interprovider per-flow QoS
BT's dilemma
distributed vs. centralised

Bob Briscoe
BT Group CTO, Networks Research
Oct 2005

context

- inter-provider QoS for inelastic applications
  - including PSTN replacement
  - ~30%-50% of the bits may be inelastic
- large-scale deployment: “national infrastructure”
  - IP-based platform: BT’s 21C network
  - state of the art, but only using technology for sale now
- wholesaling for different retail business models and access nets
  - cellular backhaul, DSL+WiFi, satellite, ...
  - free VoIP over BE, session charged VoIP over BE, admission controlled VoIP
  - and “growing demand” for inelastic apps other than VoIP

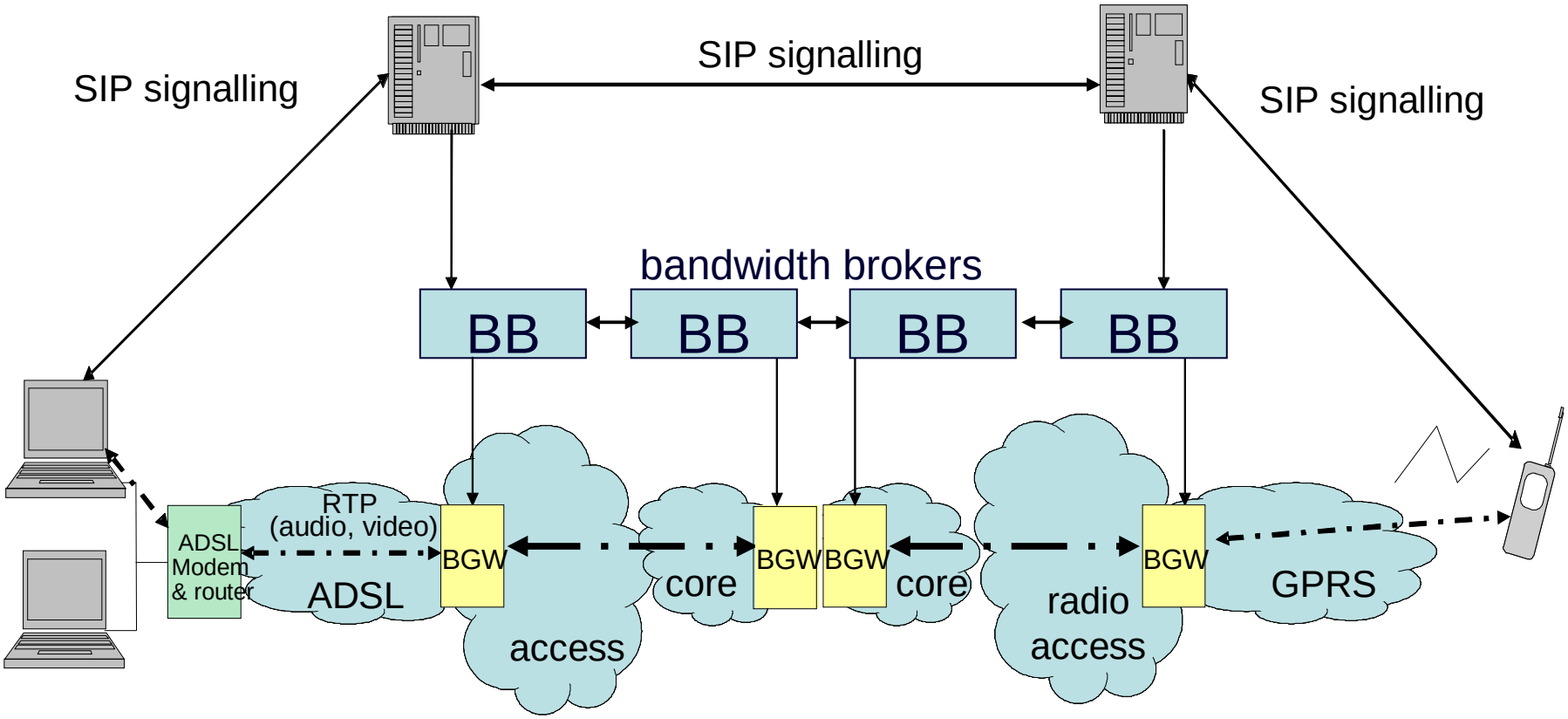
menu

- introductory remarks
- walk through the sequence of candidate solutions
  - a carrier is looking for why it wouldn’t choose a solution
  - why risk-aversity is as important as business opportunity
- simple proposal that hits sweet spot?
  - enables innovation
  - no features to scare carriers away

