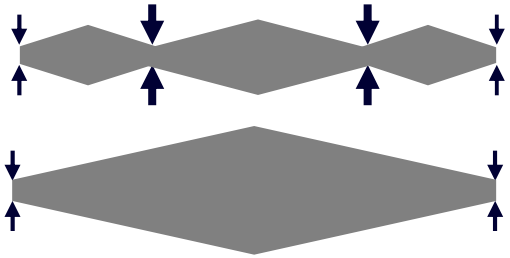
current Diffserv interior link provisioning
for voice/video expedited forwarding (EF) class

- 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)



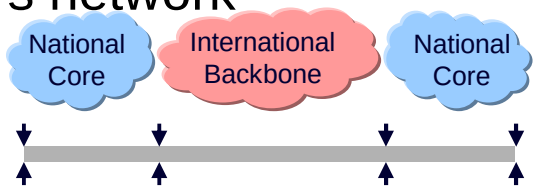
- PCN: radical cost reduction
  - compared here against simplest alternative – against 6 alternatives on spare slide
  - no need for any Diffserv generous provisioning between admission control points
    - 81% less b/w for BT's UK PSTN-replacement
    - ~89% less b/w for BT Global's premium IP QoS
    - 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

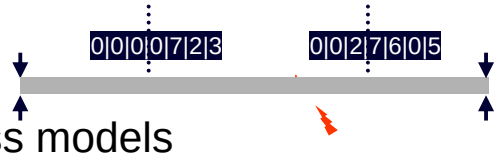
BT

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

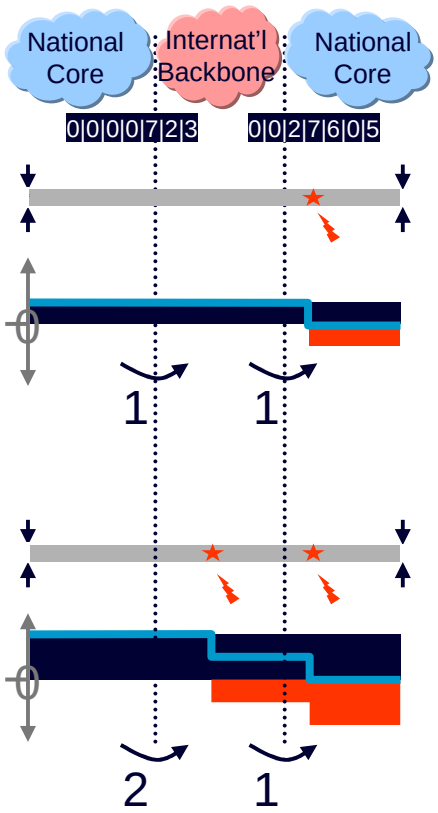


  - no change needed to edge / customer-facing business models
  - not selling same things across interconnects as is sold to end-customer
  - but bulk interconnect SLAs with penalties for causing pre-congestion can create the same guaranteed retail service

