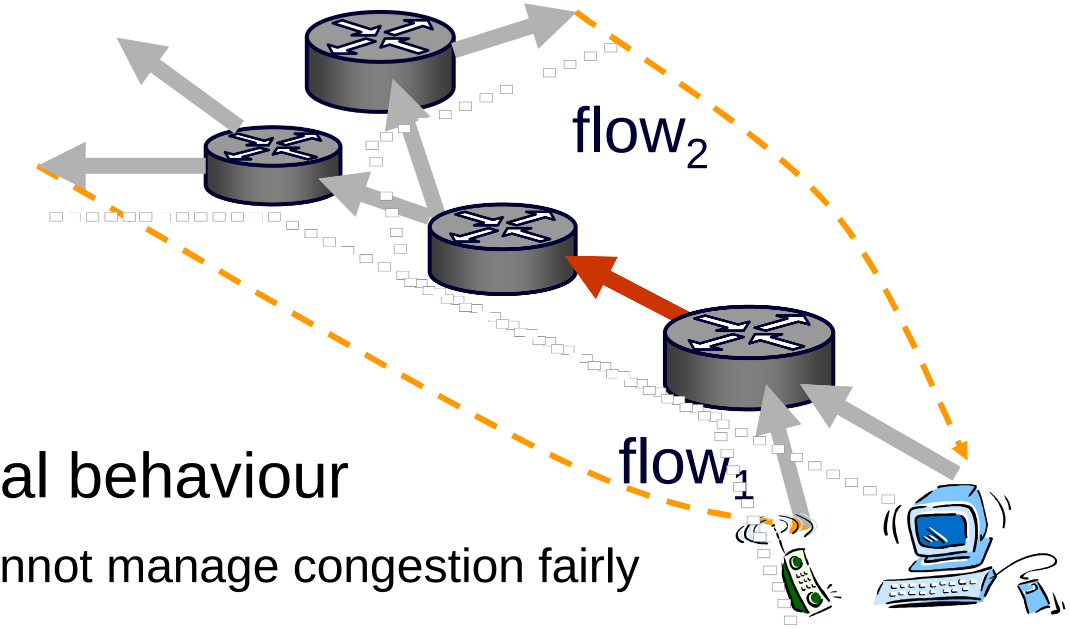
fixing Internet DDoS & net neutral QoS

using one more bit and economic policy

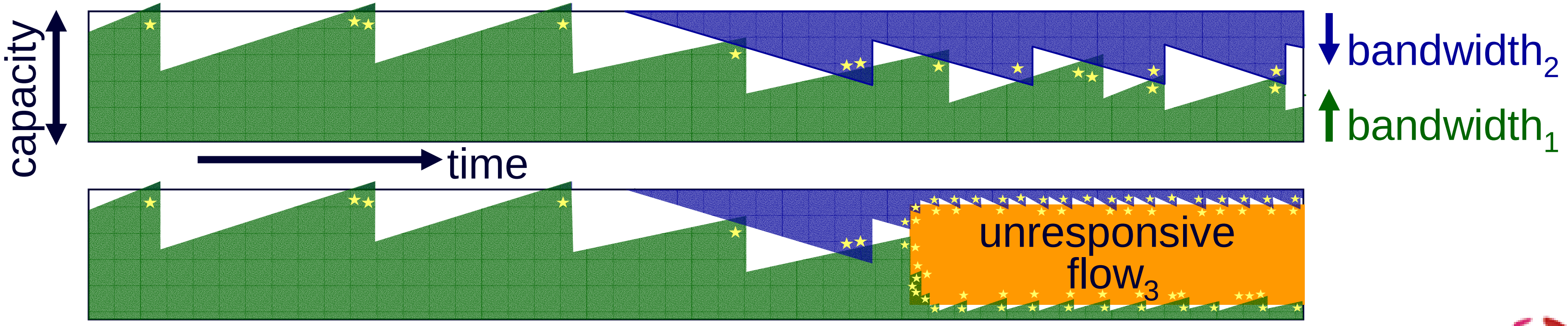
Bob Briscoe
Chief Researcher, BT Group

Nov 2006

“the big problem with the Internet”



- cannot control anti-social behaviour
  - at the network level → cannot manage congestion fairly
  - ‘cannot’ is strictly true – congestion information in wrong places
  - network reliant on voluntary politeness of all computers
  - a game of chicken – taking all and holding your ground pays



a long standing architectural vacuum

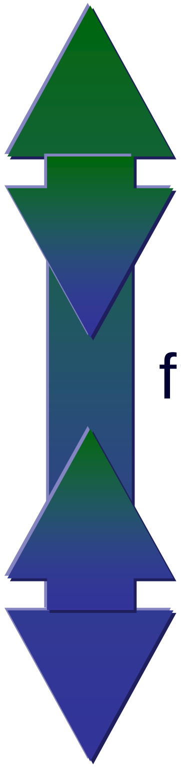
resource allocation / accountability / fairness

- on ‘to do’ list since the Internet’s early days
- isn’t enforcing ‘TCP-fairness’ the answer? No
  - anyone can create more TCP-friendly flows than anyone else
  - for much longer than anyone else (p2p file-sharing)
  - and embedding only TCP congestion control into Internet would kill evolution (VoIP)
- the community problem has been this deeply embedded dogma
  - “equal flow rates are fair” has no basis in real life, social science or philosophy
  - obscured by this idea, community can’t tell a bad fix from a good one
  - and doesn’t even realise fairness is completely out of control
- correct measure of fairness is volume of congestion (‘cost’) not flow rate
  - proof of correctness based on global utility maximisation (Kelly97 in [1])
  - answers questions like “how many flows are fair?” “for how long?”
  - rejected at the time – required congestion pricing to discourage anti-social behaviour
- this talk: users can have flat pricing *and* fairly allocate resources

[1] Briscoe “Flow rate fairness: Dismantling a religion” (Oct 2006) <http://www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#rateFairDis>

freedom vs fairness

resolving the net neutrality debate



freedom to be anti-social – demand side

- the Internet is all about the freedom to get what I want (within my line rate)
- limited by how much I impinge on the freedom of others
    - congestion

freedom within fairness

- differentiated quality of service
- you’ll get what you ask for (within the prevailing fairness policy)
- you’ll get what *we* infer you want from what you’re doing

freedom to be anti-competitive – supply side

is this important?

- working with packets depersonalises it
  - it’s about conflicts between real people
  - it’s about conflicts between real businesses
- 1st order fairness – average over time
  - 24x7 file-sharing vs interactive usage
- 2nd order fairness – instantaneous shares
  - unresponsive video streaming vs TCP
  - fair burden of preventing congestion collapse
- not some theoretical debate about tiny differences
  - huge differences in congestion caused by users on same contract
  - hugely different from the shares government or market would allocate
  - yes, there’s a lot of slack capacity, but not that much and not for ever
- allocations badly off what a market would allocate
  - eventually lead to serious underinvestment in capacity
- ‘do nothing’ will not keep the Internet pure
  - without an architectural solution, we get more and more middlebox kludges

designed for tussle

- current Internet gives freedom but no fairness
  - the more you take, the more you get; the more polite you are, the less you get
  - but we don’t want to lose freedom by enforcing fairness

solution: allow ISPs to enforce user-specific congestion control fairness

- liberal acceptable use policies
  - open access, no restrictions
- middle ground
  - might want to cap congestion caused per user (e.g. 24x7 heavy p2p sources, DDoS)
  - evolution of different congestion control (e.g. hi-dynamics; rate adaptive VoIP, video)
- conservative acceptable use policies

11. religious
10. political
9. legal
8. commercial
7. application
4. transport
3. network
2. link
1. physical

exec summary

- will range widely across religion, economics, architecture & bits
- freedom vs. fairness
- solution
  - congestion re-feedback engineered for IP (re-ECN)
- expected effect – a step to trigger evolutionary change
  - on Internet applications – aggressive behaviour proportionately throttled
  - on network interconnection market – usage charging based on congestion
  - on distributed denial of service attacks – natural extreme throttling
- strong deployment incentives

- unless there’s interest, I won’t cover:
  - protocol & algorithm detail
  - potential routing benefits
  - microeconomics of welfare maximisation
  - how to do fairness between fairnesses within sub-groups
    - NATO, commercial ISPs, universities, countries with social objectives

solution: congestion re-feedback (re-inserted feedback)