caveat

- a personal view, not the position of BT
- one step removed from BT’s architecture decisions
  - details may be sketchy
- reverse-engineered interpretation of the motivations
- based on rumour, innuendo and sometimes even the views of those with first hand knowledge
- generalised enough to be any telco

2004/5: centralised bandwidth brokers



- note: this whole inelastic transport service is itself a VPN coexisting alongside other VPNs

why bandwidth brokers?

- every BT QoS expert thinks someone else decided
- a given before any decision was requested
- my reverse-engineered suspicion:
  “outsource the hard bit”
  - buying a box means QoS not so dependent on own design
    - responsibility of box vendor
  - box vendor gets a bigger cut by taking more responsibility
    - sold to technical management rather than technical experts
- decision is truly burned-in now
- summary: de-risk a risky area
- perhaps I’m cynical

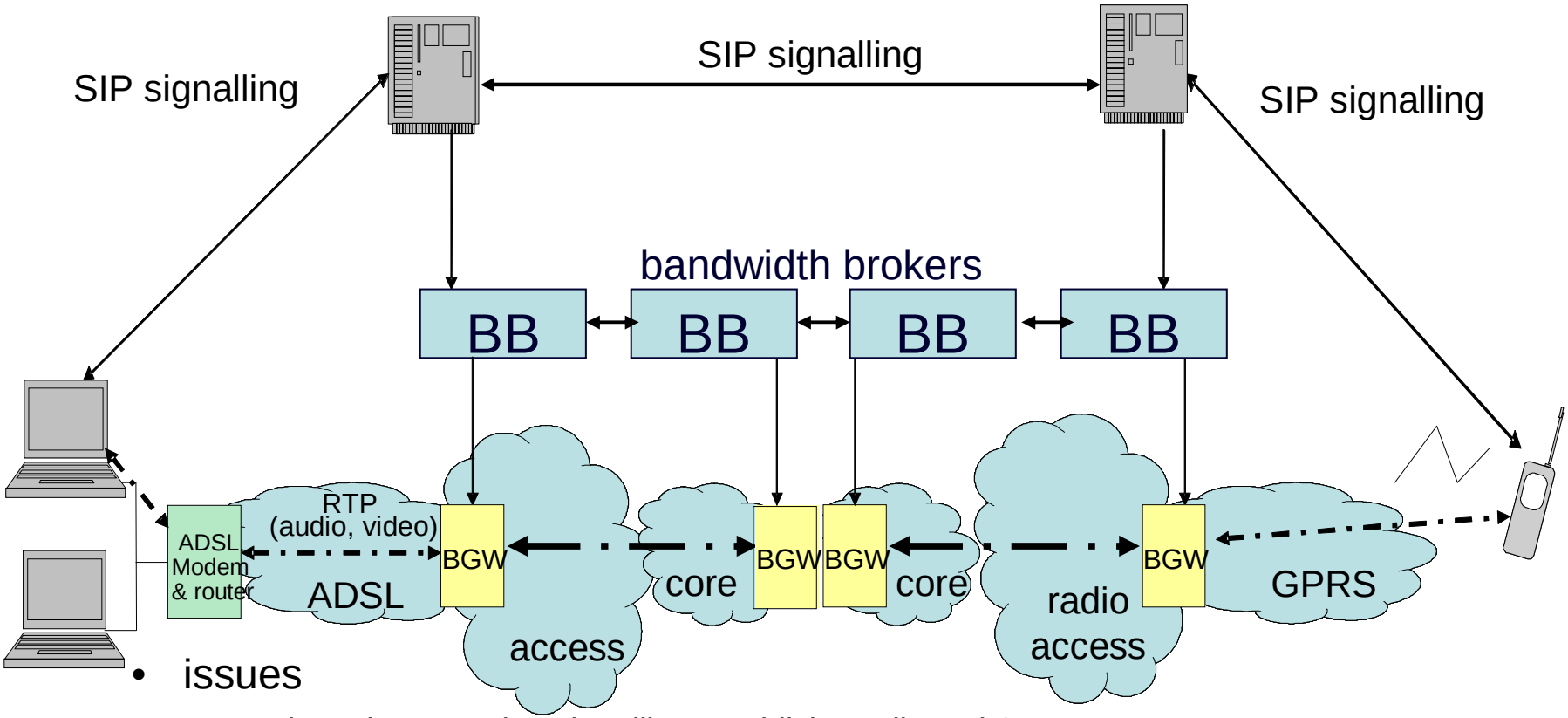