BT

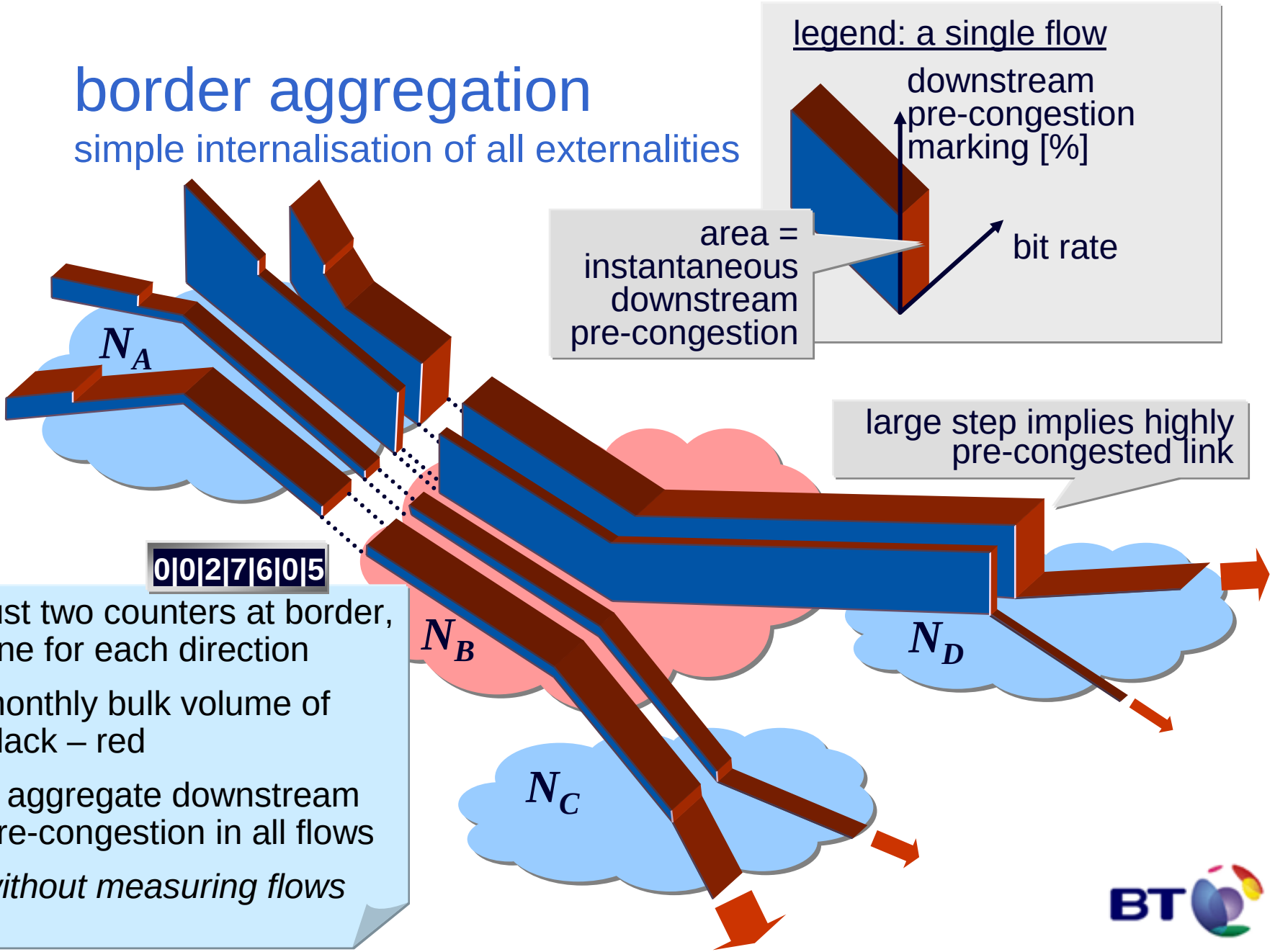
accountability of sending networks

- 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)
  - 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

