Using Self-interest to Prevent Malice

Fixing the Denial of Service Flaw of the Internet

Bob Briscoe
Chief Researcher, BT Group

Oct 2006

Credits: Martin Koyabe, Carla Di Cariano-Gilfedder, Arnaud Jacquet

context & problem

defend against what attackers *could* do
not what they do

need to win the last battle
not just the next one

- infrastructure must serve v large population
  - even during genuine flash crowds of demand
- most cost-effective attack: flood requests during flash crowd
  - when most people need/value a service most
  - when least effort needed to tip it over the edge
- assume virus-prone end systems won't go away
  - cell phones, TVs, MP3 players, game boxes, domestic control systems
- attackers can amass 100,000s into zombie botnets
  - can and do saturate even the biggest links in the Internet at will

---

- other approaches all try to detect attack traffic
  - then block future attempts from same source address
  - they need to stop attackers faking different source addresses for each packet
  - still problem with floods of single packets
  - with this mindset, researchers have defined success as
    - forcing an attacker to imitate a flash crowd

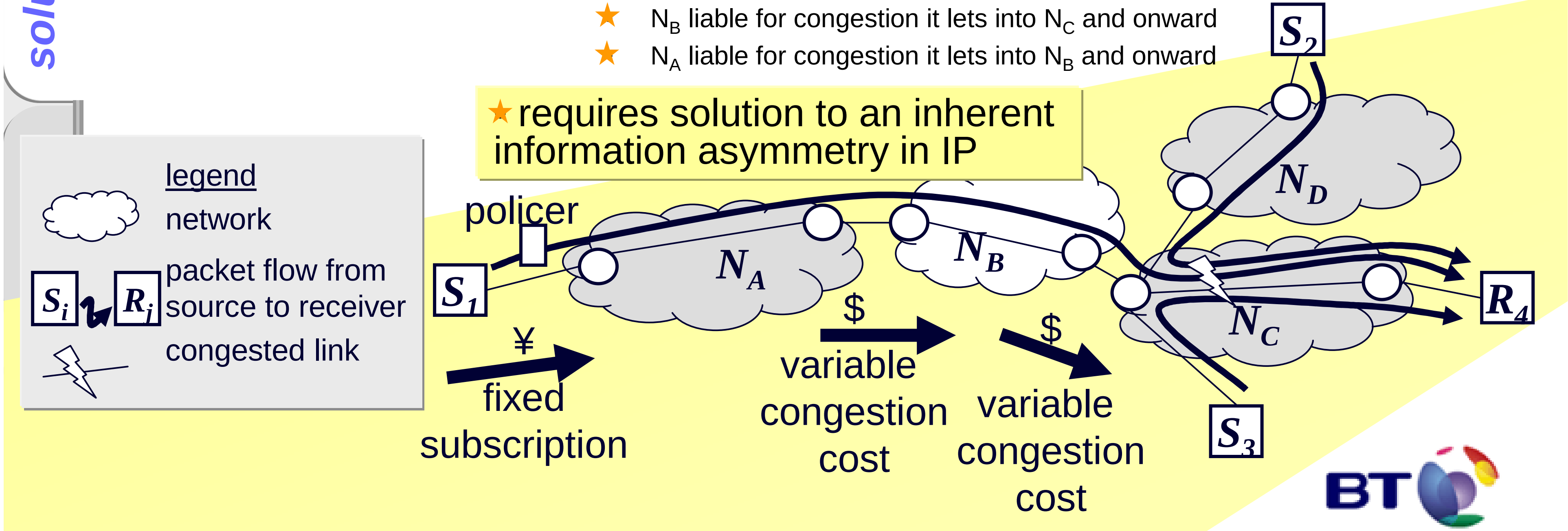
status

- one result of 3yrs research to fix Internet architecture
  - prime directive: don’t unduly restrict Internet’s ability to foster surprises
  - fixed Internet resource sharing – DDoS fix a pleasant consequence
- plan to do whatever it takes to standardise into IP
  - 2005 full standards specs drafted
    - been progressing them through IETF
  - propose to use last undefined bit in IP packet header
    - we don’t underestimate the task ahead
- huge effort trying to pervert protocol
  - two major flaws successfully fixed without additional complexity
- seeking wider collaboration
  - co-operative or adversarial