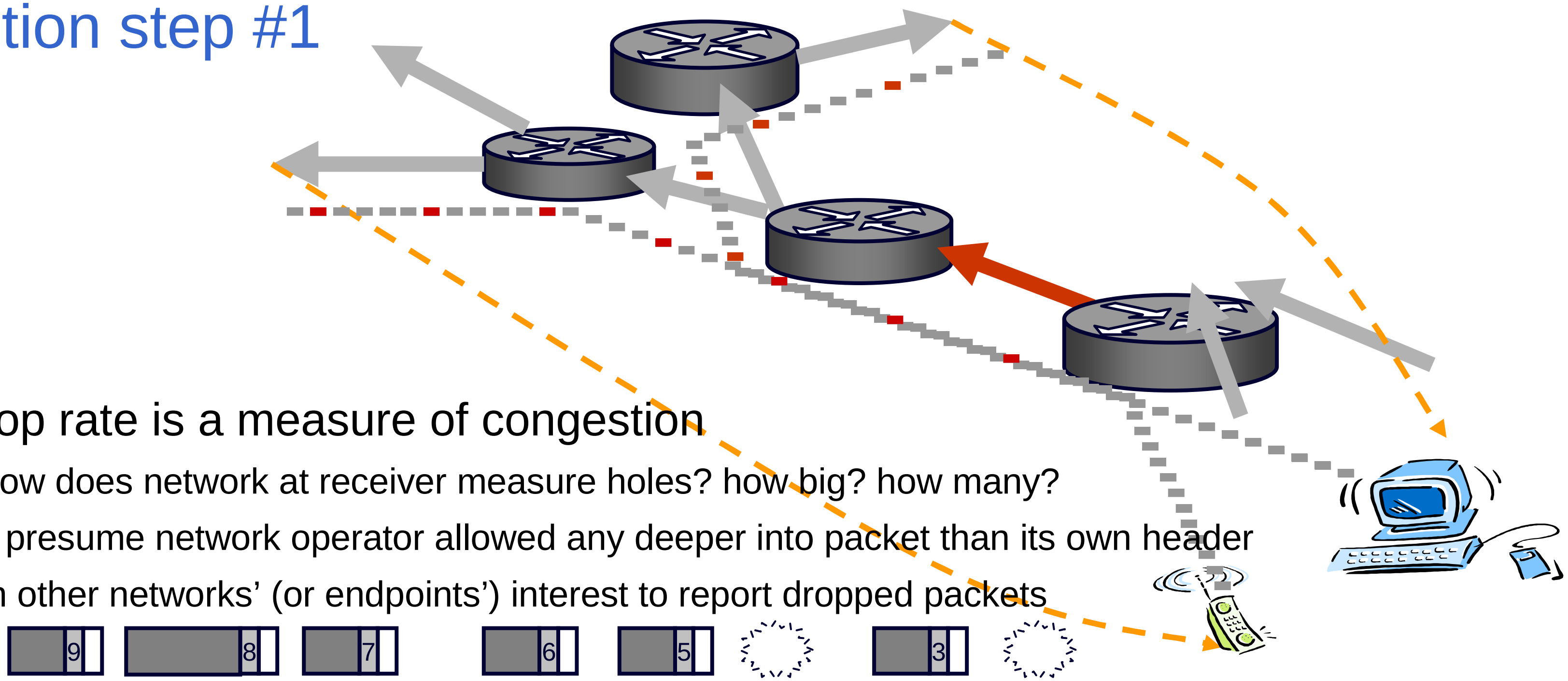
status

- culmination of over a decade of research (mainly Cam, BT, M$, UCL +)
  - addition of information missing from packet - essential to network economics
  - even if our specific protocol (re-ECN) has flaws, it will be worth finding another
- progressing through IETF – long haul – requires change to IP
  - fully spec'd protocol - last week: 4th presentation since Sep 05
  - also great progress dismantling the prevailing fairness religion (IETF and wider)
- intellectual property rights
  - originally recognised by BT as key patent
  - agreed to freely license aspects essential to IETF standardisation
- working to get on roadmaps for
  - NGN interconnection; IETF pre-congestion notification (PCN) w-g; 3GPP
- support / interest
  - BT's wholesale & retail divisions & CTO, big 5 network operators (senior level)
  - broadband, interconnection & net neutrality w-gs of MIT comms futures programme (FT, BT, DT/T-Mobile, Cisco, Comcast, Intel, Motorola, Nokia, Nortel, MIT, Cam, +)

a change to IP needs to be 'owned' by Internet community
please take it, break it, analyse it, re-design it

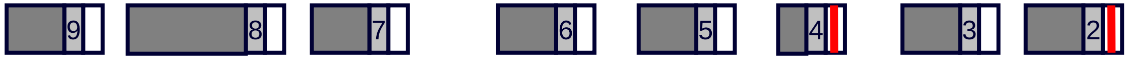
measurable incipient congestion

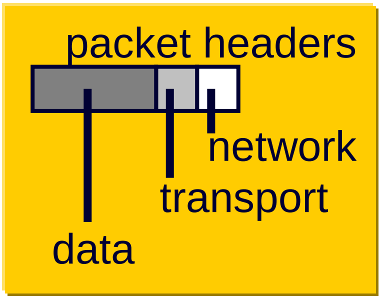
solution step #1



- 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

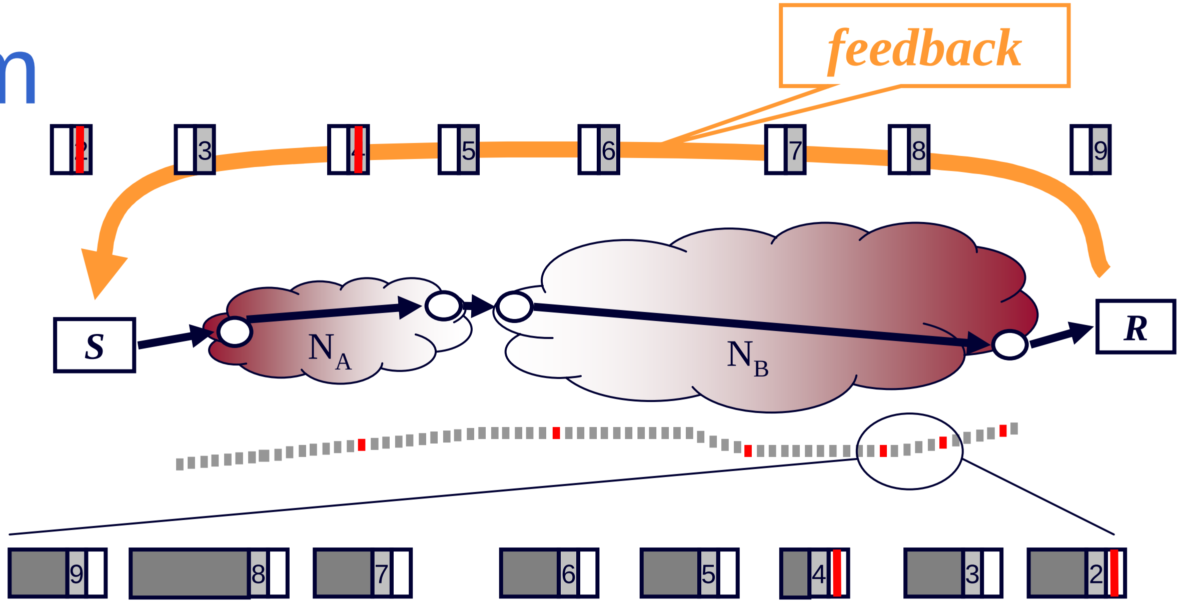
- solution: Explicit Congestion Notification (ECN)
  - mark packets as congestion ***approaches*** - to avoid drop
  - already standardised into IP (2001)
  - implemented by all router vendors – very lightweight mechanism
  - but rarely turned on by operators (yet) – mexican stand-off with OS vendors

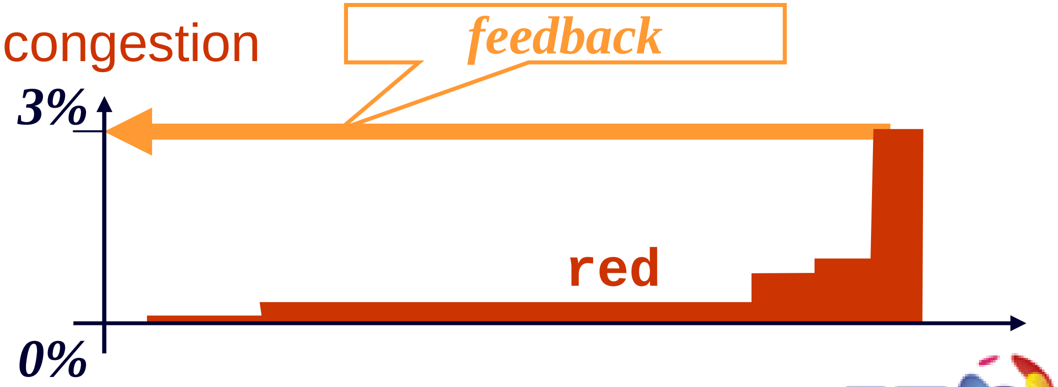




new problem

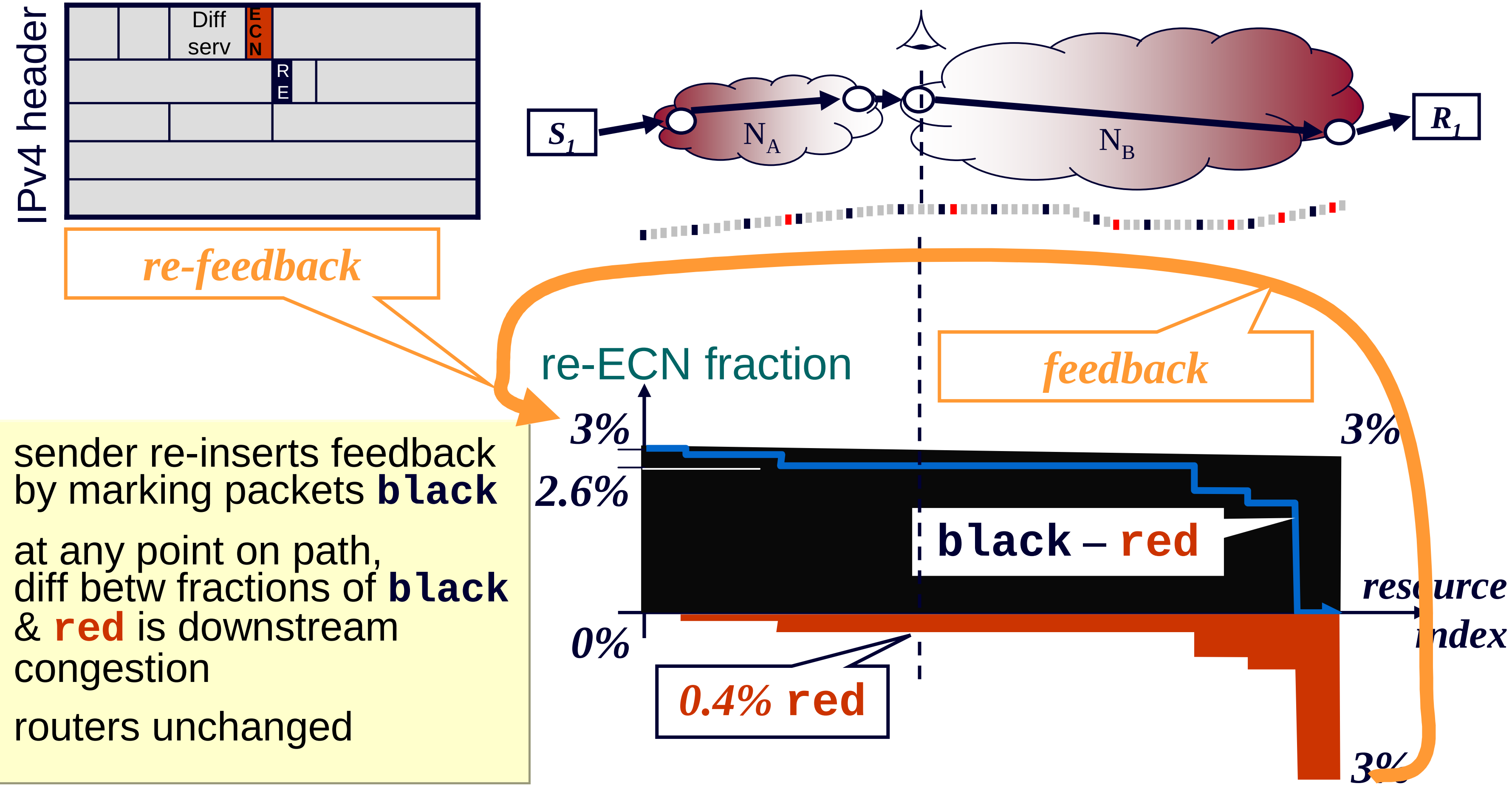
- congestion only measurable at exit
- can’t measure congestion at entry
  - can’t presume allowed deeper into feedback packets