BT

step back: why CAC in the first place?

- FUD: fear, uncertainty and doubt
- Diffserv only: out of control if unforeseen events
  - link failures, flash crowds
- would you be the one responsible for replacing the national infrastructure with something that has even a tiny risk of 100,000s of calls all failing at once?
  - perhaps live on a phone-in show
- telco instinct for robustness by engineering
  - can’t avoid Diffserv’s occasional episodes
  - can engineer down the possibility of a centralised box failing
    - by replication

BT

2004/5: centralised bandwidth brokers



- issues
  - how does session signalling establish media path?
    - to place border media controllers (also a problem with just two)
    - and determine BB path
  - b/w broker interworking isn't being standardised
  - b/w broker for core untried scaling challenge: expensive

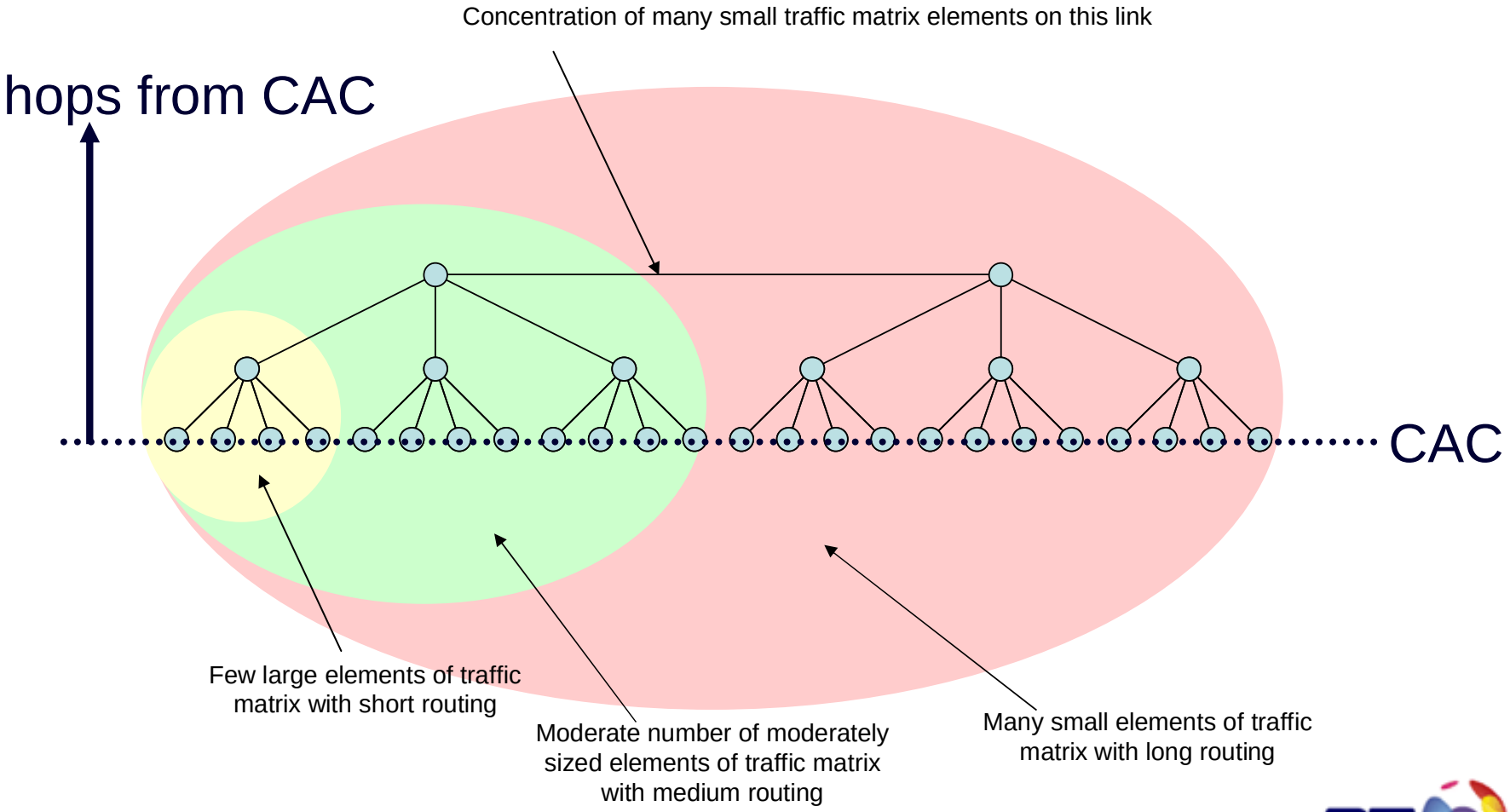
provisioning for inter-provider Diffserv [Reid05]