intro
status
solution

approach

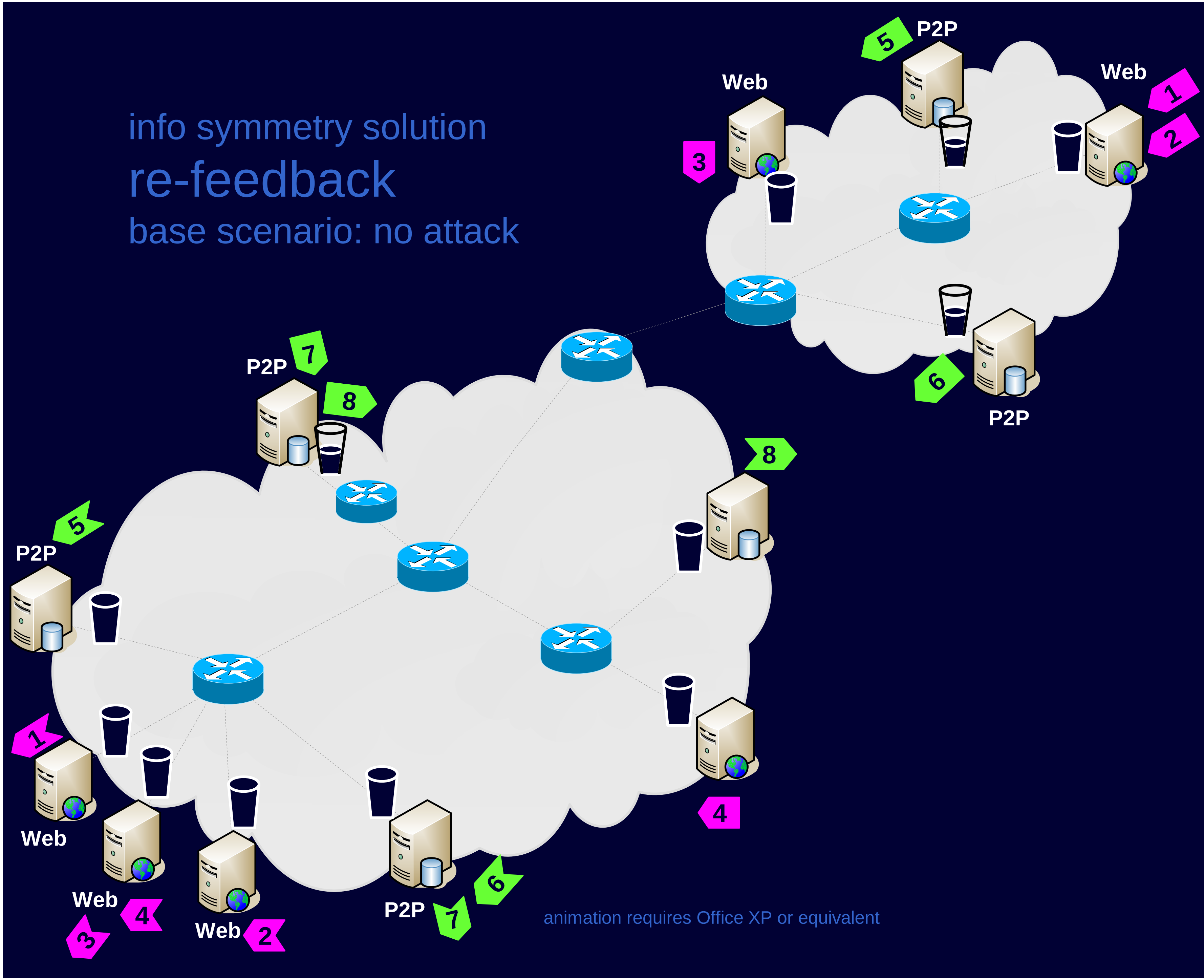
- fix generic IP layer first
  - will raise the bar (increasing attacks on higher layer vulnerabilities)
- treat DoS for what it is: extreme congestion – an externality
- genuine sources should slow down in response to congestion
  - voluntary response inherent to current Internet design
  - persistently sending fast into high congestion is never genuine behaviour
  - ★ don’t need to judge good/bad, ISP can just force response to congestion
  - ★ stability of Internet depends on congestion response anyway
- ★ designers don’t mandate congestion response, each ISP does
  - ★ market decides
  - ★ but relevant ISP liable for externality if it doesn’t act
- ★ focus on liabilities between networks
- ★ enforce liability for congestion externality, but recursively
  - ★ N_B liable for congestion it lets into N_C and onward
  - ★ N_A liable for congestion it lets into N_B and onward



★ requires solution to an inherent information asymmetry in IP info symmetry solution

re-feedback

base scenario: no attack

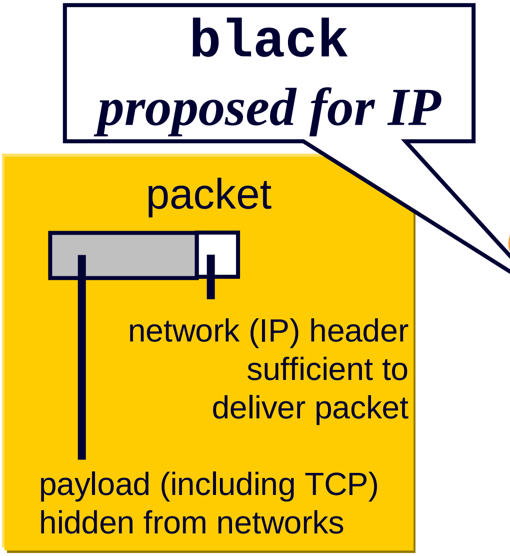


animation requires Office XP or equivalent

solution

information symmetry 're-feedback'

- currently N_A contracts with N_B to deliver packets but without information about N_B's quality (congestion)
- S_1 has this information, so make it reveal it