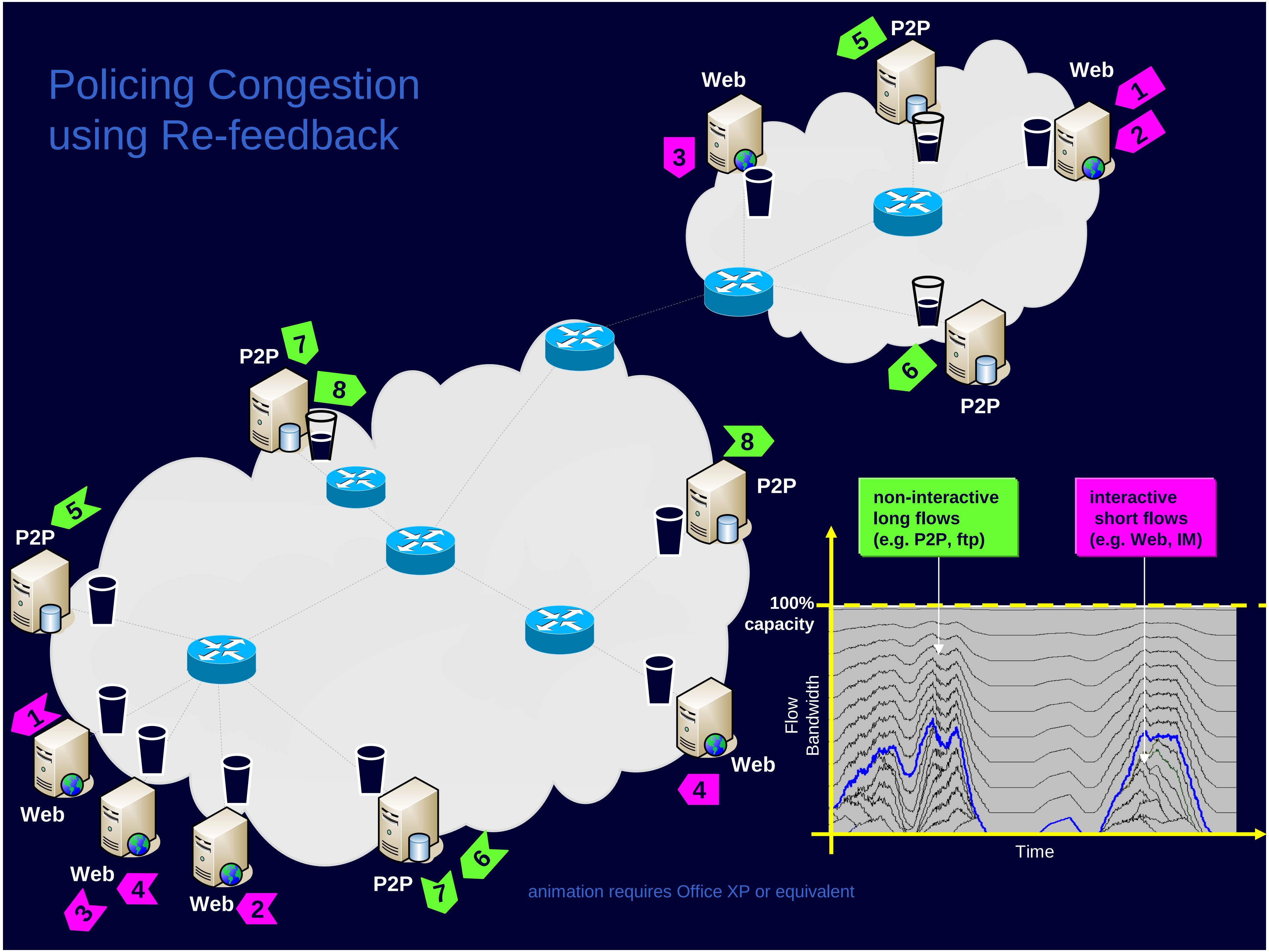
measurable downstream congestion

solution step #2



- sender re-inserts feedback by marking packets **black**
- at any point on path, diff betw fractions of **black** & **red** is downstream congestion
- routers unchanged