- scenario:
  - CAC in ingress and egress access networks
  - interconnected Diffserv in cores/backbones
- idea: limit variance of aggregates on interior links
  - by CAC limiting variance at ingress and egress
- but how fast does the effect of CAC wear off, the more hops away it is?
  - variance grows ~linearly with hops from where CAC is applied (ingress & egress)
  - congestion probability may* grow ~exponentially with variance
- to achieve same congestion probability on interior as edge links
  - must provision disproportionately more generously, the more hops from CAC
- exacerbated by targeted marketing confining largest flows locally
  - leaving bias toward more smaller flows on interconnect
  - correlation effects worse if there are more flows to correlate

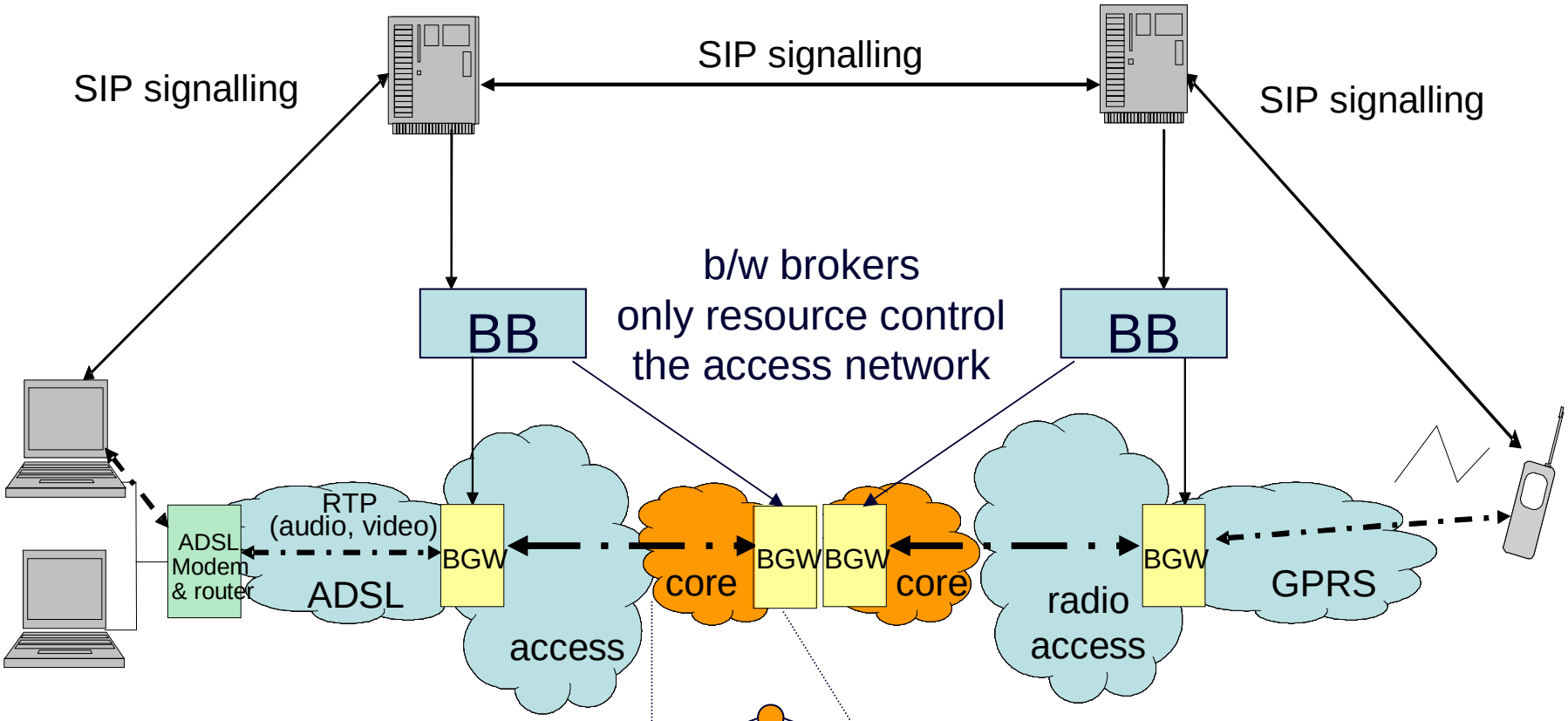
* exact growth depends on shape of traffic probability distribution
  - so simulation results depend heavily on distribution chosen
  - meaning: we won't know for sure until we've tried it for real