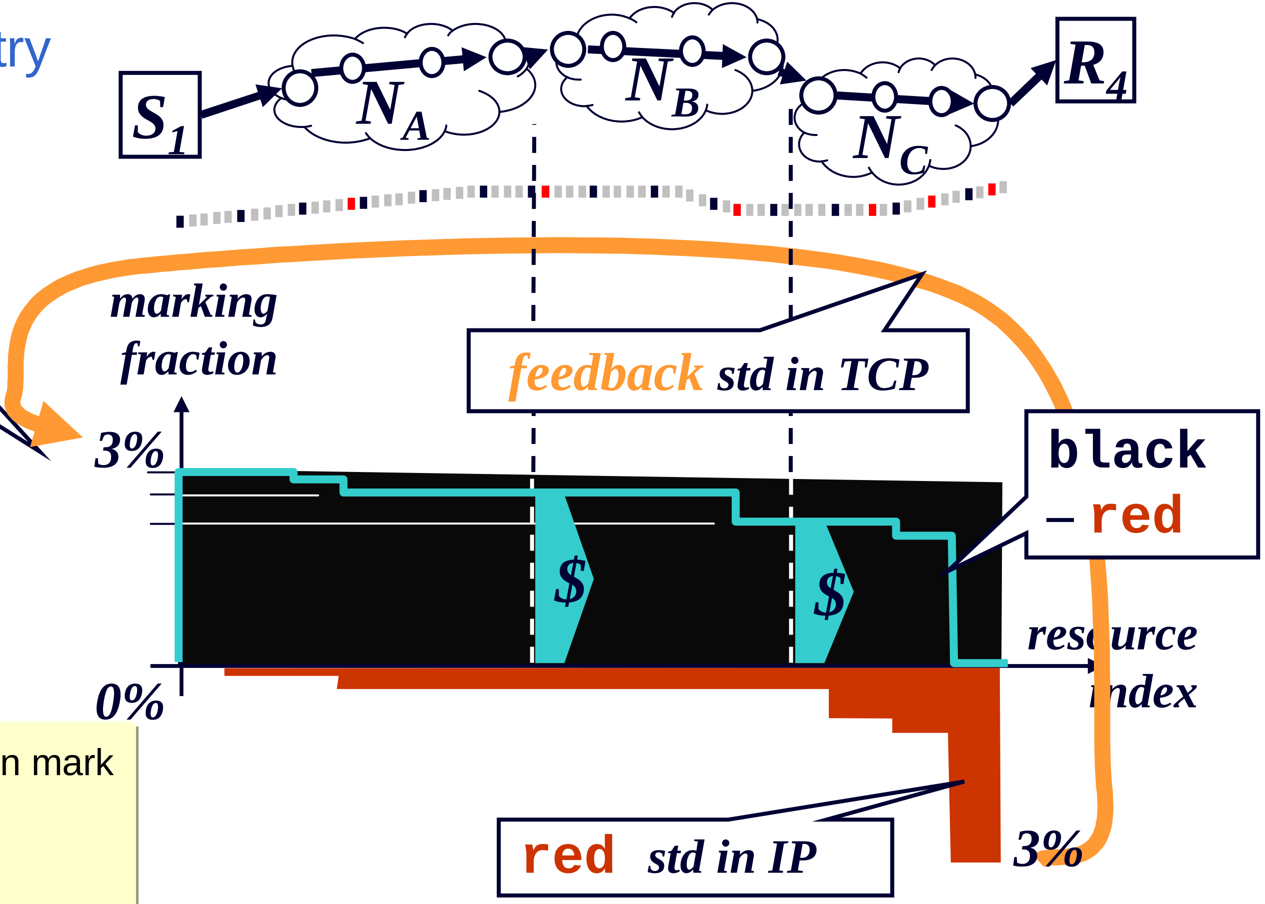
black

proposed for IP

packet

network (IP) header sufficient to deliver packet

payload (including TCP) hidden from networks



***feedback* std in TCP**

black – red

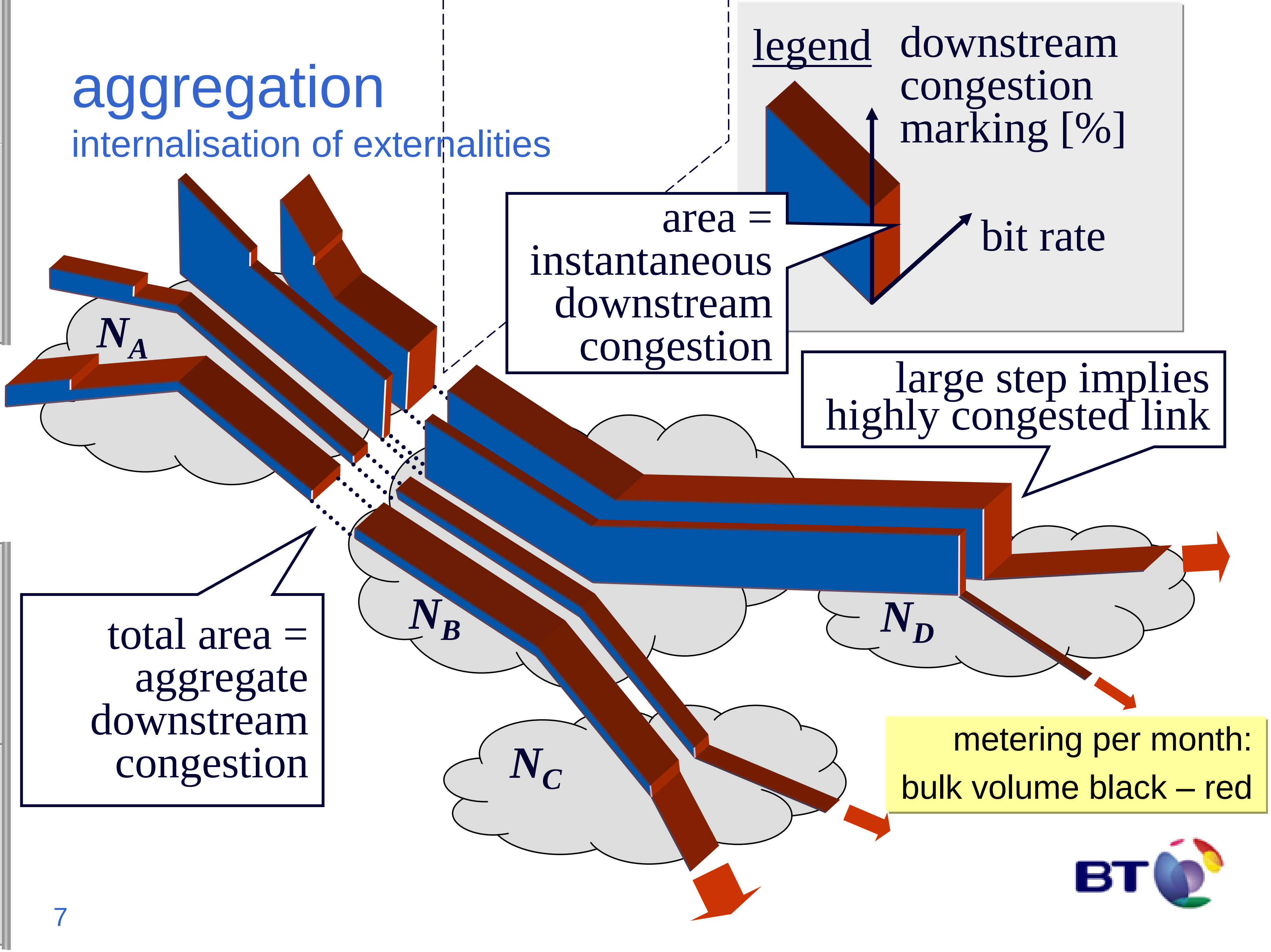
red *std in IP*

- routers approaching congestion mark some packets **red**
  receiver **feeds back** to sender
  - already standardised & implemented
  - not generally turned on by operators
- sender re-inserts **feedback** by marking packets **black**
  - **re-feedback** requires standardisation

- flows get no further than their 'fare' pays for
- routers discard persistent negative balance

aggregation

internalisation of externalities



legend downstream congestion marking [%]

bit rate

area = instantaneous downstream congestion

large step implies highly congested link

total area = aggregate downstream congestion

metering per month:

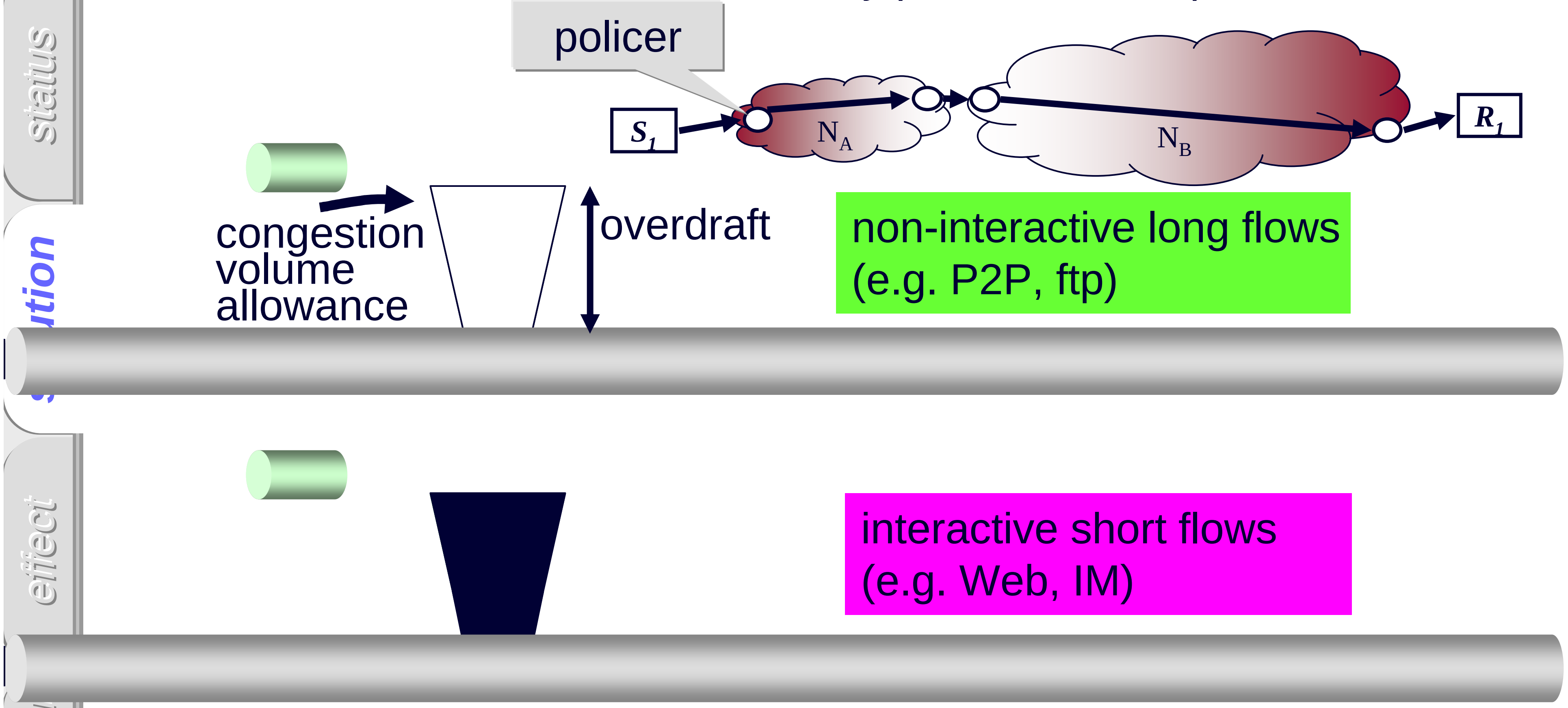
bulk volume black – red

congestion policer

base scenario: no attack

one example: per-user policer

many possibilities – up to ISPs



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

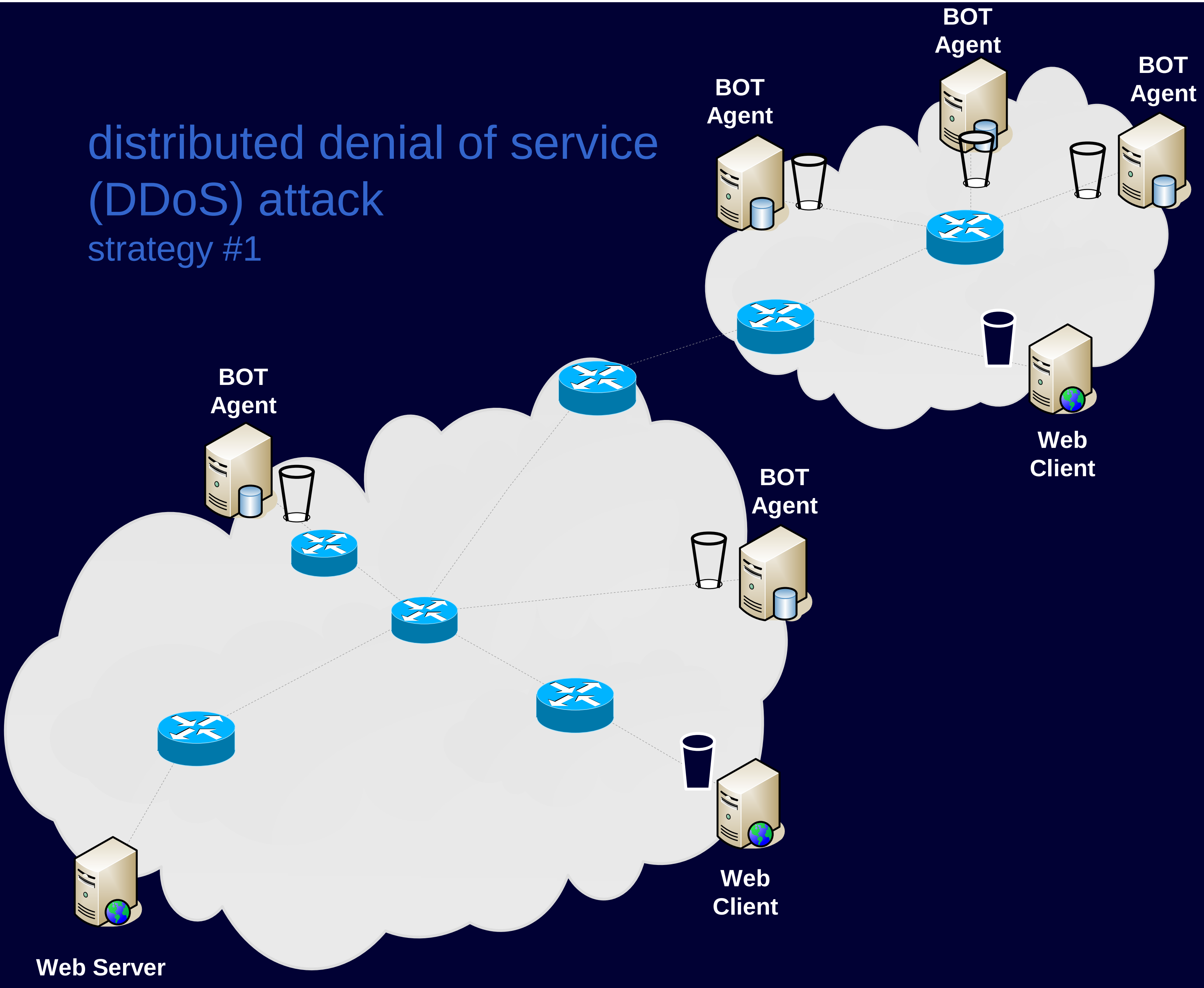
interactive short flows (e.g. Web, IM)