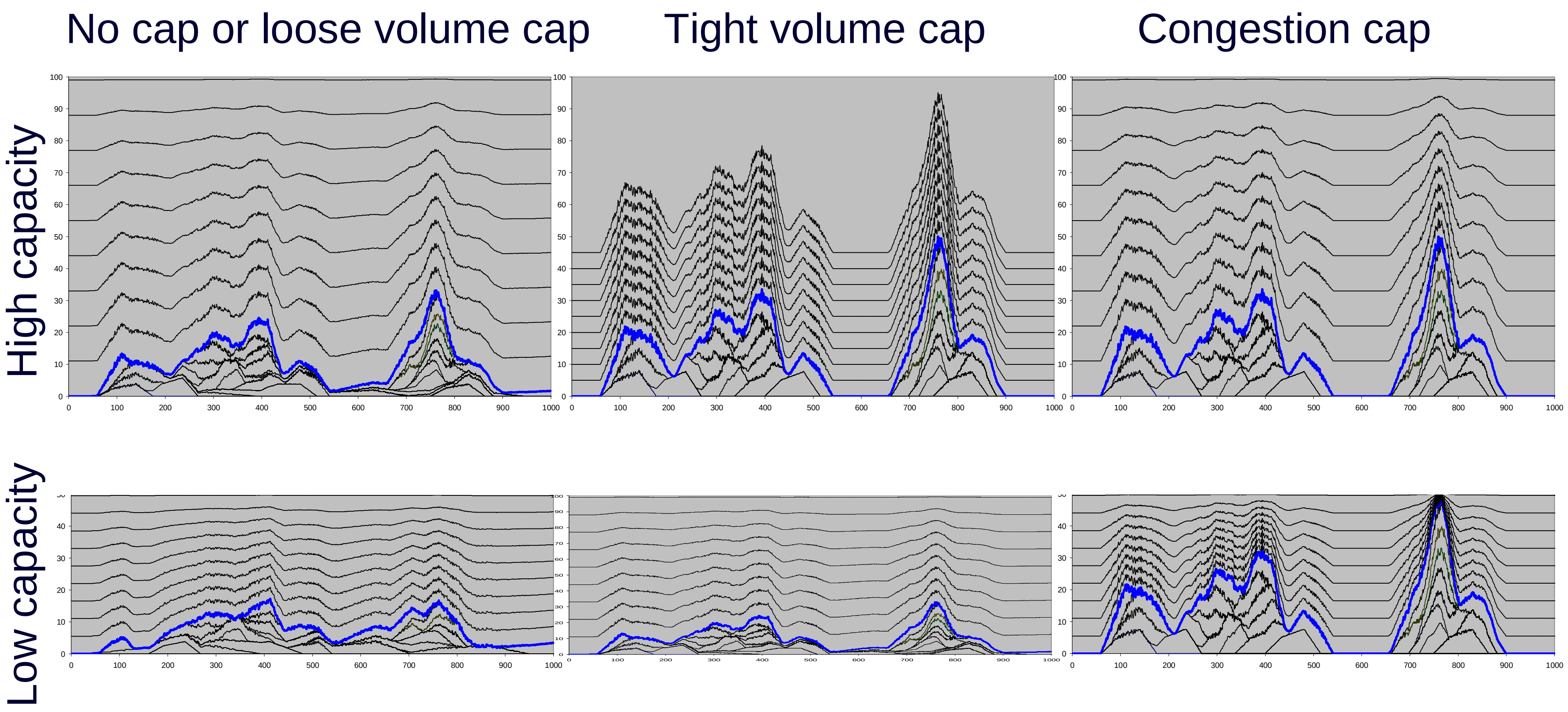
Policing Congestion using Re-feedback



animation requires Office XP or equivalent

congestion cap auto-adjusts

volume cap always a hard compromise



No cap or loose volume cap

Tight volume cap

Congestion cap

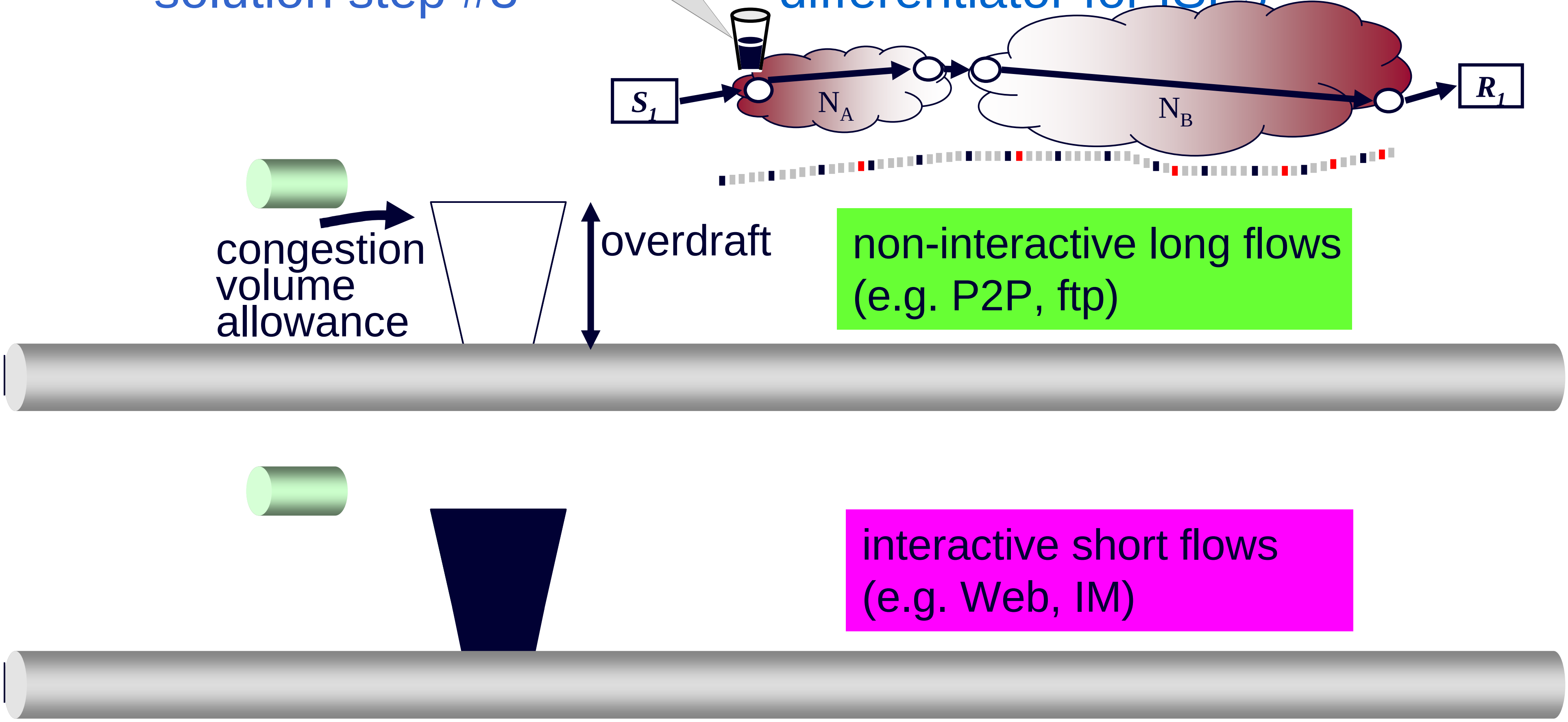
High capacity

Low capacity

congestion policer – one example: per-user policer

solution step #3

differentiator for ISPs



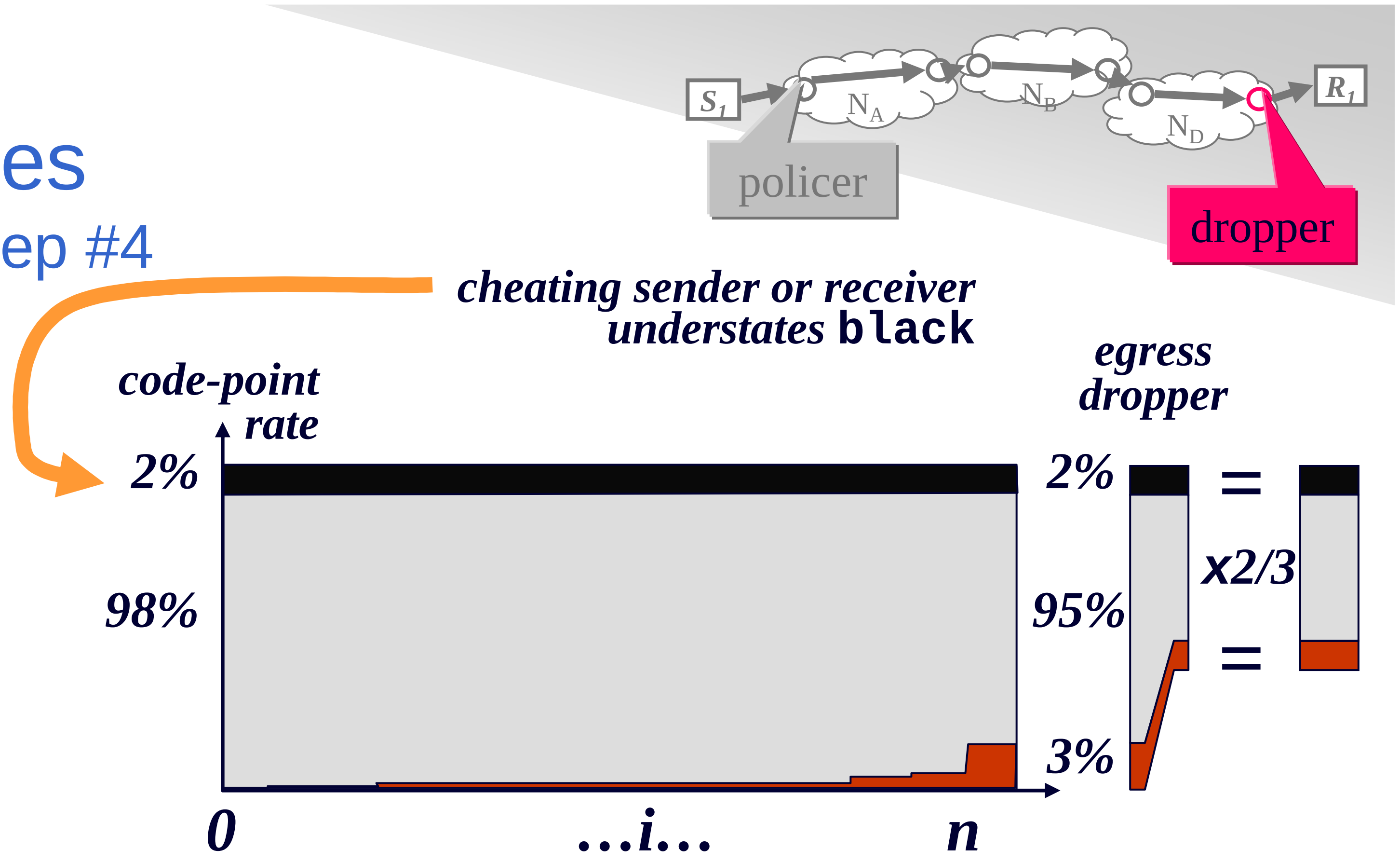
non-interactive long flows (e.g. P2P, ftp)

interactive short flows (e.g. Web, IM)

two different customers, same deal

incentives

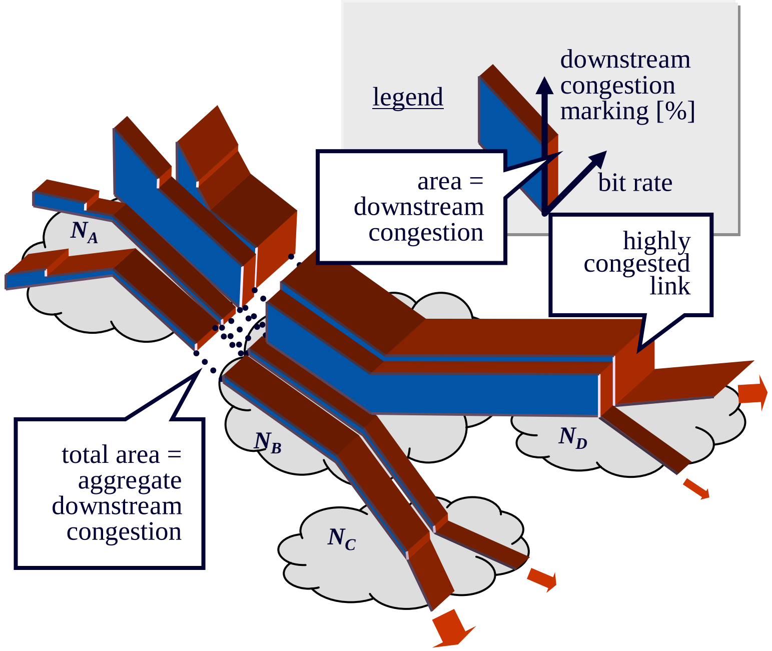
solution step #4

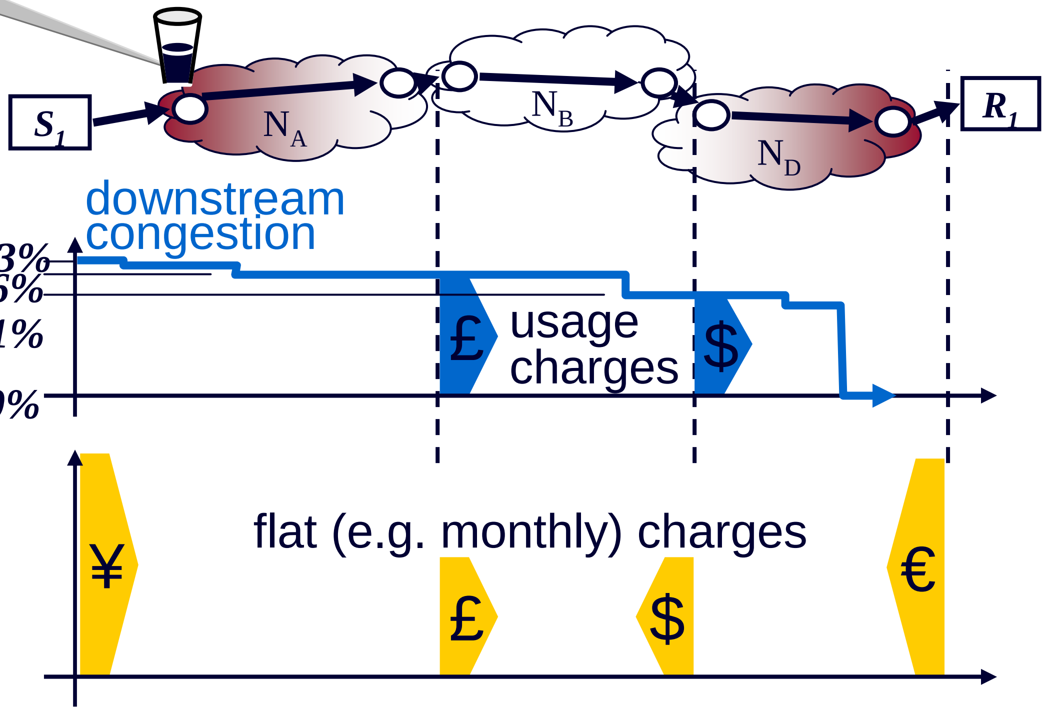


- won’t sender or receiver simply understate congestion?
- no – drop enough traffic to make fraction of **red** = **black**
- goodput best if rcvr & sender honest about feedback & re-feedback