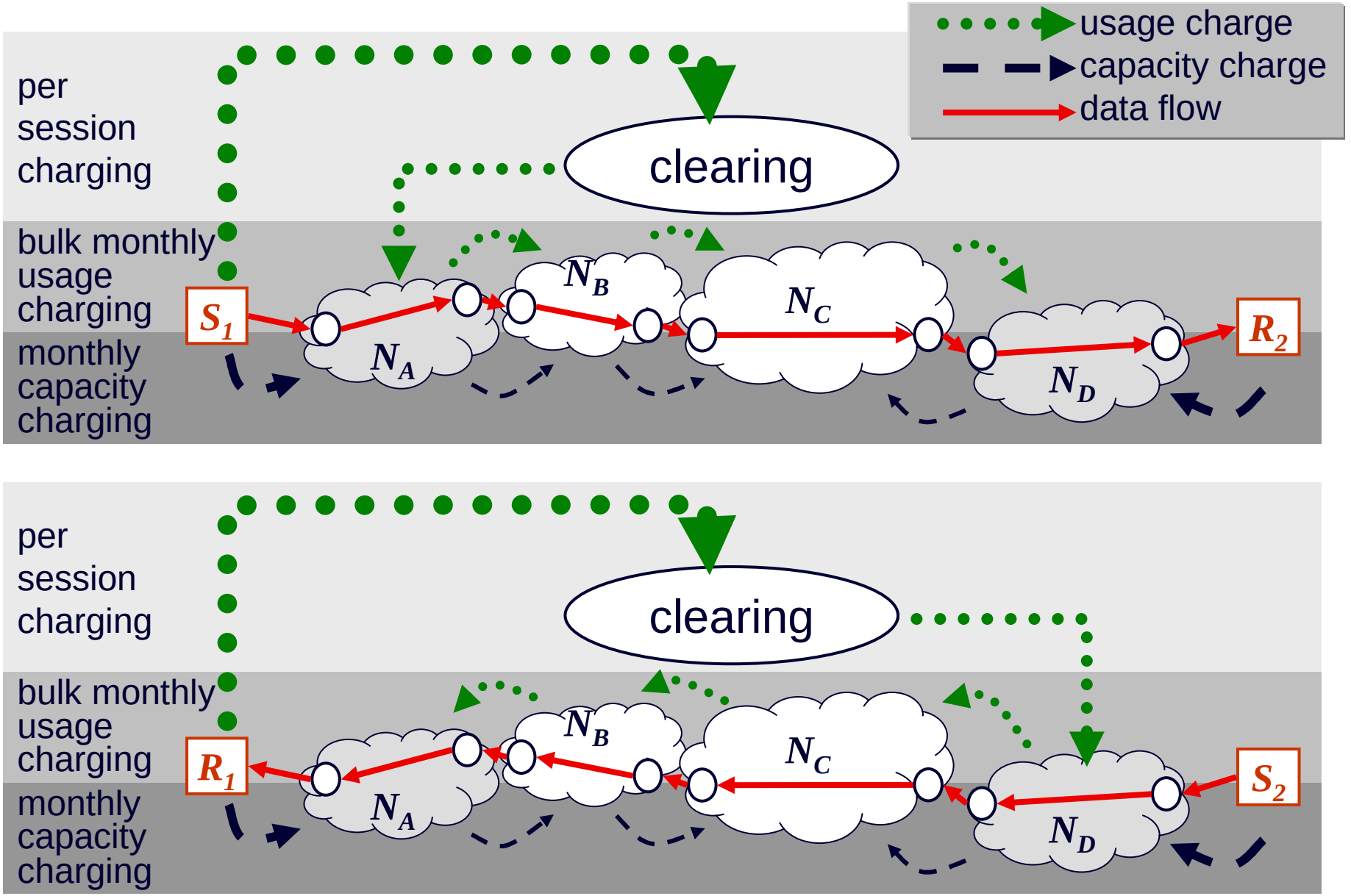
next steps
where the IETF stops

- IETF supplies the metric
  - chosen based on economics: competition driving to marginal cost
- operators build/agree interconnect business models
  - will need to thrash out the business implications in depth
- the necessary *downstream* pre-congestion metric
  - requires a valuable packet header bit that others want
  - debate will come to a head during 2008

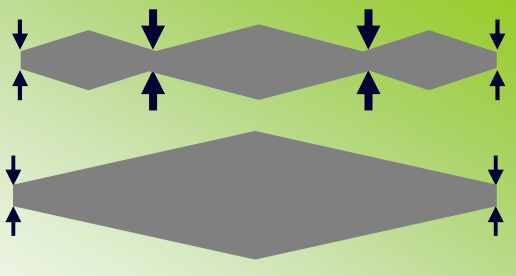
BT

possible business model around edge-edge PCN

duplex call with edge-to-edge clearing



in conclusion



- a new IETF technology: pre-congestion notification (PCN)
  - carrier-grade QoS but intrinsic cost competes with no-QoS services

- scheduled for 2008
  - intra-domain standards Q1’08
  - interconnect depends on outcome of IETF debate during 2008
    - tremendous achievement: grail of last 14 years of Internet QoS effort
    - fully guaranteed global *inter*-network QoS with economy of scale
- business model implications
  - core networks could fully guarantee sessions without touching sessions
  - some may forego falling session-value margins to compete on cost

more info

- Diffserv's scaling problem
  - 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:
  - 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:
  - Phil Eardley (Ed), *Pre-Congestion Notification Architecture*, Internet Draft <www.ietf.org/internet-drafts/draft-ietf-pcn-architecture-00.txt> (Aug'07)
  - Bob Briscoe, *Emulating Border Flow Policing using Re-ECN on Bulk Data,* Internet Draft <www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#repcn> (Jun'07)
- These slides <www.cs.ucl.ac.uk/staff/B.Briscoe/present.html#0709ixqos>

BT

IP QoS interconnect

business impact of new IETF simplification

spare slides

QoS trade-offs FAQ -

comparative evaluation -

how PCN works -

usage charging model today -

Q&A

classic cost trade-offs for assured QoS FAQ

Q Why are IP admission control smarts costly at trust borders?