two different customers, same deal

animation requires Office XP or equivalent

distributed denial of service (DDoS) attack

strategy #1

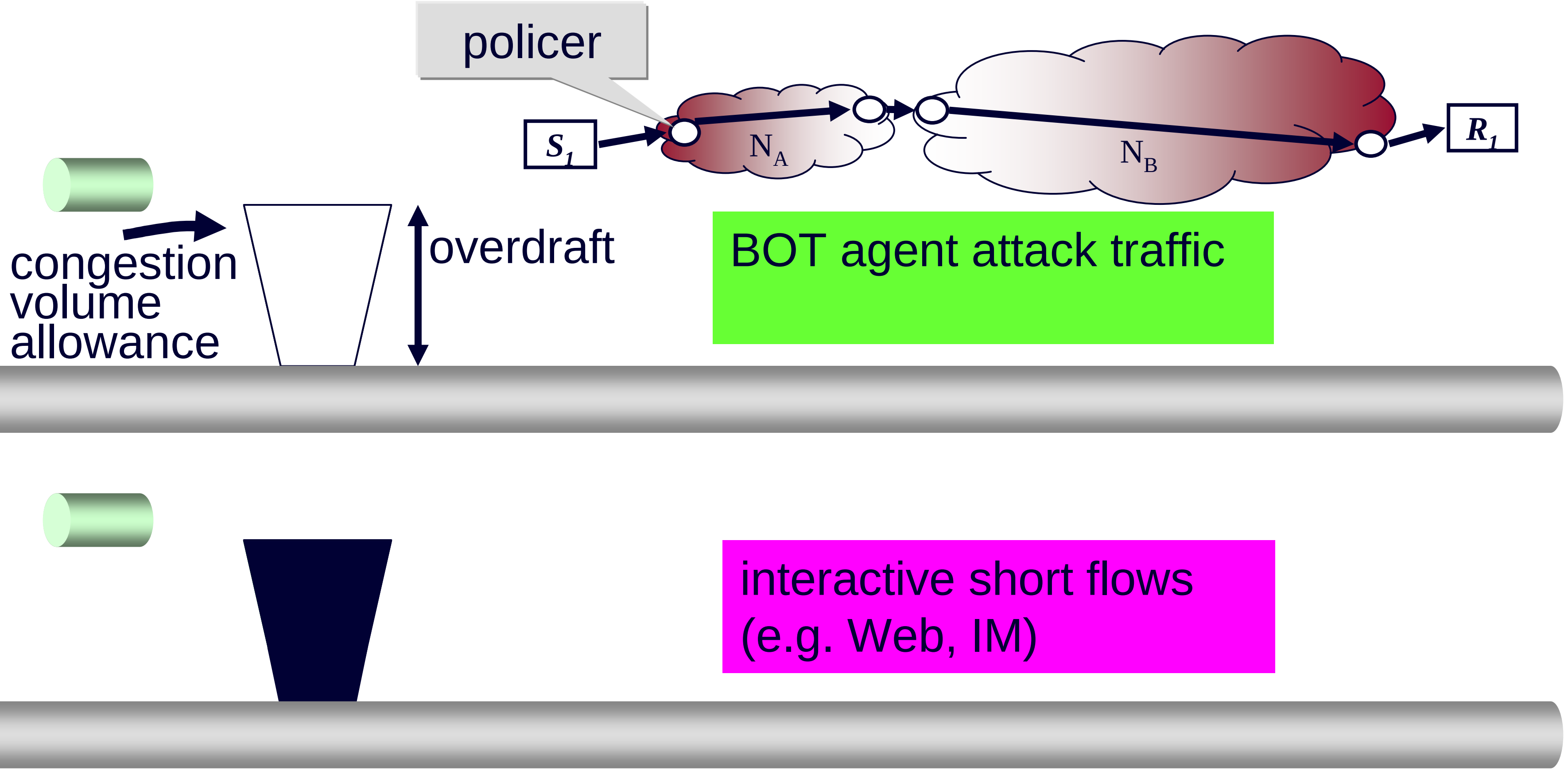


animation requires Office XP or equivalent

intro
status
ution
effect
deployme

per-user congestion policer

DDoS attack strategy #1



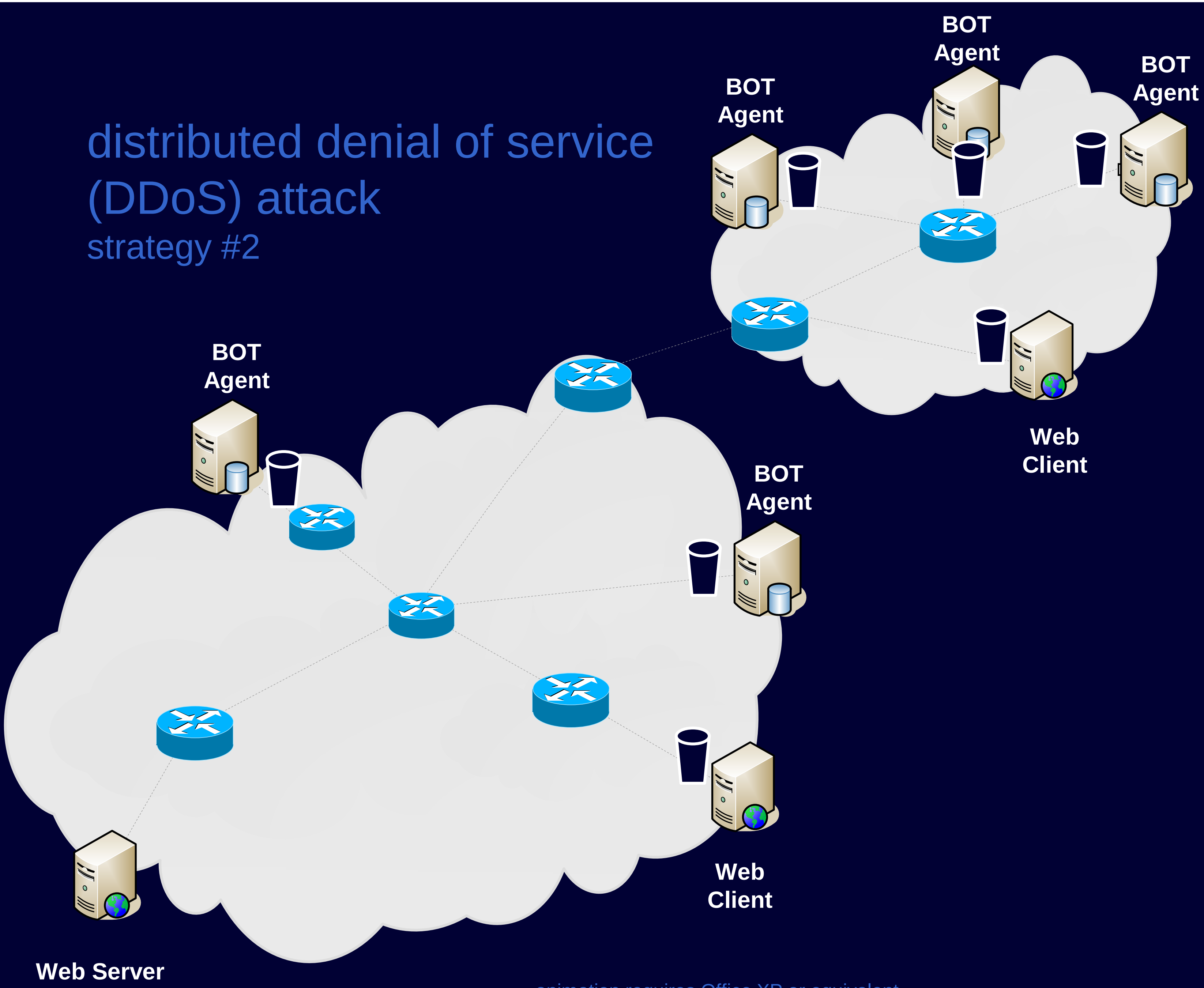
BOT agent attack traffic

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

will re-feedback prevent DDoS?
≡ will it be deployed widely *enough*?

- deployment bootstrap incentives

- deployment closure incentives
  - doesn’t have to finish the job itself
  - can create right incentives to deploy complementary solutions
- once fully deployed, winning the war
  - distinguishing genuine flash crowd from simultaneous attack

deployment bootstrap incentives

- deployment effectively involves architectural change
  1. (minor) change to sender's Internet stack
  2. network deploys edge/border incentive functions