inter-domain accountability for congestion

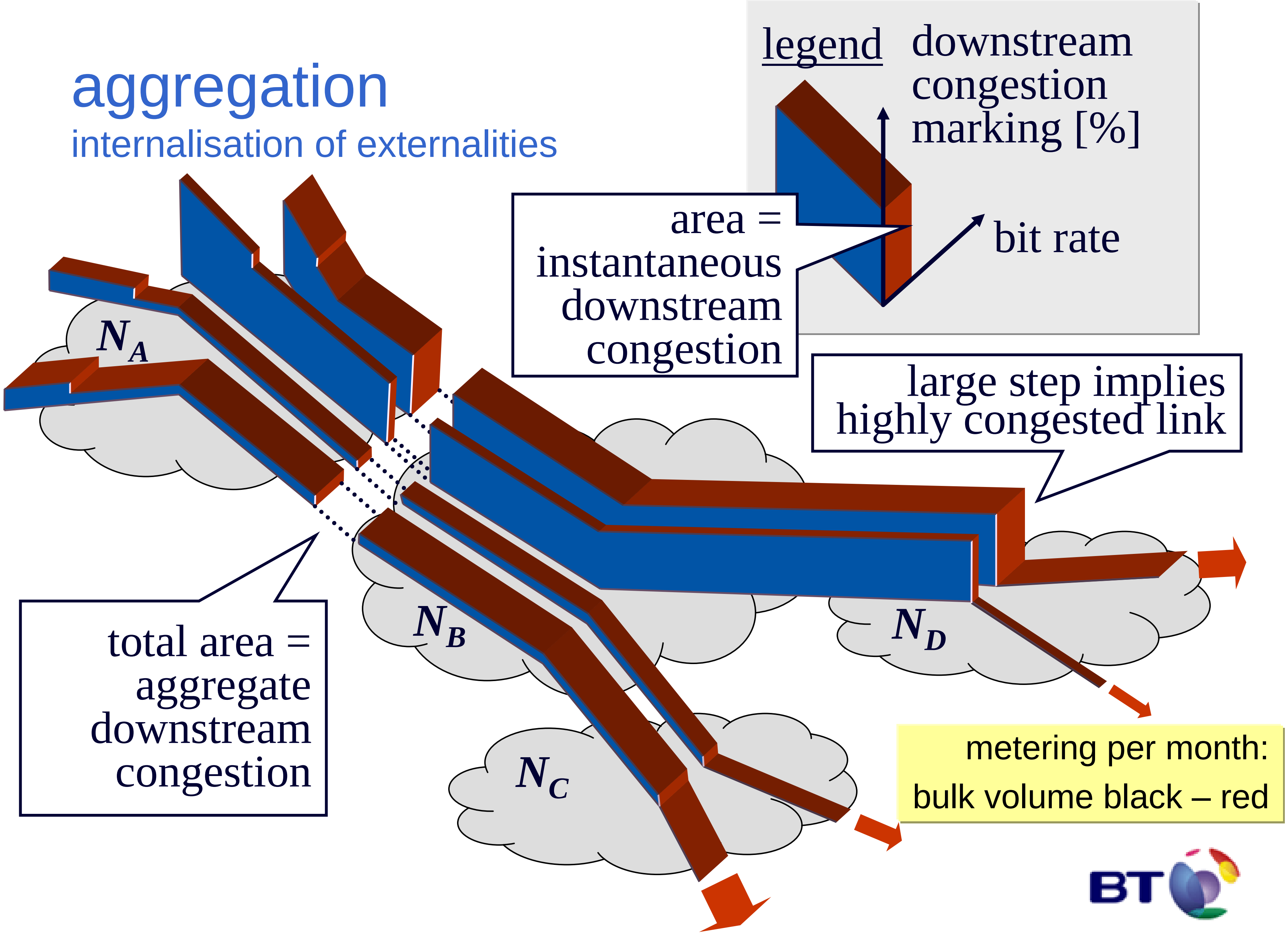
- metric for inter-domain SLAs or usage charges
  - N_B applies penalty to N_A in proportion to bulk volume of **black** less bulk volume of **red** over, say, a month
  - could be tiered penalties, directly proportionate usage charge, etc.
  - flows de-aggregate precisely to responsible networks
  - N_A deploys policer to prevent S_1 causing more cost than revenue


legend
downstream congestion marking [%]
bit rate
area = downstream congestion
highly congested link
total area = aggregate downstream congestion
N_A
N_B
N_C
N_D



aggregation

internalisation of externalities



area = instantaneous downstream congestion

large step implies highly congested link

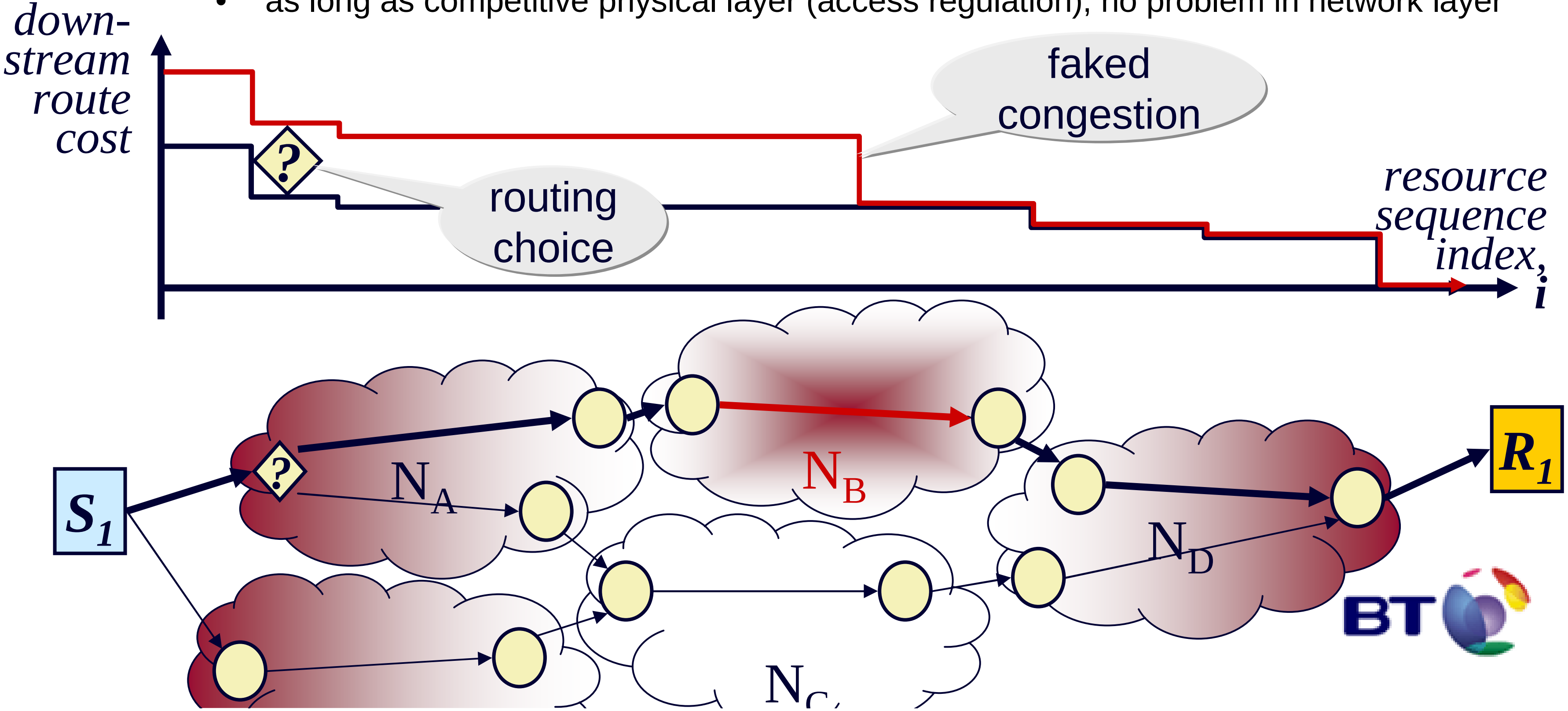
total area = aggregate downstream congestion

metering per month:

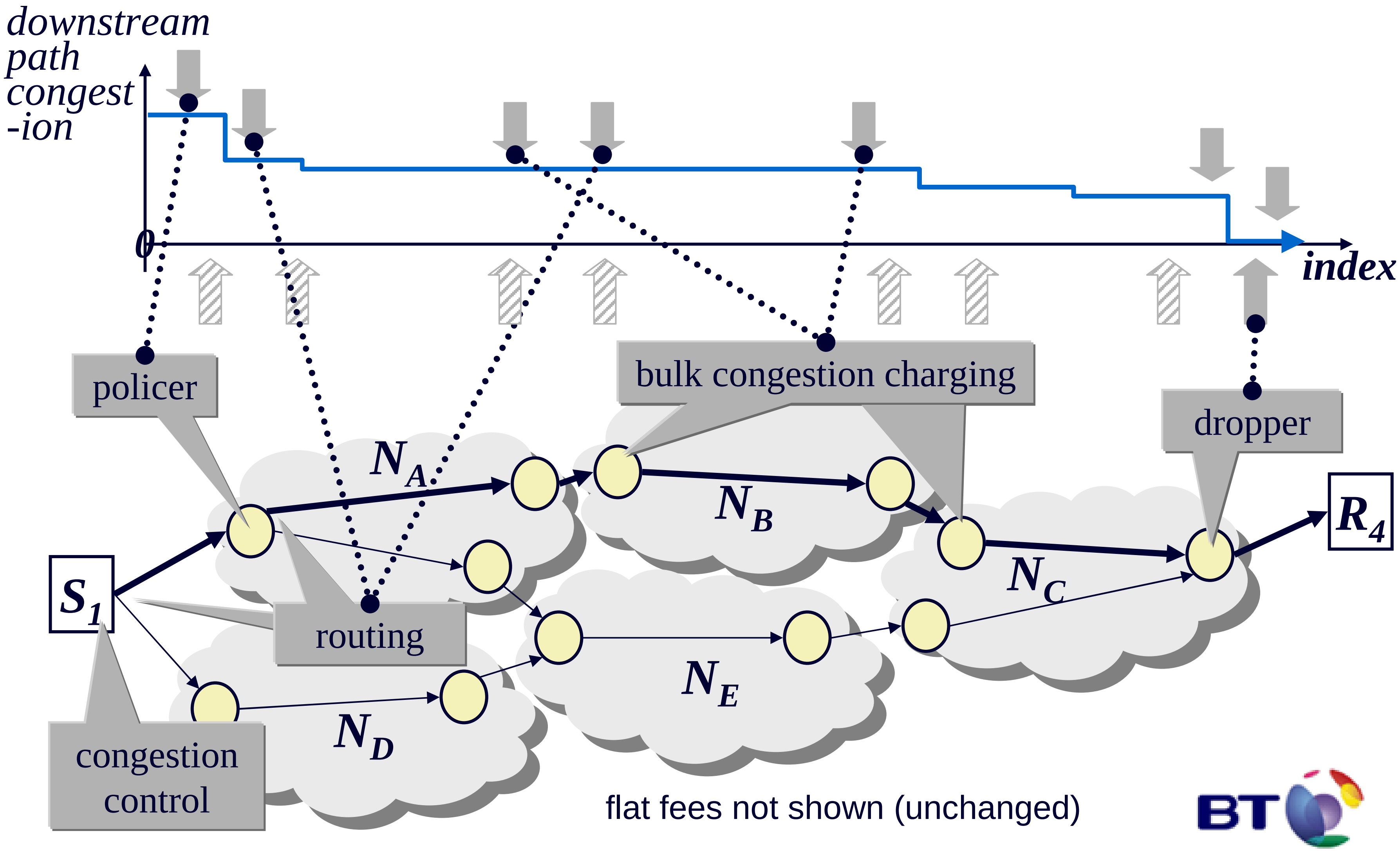
bulk volume black – red

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



incentive framework



flat fees not shown (unchanged)