BT

long topologies for inter-provider QoS



mid-2005: non-blocking core



non-blocking core
- fully meshed
- fully load balanced

costs wouldn't scale
- but OK for BT's size

centralised b/w broker + distributed non-blocking core

outstanding issues

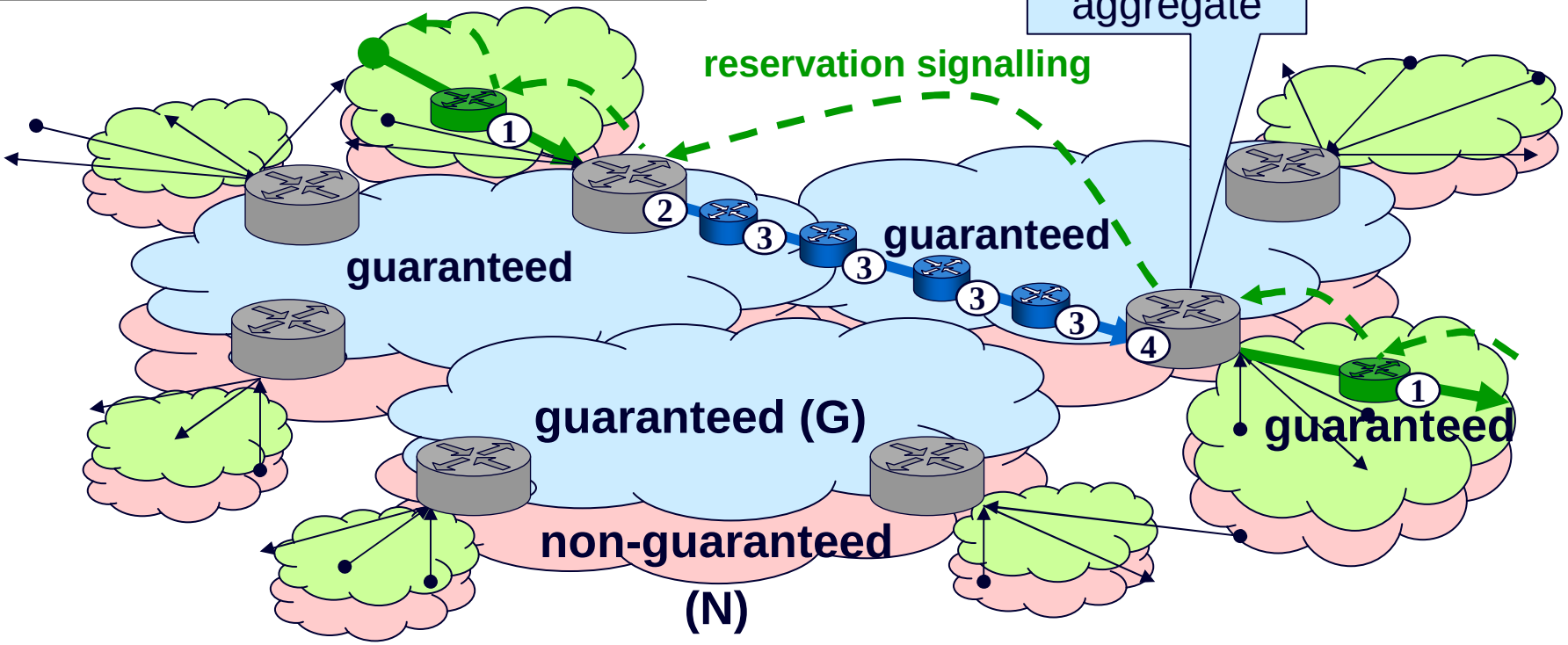
- interconnecting cores
  - two non-blocking cores don't make a non-blocking interconnect
  - unless you connect every BT core node to every core node of the other operator
  - current solution: per-flow CAC at border gateway
- if backbone transit between cores
  - requires multiple border gateways to divide load
- currently border gateway boxes can cope (?)
  - designed for transcoding to PSTN per flow anyway
  - longer term still need radical cost reduction
- PSTN replacement only
  - not for general inelastic flows, range of mean bandwidths, VBR etc