- preventing gridlock between these actors requires strong incentives

deployment bootstrap incentives

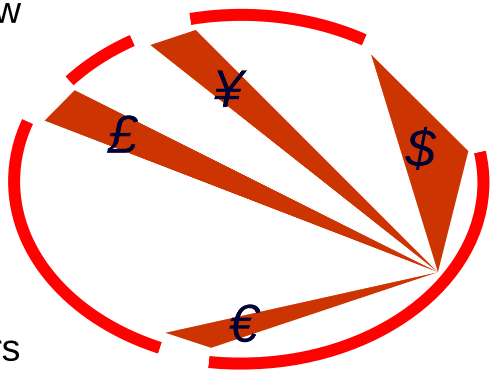
- ★ bundling with itself
  - re-feedback solves central cost control problem of ISPs
    - third party services competing with ISP pay below network cost
    - ISP has to compete *while* paying balance of competitor’s costs
  - hits very big fear and button and greed button
  - but keeps moral high ground
    - net neutral and doesn’t help lock-in or lock-out
  - re-f/b as a solution to DDoS bundled with re-f/b as cost-control
- alliance deployment strategy
  - 3GPP alliance has most to lose from not deploying, followed by NGNs
  - controls vertically integrated network and mobile terminal market
- ★ deployment by cross-infection
  - nomadic, roaming devices
- ★ inverse bundling
  - can degrade a substitute product (legacy network service without re-feedback)
  - generally useful model for security products – tend to restrict rather than enhance