grounded in economic theory

not just arbitrary bit twiddling

demand side

- applying a price to congestion causes users to maximise Internet-wide utility [Kelly97]
  - reasonable assumptions: concave utility; competitive market with price taking users
- but without re-feedback, had to congestion charge and had to charge receiver
- with re-feedback can keep traditional flat fee
  - use engineered mechanism (policer) not pricing
    - limit the cost of congestion the *sender* can cause to the flat fee she paid
- accountability without usage charging

supply side

- incipient congestion stats drive provisioning
  - congestion marking represents real (paid for) demand
  - volume of congestion marking at each resource proportional to investment that resource needs
- network knowledge of downstream congestion hugely simplifies control & mgmt

fixes market failures

- balances information asymmetry between endpoints and network
- congestion externality internalised by those that cause congestion
  - *and those that allow it to be caused*

differential quality of service (QoS) control

without all the complicated stuff

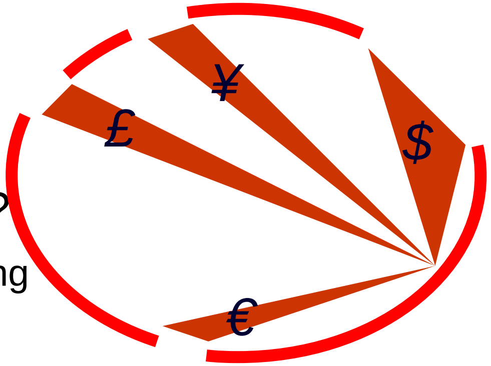
- QoS only relevant when there’s a risk of congestion
- enforcing congestion control is equivalent to QoS
  - allowing one app’s rate to slow down less than others in response to incipient congestion (ie. still low delay)
  - is equivalent to giving scheduling priority on routers*
- even if user pays a flat monthly fee
  - better QoS for some apps leaves less congestion ‘quota’ for rest
- making users accountable for not slowing down as much as others during congestion
  - is a sufficient mechanism both for QoS and for ‘paying’ for QoS
- incredible simplification of mechanisms for QoS control & mgmt
  - and, unlike other QoS mechanisms
  - it also prevents users ‘stealing’ QoS at everyone else’s expense

* except within a round trip time – implies two priority classes would be sufficient

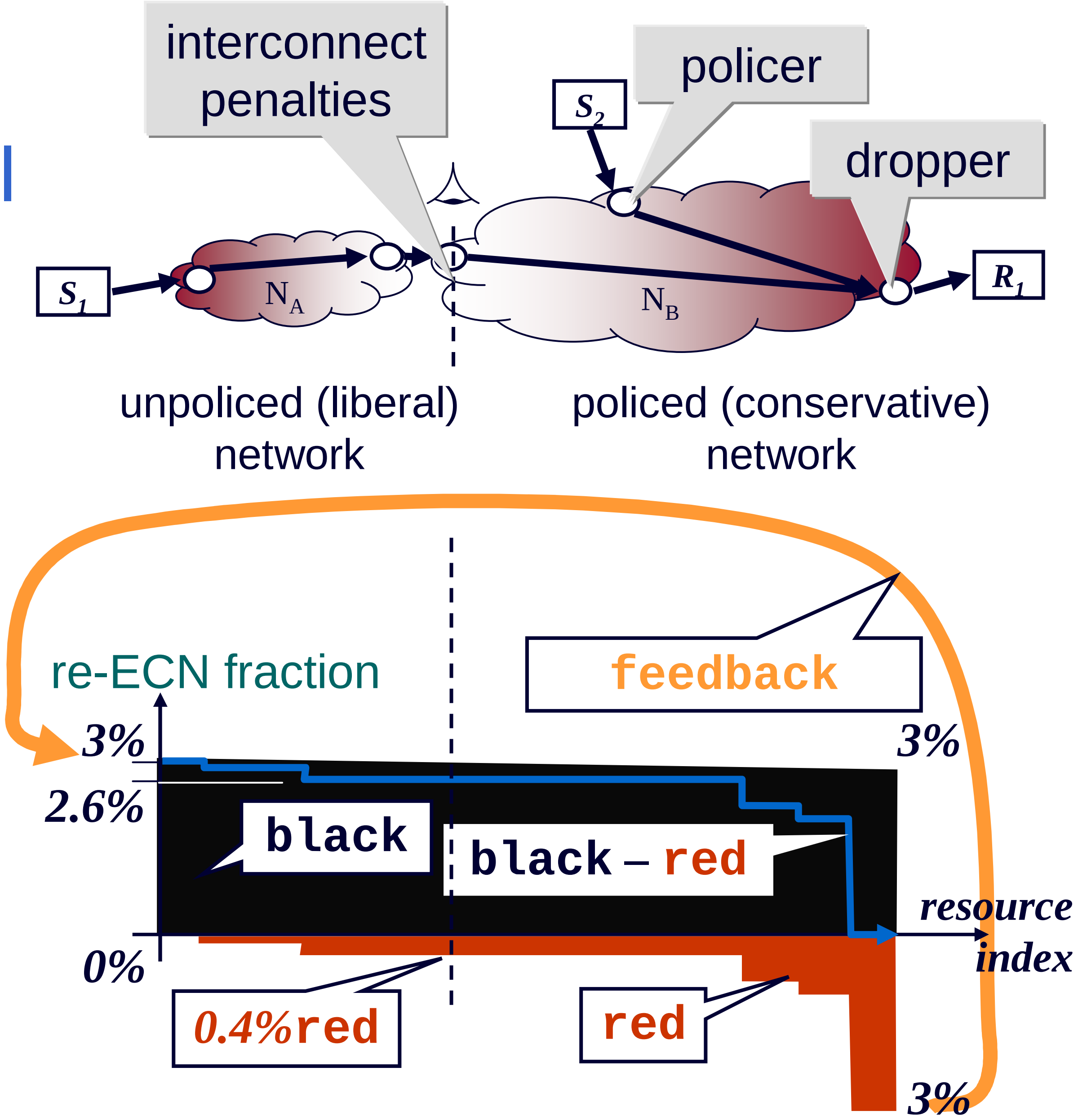
(can also determine relative congestion marking rates of each class using economics)

deployment incentives
bootstrap then chain reaction

- deployment effectively involves architectural change
  1. (minor) change to sender's Internet stack
  2. network deploys edge/border incentive functions
  - breaking the stand-off between 1 & 2 requires strong incentives
- re-feedback solves ISPs' main cost control problem
    - – third party services competing with ISP pay below network cost
    - – ISP has to compete *while* paying balance of competitor's costs
  - hits big fear button and big greed button
  - but keeps moral high ground
    - – net neutral: managing congestion not app discrimination
- first movers: vertically integrated cellular operators?
  - 3GPP devices leak deployment to other networks by roaming
- 2nd movers (NGNs?) continue chain reaction
  - adopters' incoming border charges focus on non-adopters



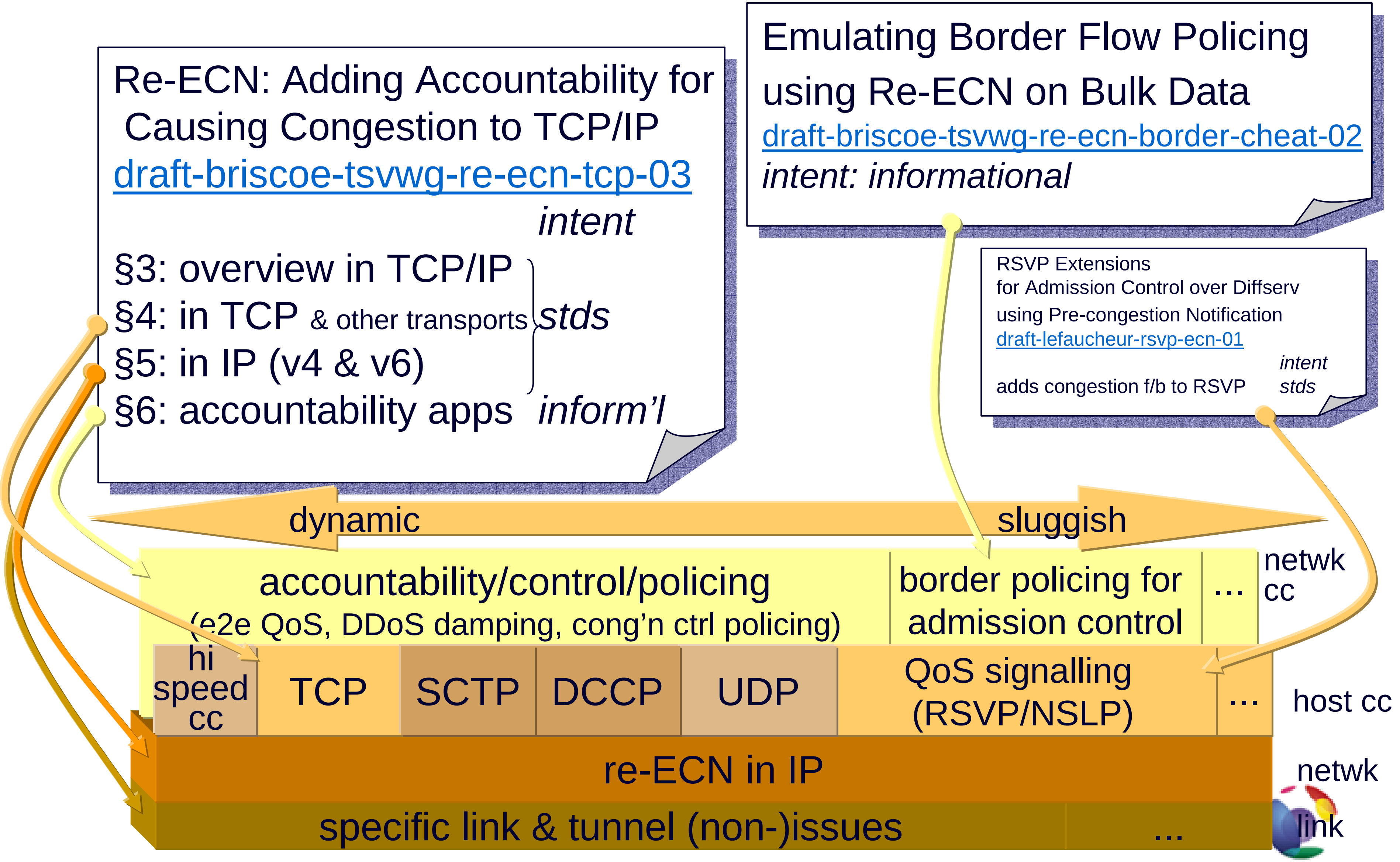
re-ECN partial deployment



other steps to deploy re-feedback

- customer contracts
  - include congestion limit
- 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 internet draft roadmap