A Flows switch between aggregates at borders
so must police packet rate in each microflow,
otherwise cheating networks request low b/w but take high.

Q Why does generous sizing have to be so costly?

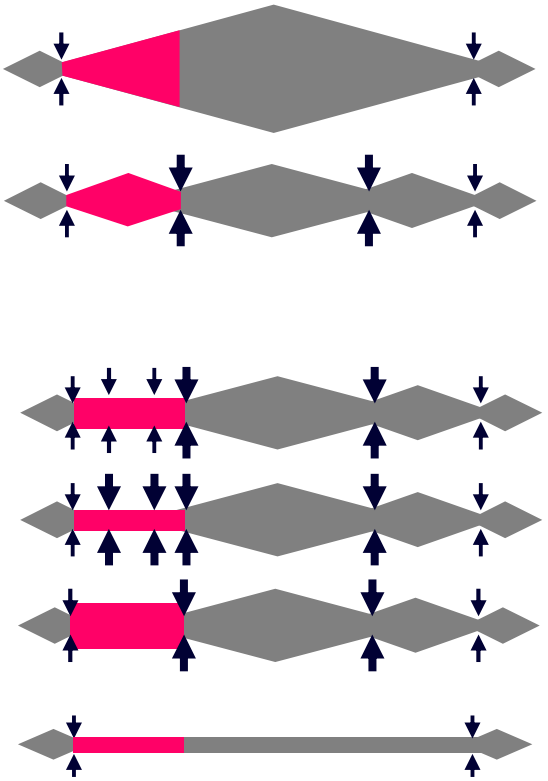
A Sufficient capacity for anomalies: failures, disasters, flash crowds.
No matter how much oversizing,
always residual risk of overload breaking up *all* calls in progress