GQS system arrangement

IP routers	Data path processing
Reservation enabled	1 Reserved flow processing
RSVP/ECN gateway	2 Policing flow entry to G 4 Meter congestion per peer
ECN only	3 Bulk ECN marking G prioritised over N

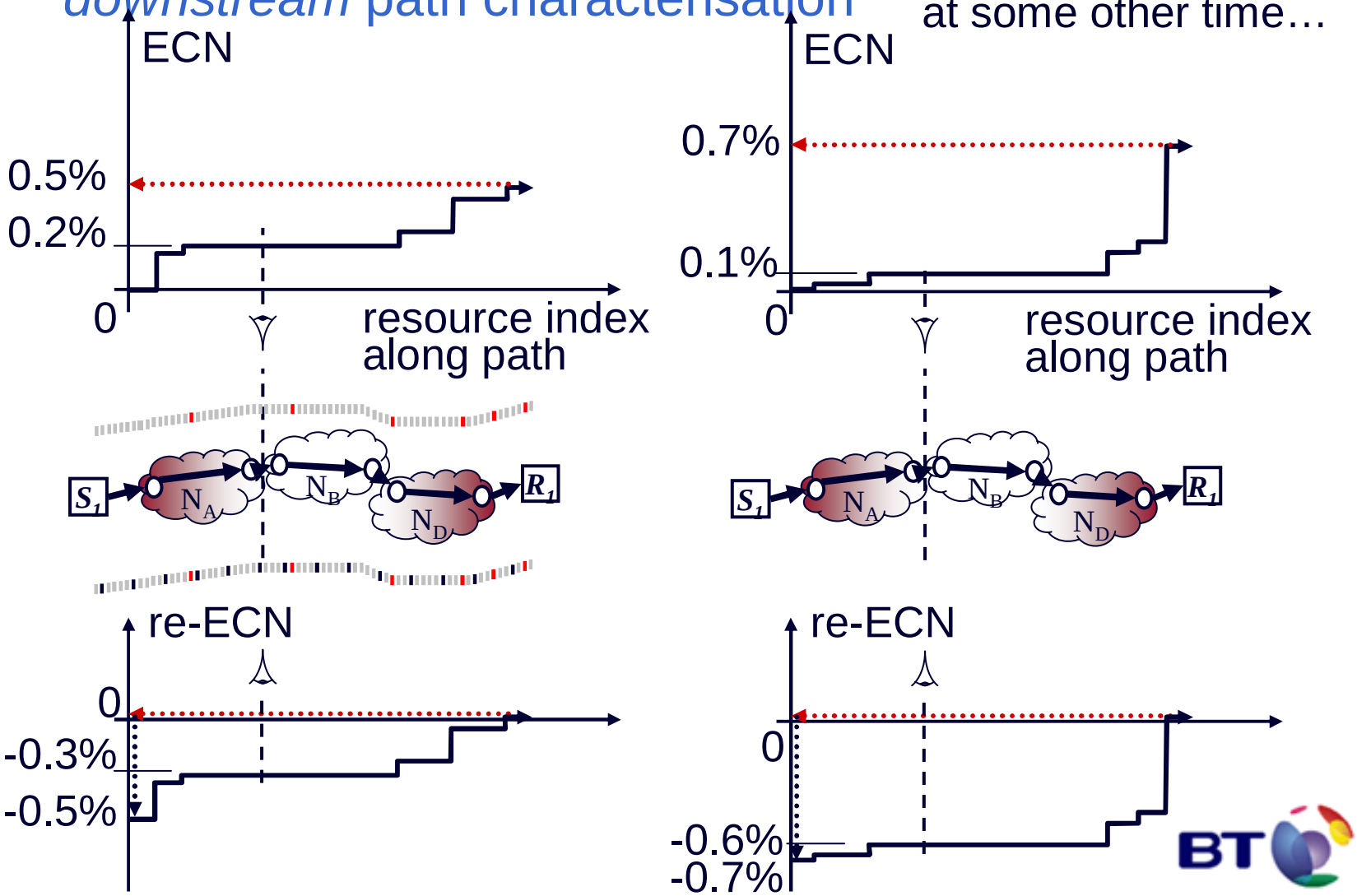
table of ECN fraction per previous RSVP hop aggregate