Re-ECN: Adding Accountability for Causing Congestion to TCP/IP
draft-briscoe-tsvwg-re-ecn-tcp-03

	intent
§3: overview in TCP/IP	*stds*
§4: in TCP & other transports	
§5: in IP (v4 & v6)	
§6: accountability apps	*inform'l*

Emulating Border Flow Policing using Re-ECN on Bulk Data
draft-briscoe-tsvwg-re-ecn-border-cheat-02
intent: informational

RSVP Extensions
for Admission Control over Diffserv
using Pre-congestion Notification
draft-lefaucheur-rsvp-ecn-01
adds congestion f/b to RSVP *intent* *stds*

dynamic ⟷ sluggish

accountability/control/policing (e2e QoS, DDoS damping, cong'n ctrl policing)					border policing for admission control	...	netwk cc
hi speed cc	TCP	SCTP	DCCP	UDP	QoS signalling (RSVP/NSLP)	...	host cc
re-ECN in IP							netwk
specific link & tunnel (non-)issues						...	link

extended ECN codepoints: summary

- extra semantics backward compatible with previous ECN codepoint semantics

ECN code-point	ECN [RFC3168] codepoint	RE flag	Extended ECN codepoint	re-ECN meaning	`worth’
00	not-ECT	0	Not-RECT	Not re-ECN capable transport	
		1	FNE	Feedback not established	**+1**
01	ECT(1)	0	Re-Echo	Re-echo congestion event	**+1**
		1	RECT	Re-ECN capable transport	**0**
10	ECT(0)	0	---	‘Legacy’ ECN use	
		1	--CU--	Currently unused	
11	CE	0	CE(0)	Congestion experienced with Re-Echo	**0**
		1	CE(-1)	Congestion experienced	**-1**

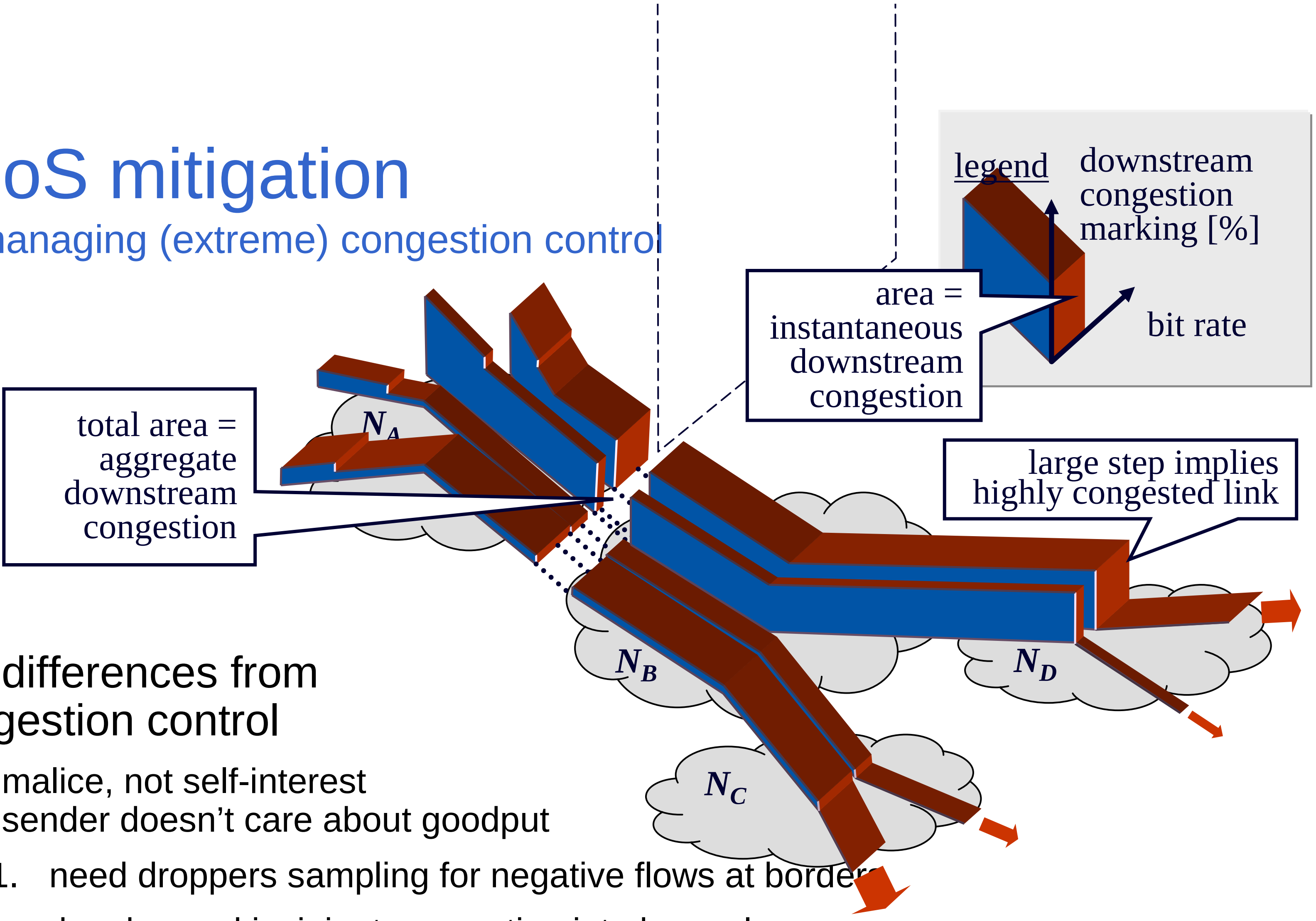
flow bootstrap

- **`green`** packet(s) at start of flow
  - ‘worth’ **`+1`** same as **`black`**
  - credit for safety due to lack of feedback
  - a deposit
- after idle >1sec
  next packet MUST be **`green`**
  - enables deterministic flow state mgmt (policers, droppers, firewalls, servers)

- **`green`** also serves as state setup bit [Clark, Handley & Greenhalgh]
  - protocol-independent identification of flow state set-up
  - for servers, firewalls, tag switching, etc
  - don’t create state if not set
  - may drop packet if not set but matching state not found
  - firewalls can permit protocol evolution without knowing semantics
  - some validation of encrypted traffic, independent of transport
  - can limit outgoing rate of state setup
- to be precise **`green`** is ‘idempotent soft-state set-up codepoint’

DDoS mitigation

just managing (extreme) congestion control



total area = aggregate downstream congestion

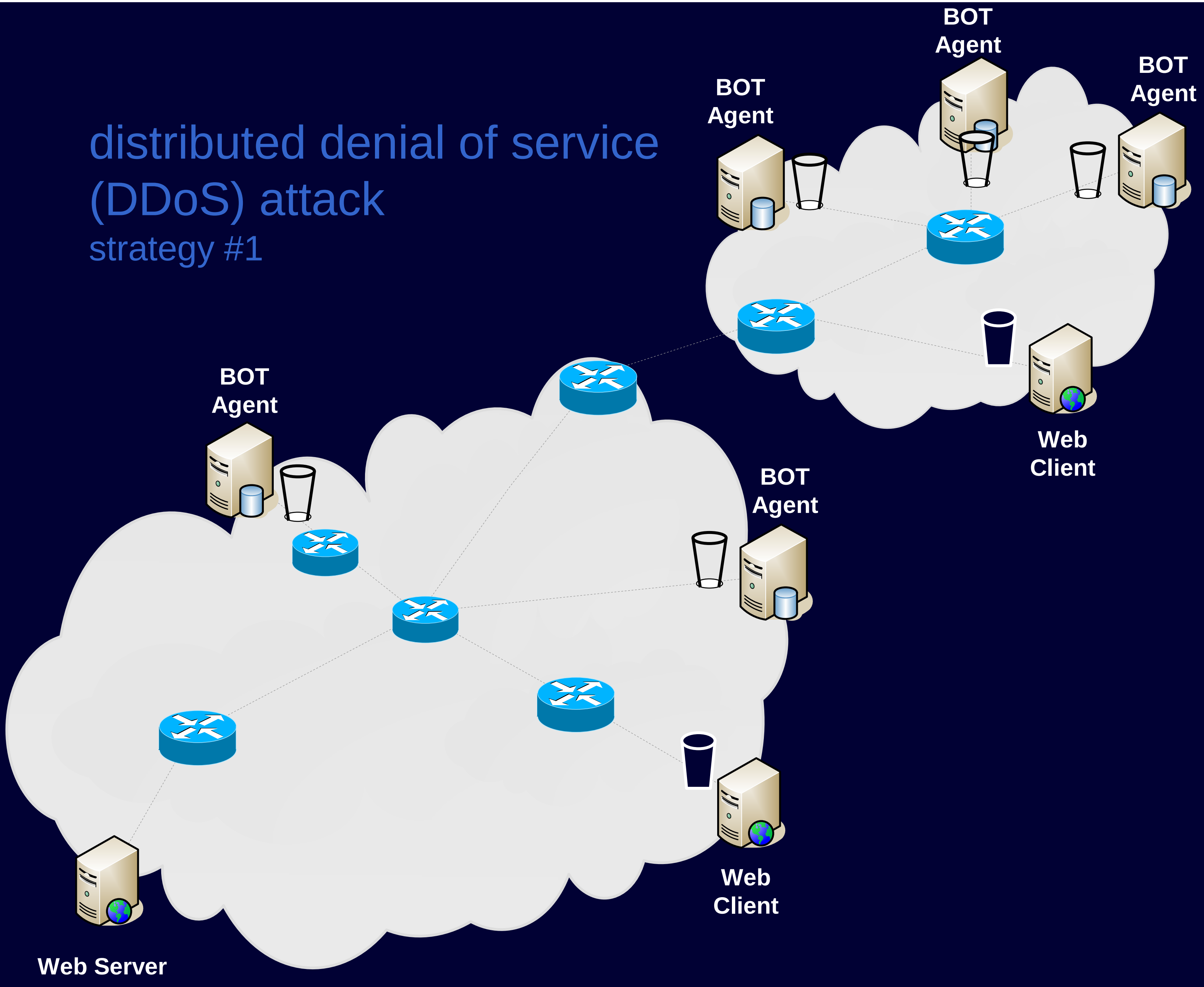
area = instantaneous downstream congestion