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	£	£	£

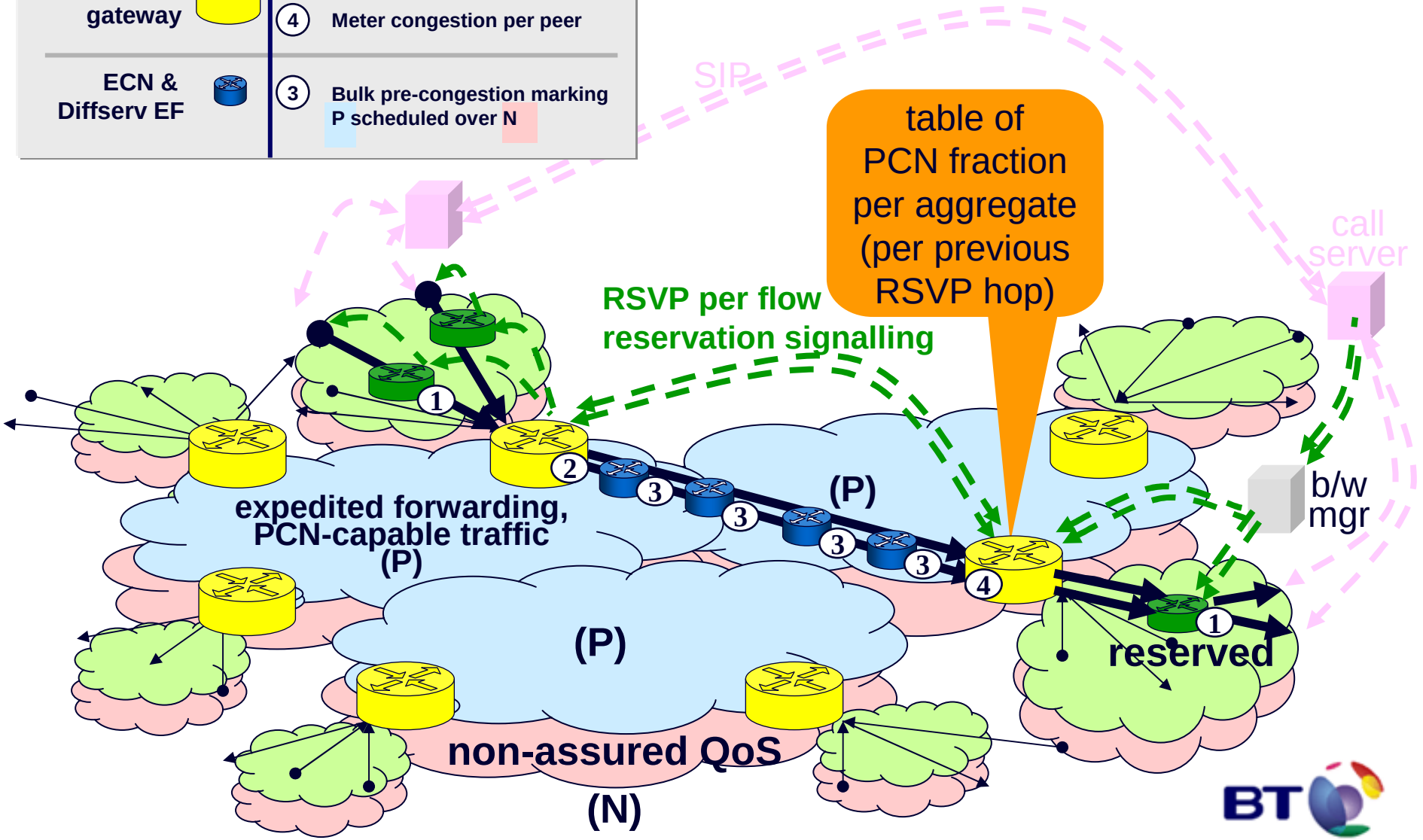


downside to PCN: not available yet

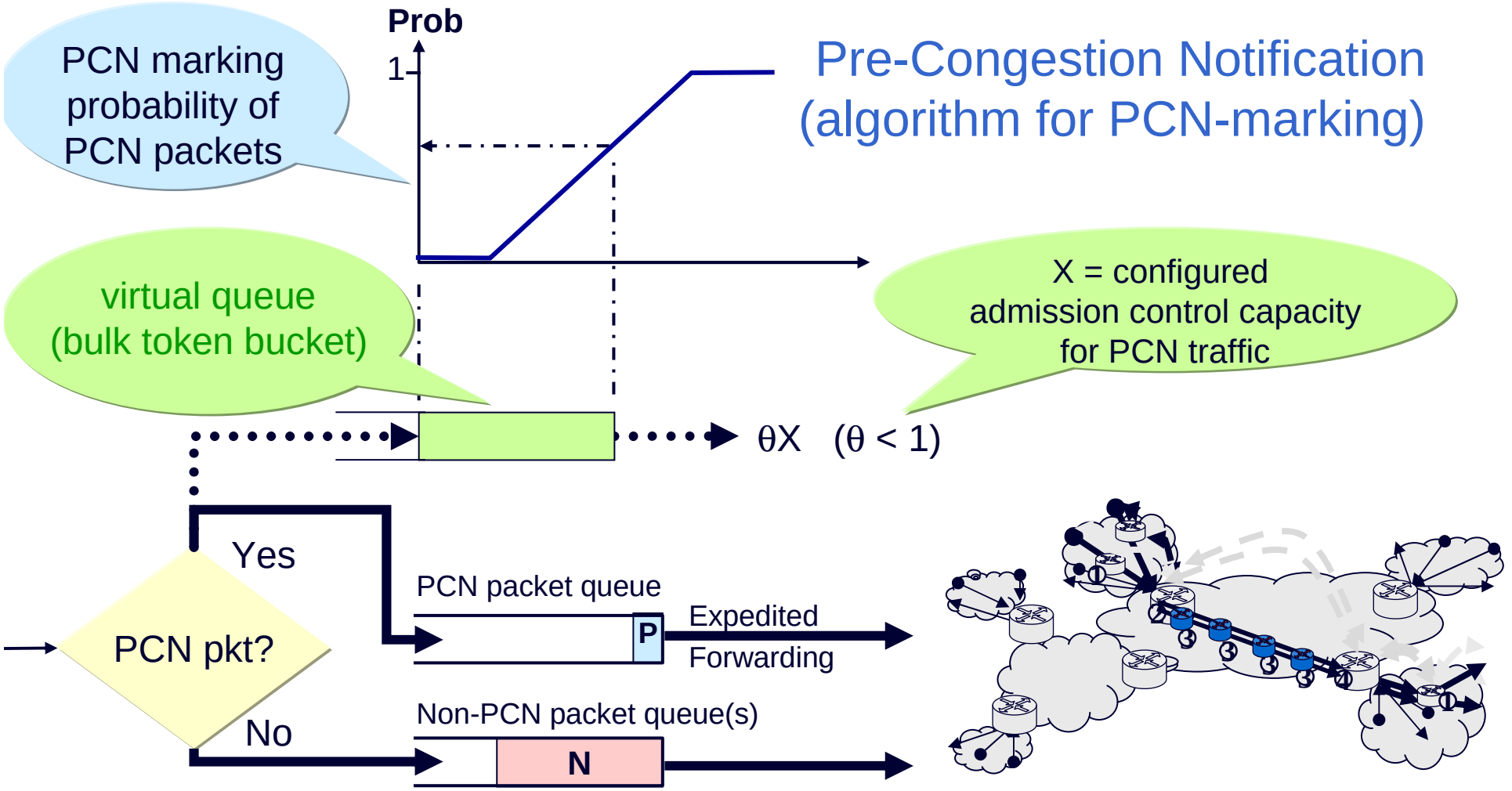
PCN system arrangement

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
ECN & Diffserv EF	3 Bulk pre-congestion marking P scheduled over N

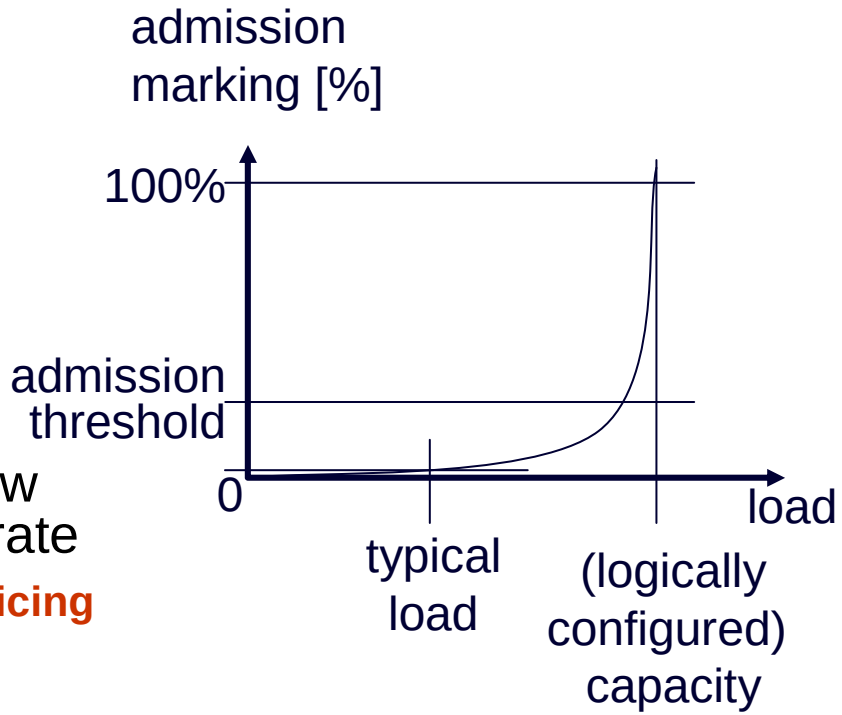


Pre-Congestion Notification (algorithm for PCN-marking)



- 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

solution rationale



- <0.01% packet marking at typical load
  - addition of any flow makes little difference to marking
- penalties to ingress of each flow appear proportionate to its bit rate
  - **emulates border flow rate policing**
- as load approaches capacity
  - penalties become unbearably high (~1000x typical)
  - insensitive to exact configuration of admission threshold
  - **emulates border admission control**
- neither is a perfect emulation
  - but should lead to the desired behaviour
  - fail-safes if networks behave irrationally (e.g. config errors) – see draft

possible current business model

edge-to-edge clearing