[Briscoe05a]



- distributed but deterministic CAC: meets carrier-scale reqs
- handles unexpected interior events gracefully
- still research, but gaining traction within BT for some time, and recent strong wider interest

accountability architecture
re-ECN: receiver-aligned ECN [Briscoe05]
downstream path characterisation



summary

- Diffserv with edge CAC will occasionally fail large numbers of inelastic flows simultaneously [Reid05]
- unlikely to be solution of choice for those with carrier-scale obligations
- even if in practice the system will fail nearly as often for other reasons (human error, natural disaster)
- current solution:
  - bandwidth brokers for access and non-blocking topology for core
- carrier-scale QoS interconnect for inelastic flows
  - still problematic
- distributed measurement-based admission control (MBAC)
  - current focus of attention [Briscoe05a]
  - part of wider, principled approach to Internet QoS [Briscoe05b]

more info

- [Reid05] Andy B. Reid, Economics and scalability of QoS solutions, BT Technology Journal, 23(2) pp97 – 117 (April 2005)
- [Briscoe05] Bob Briscoe et al, Policing Congestion Response in an Inter-network using Re-feedback, in Proc ACM SIGCOMM'05, Computer Communications Review **35**(4) (Sep 2005) <http://www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#refb>
- [Briscoe05a] Bob Briscoe et al, An architecture for edge-to-edge controlled load service using distributed measurement-based admission control, Internet Draft <draft-briscoe-tsvwg-cl-architecture-00.txt> (Jul 2005)
- [Briscoe05b] Bob Briscoe and Steve Rudkin, Commercial Models for IP Quality of Service Interconnect, in BTTJ Special Edition on IP Quality of Service, 23(2) (Apr 2005) <http://www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#ixqos>

inter-provider per-flow QoS next steps?

Bob Briscoe
BT Group CTO, Networks Research
Oct 2005

next steps?

- a white paper on inter-provider per-flow QoS?
- proving the ideas in the large
  - an inter-operator test bed
- which standards bodies and industry fora for which issues?
  - IETF for component technologies
  - ITU/ETSI/PacketCable/DSLForum for component selection
  - ETNO etc for business model, regulatory
  - CFP for all at once?
- thorny technical ‘detail’: ECN in MPLS
- edge-edge CAC: first step to something more open?
- …?

BT

suggested agenda if next CFP meeting

- Inter-provider business models in depth
  - per-flow, per-session or bulk accounting;
  - simplex or duplex, multi-flow sessions, conferencing & multipoint
  - pricing metrics: per volume? per congestion? time of day?
  - Sender/originator pays, 800 service
  - sessions spanning multiple models: over enterprise & public networks with and without QoS support
  - layered business models
  - charging after partial failure, etc
- Security, policing and anti-cheating issues,
- Provisioning/management/accounting/metering issues,
- ??

BT