large step implies highly congested link

- two differences from congestion control
  - malice, not self-interest
    sender doesn't care about goodput
    1. need droppers sampling for negative flows at borders
  - pushes beyond incipient congestion into heavy loss
    2. need preferential drop on routers
- provides incentives to deploy complementary DDoS solutions

distributed denial of service (DDoS) attack

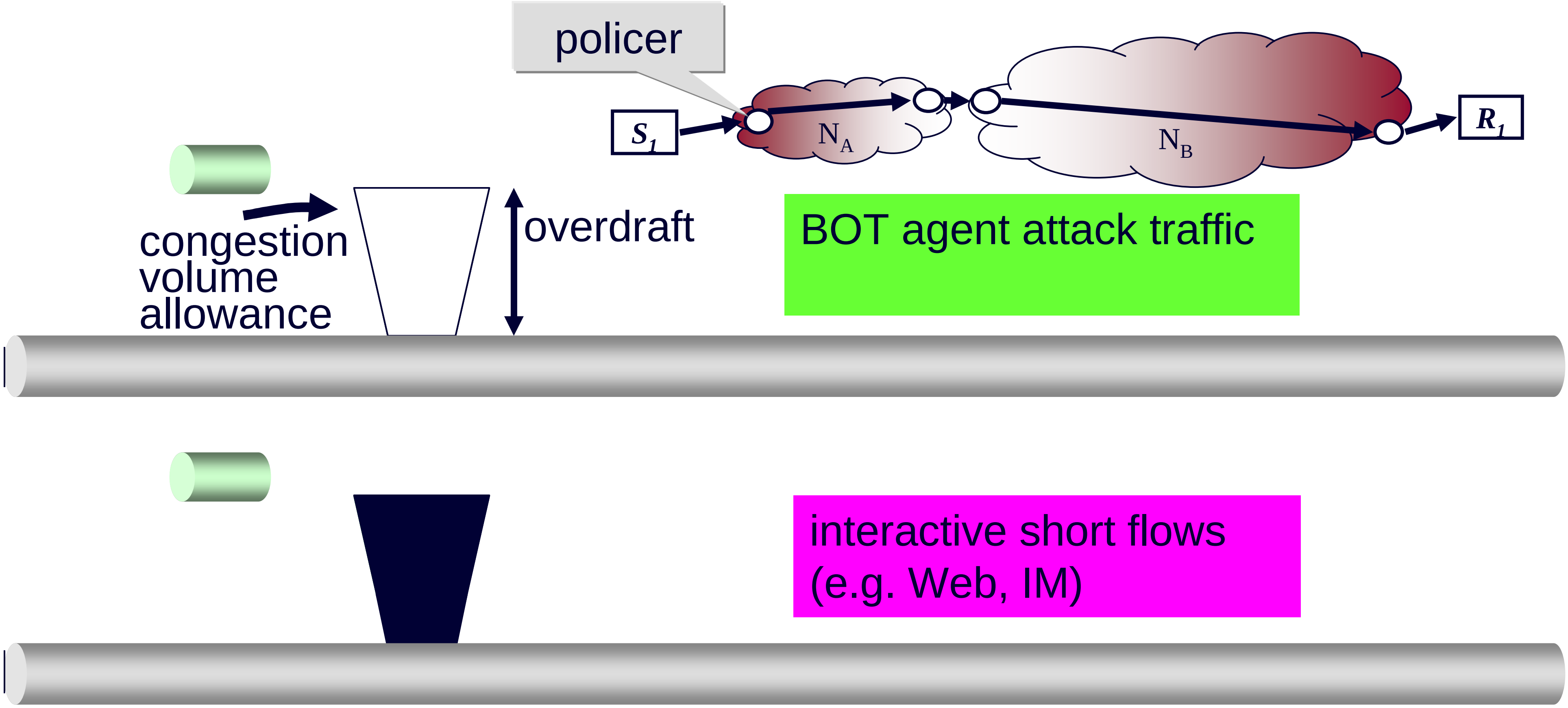
strategy #1



animation requires Office XP or equivalent

per-user congestion policer

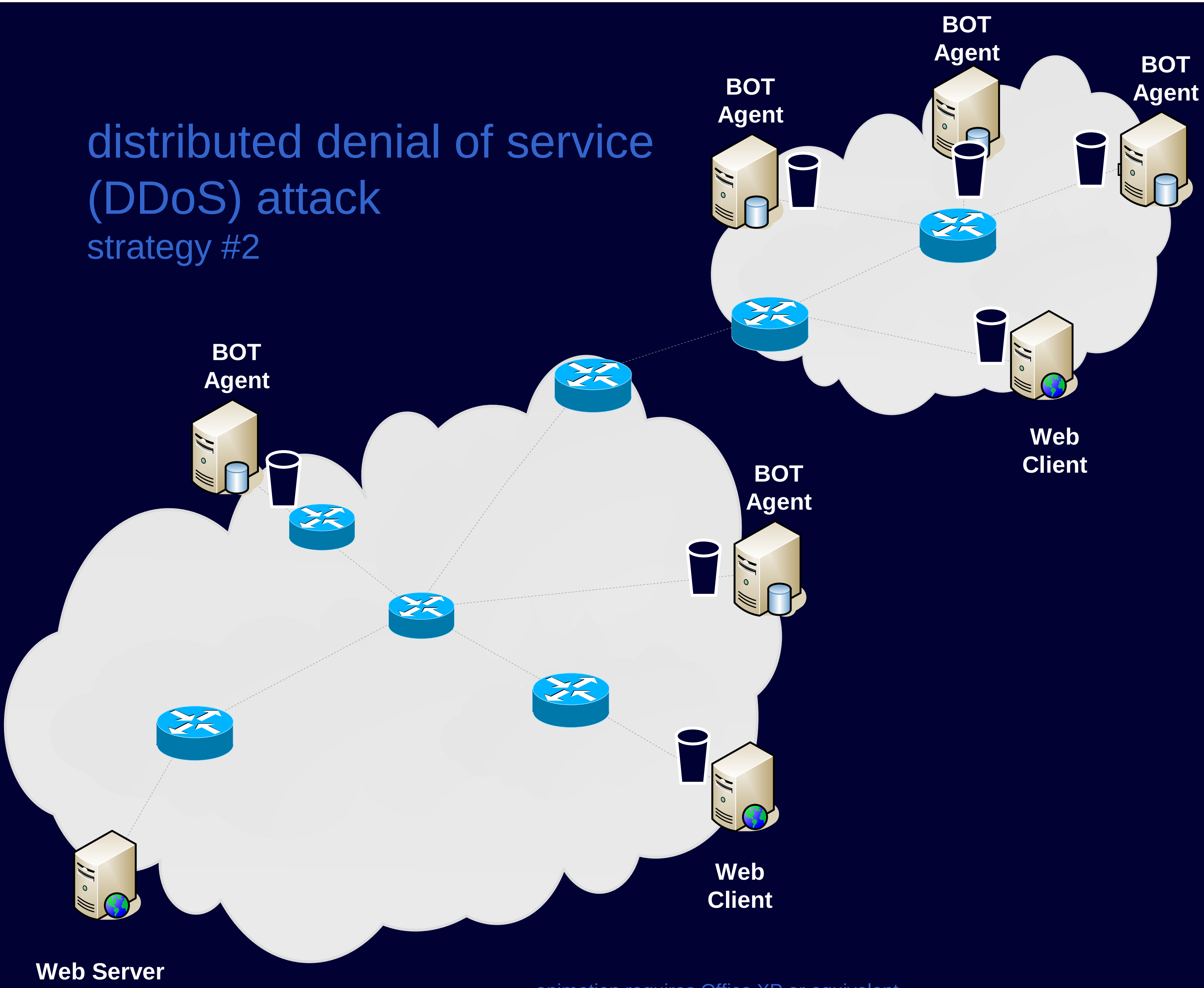
DDoS attack strategy #1


policer
S_1
N_A
N_B
R_1
congestion volume allowance
overdraft
BOT agent attack traffic
interactive short flows (e.g. Web, IM)

animation requires Office XP or equivalent

distributed denial of service (DDoS) attack

strategy #2



animation requires Office XP or equivalent

outstanding issues

- technical
  - ✗ a lot more verification of all the claims to do
  - ✗ community found a few nasty vulnerabilities over last year
    - ✓ fixed (added minor complexity in only one case)
  - ✗ connection spoofing attack still outstanding
    - ✓ possible solution recently brainstormed
- religious
  - ✗ underlying problem has been dogma that equal flow rates are fair
    - ✓ groundswell change in community thinking since mid Oct'06
    - ✗ dismantling a religion not so easy – people fall into their old ways
- community
  - ✗ a lot of passive support, but consensus needs a lot more active interest

conclusions

- resolution of tensions in net neutrality debate
  - freedom to use the Internet, until you congest freedom of others
  - proportionate restriction of freedom during congestion
- an architectural change with grand implications
  - simple management and control of QoS
  - naturally mitigates DDoS
  - generates correct capacity investment incentives and signals
- but conceptually simple and trivial to implement
- strong deployment incentives
  - bootstrap and onward chain reaction
- where’s the catch?
  - invite you to analyse it, break it, re-design it

Q&A
and more info...

- Fixing the broken mindset (polemical)
  - Flow Rate Fairness: Dismantling a Religion IETF Internet draft (Oct 2006)
- Overall intention
  - Policing Congestion Response in an Inter-Network Using Re-Feedback (SIGCOMM'05 – mechanism outdated)
- Mechanisms and rationale
  - Re-ECN: Adding Accountability for Causing Congestion to TCP/IP IETF Internet Draft (Oct 2006)
- Effect on DDoS
  - Using Self-interest to Prevent Malice; Fixing the Denial of Service Flaw of the Internet Workshop on the Economics of Securing the Information Infrastructure (Oct 2006)
- more papers referenced in the above
- Bob Briscoe <http://www.cs.ucl.ac.uk/staff/B.Briscoe/>