---

★ novel deployment models wrt Ozment & Schechter

deployment closure incentives

- assume 1st mover (cellular industry?) has deployed
- 2nd movers (NGNs?) didn’t because benefit lower than cost (if rational)
  - but first mover removed costs (risks of unknown, R&D recovered)
  - early adopters also change operational finances for non-adopters...
- money valve effect
  - between adopters and non-adopters
  - re-feedback controls congestion costs for adopters
  - peaks in incoming traffic demand drive money inward
  - outgoing traffic peaks only generate averaged money flow
    - costs of non-adopters depend on peak not average
  - stronger effect, the more variance in demand
  - DDoS is extreme variance in demand
  - like alternating current through a diode/valve
- chain reaction
  - adopters’ incoming border charges focus on non-adopters
  - bots concentrate into smaller non-adopter space
  - money valve effect surrounds more of non-adopters’ borders



winning the last battle (not just the next)

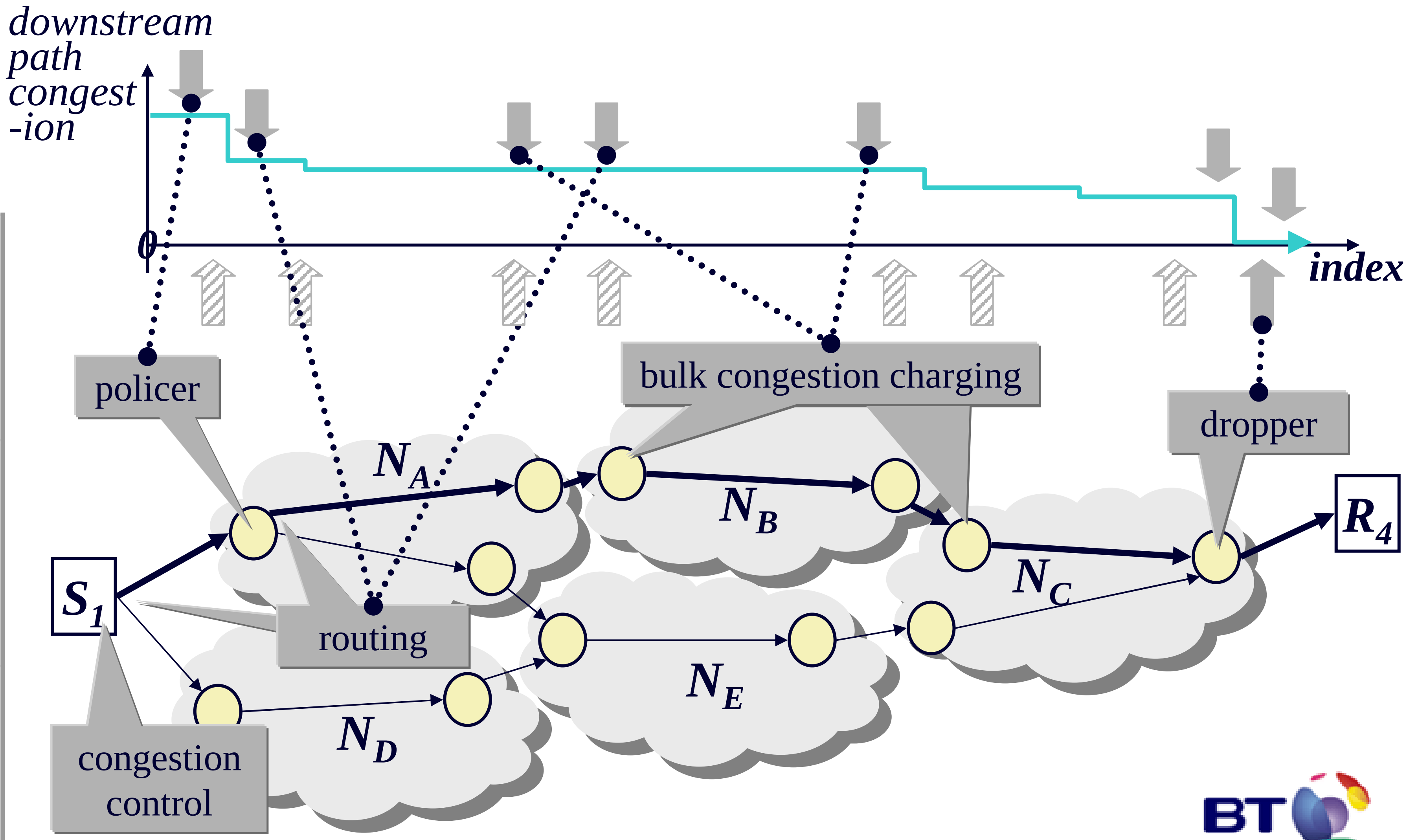
distinguishing flash crowds from attacks

- incentives not to be too greedy
  - a rate policer is effectively a revenue limiter
  - if policer allows DDoS attacks, customer has to buy bigger quota
  - why would operators try to distinguish the two?
- customers will switch to responsible operators
  - distinguishing true demand form zombies is in operator’s interest
- fortunately society still civilised enough
  - huge white market revenue not worth risking
    - just to capture marginal gains from black market
  - strategic greed overcomes myopic greed

Self-interest can Prevent Malice

Q&A

incentive framework


downstream
path
congest
-ion
0
index
policer
bulk congestion charging
dropper
routing
congestion control
N_A
N_B
N_C
N_D
N_E
S_1
R_4