The Concept of Quality of Service in the Internet

Geoff Huston
APNIC
RIPE 65, September 2012

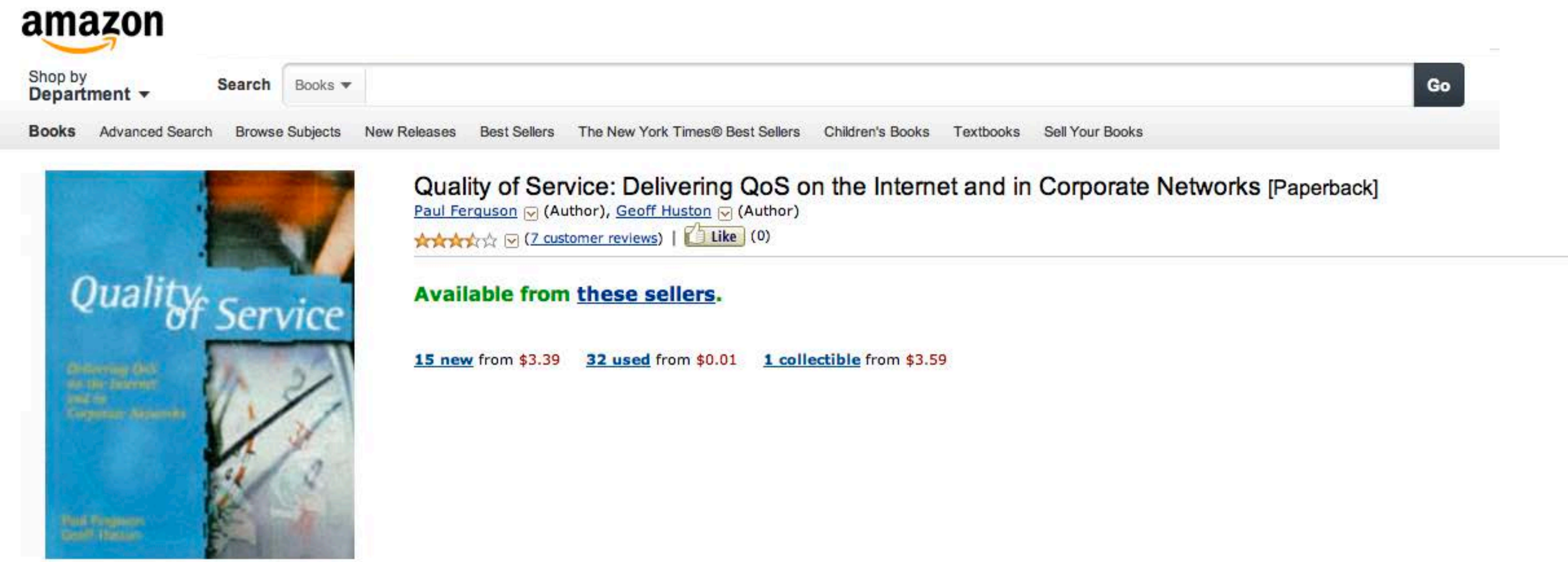
31 August 2012

THIRD DRAFT OF THE SECRETARY-GENERAL'S REPORT

for the
Fifth World Telecommunication/Information and Communication Technologies Policy Forum
2013

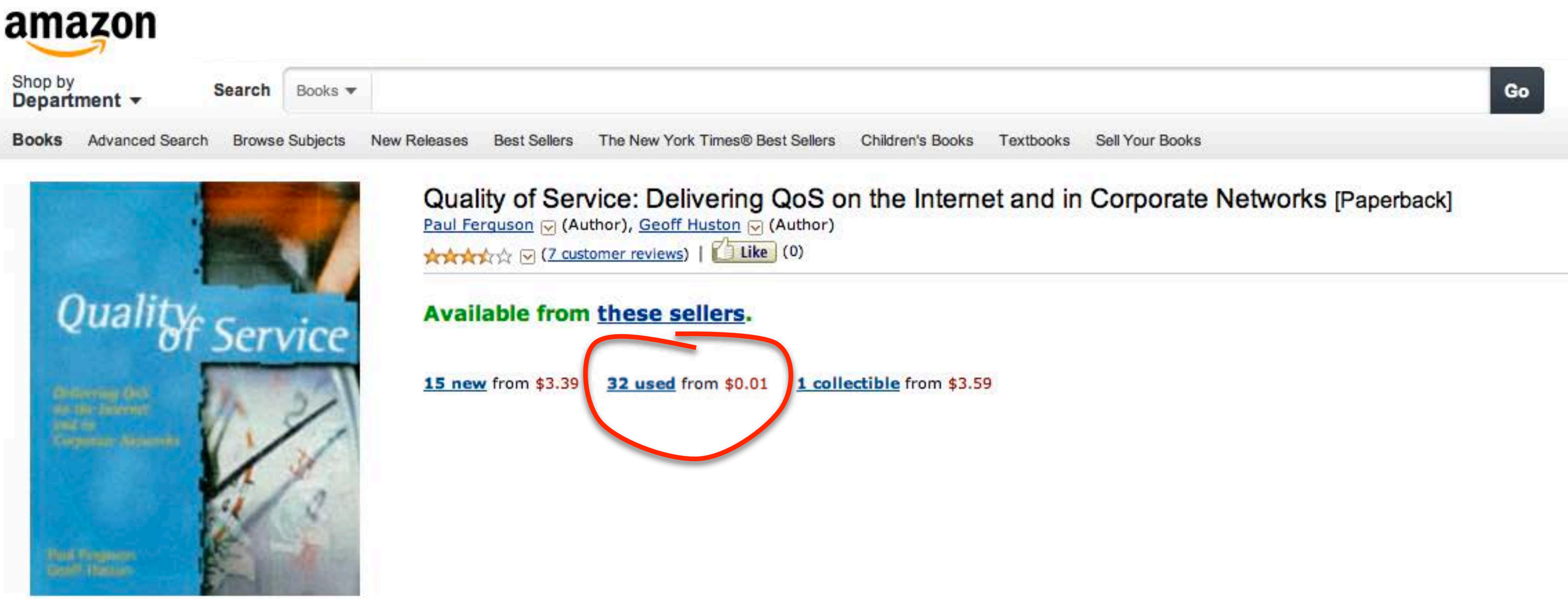
"With the move from traditional networks (based on dedicated service-channels and/or separate networks for each service) to integrated (transport) services on a single packet-based transport infrastructure, pre-defined transmission planning of Quality of Service (QoS) has become a major challenge, since many IP-based networks might not provide for self-standing end-to-end QoS, but only transport classes, which enable QoS differentiation. IP-based networks can support end-to-end QoS if the routers in between support the mechanisms and the network is designed for QoS."

A little while ago…



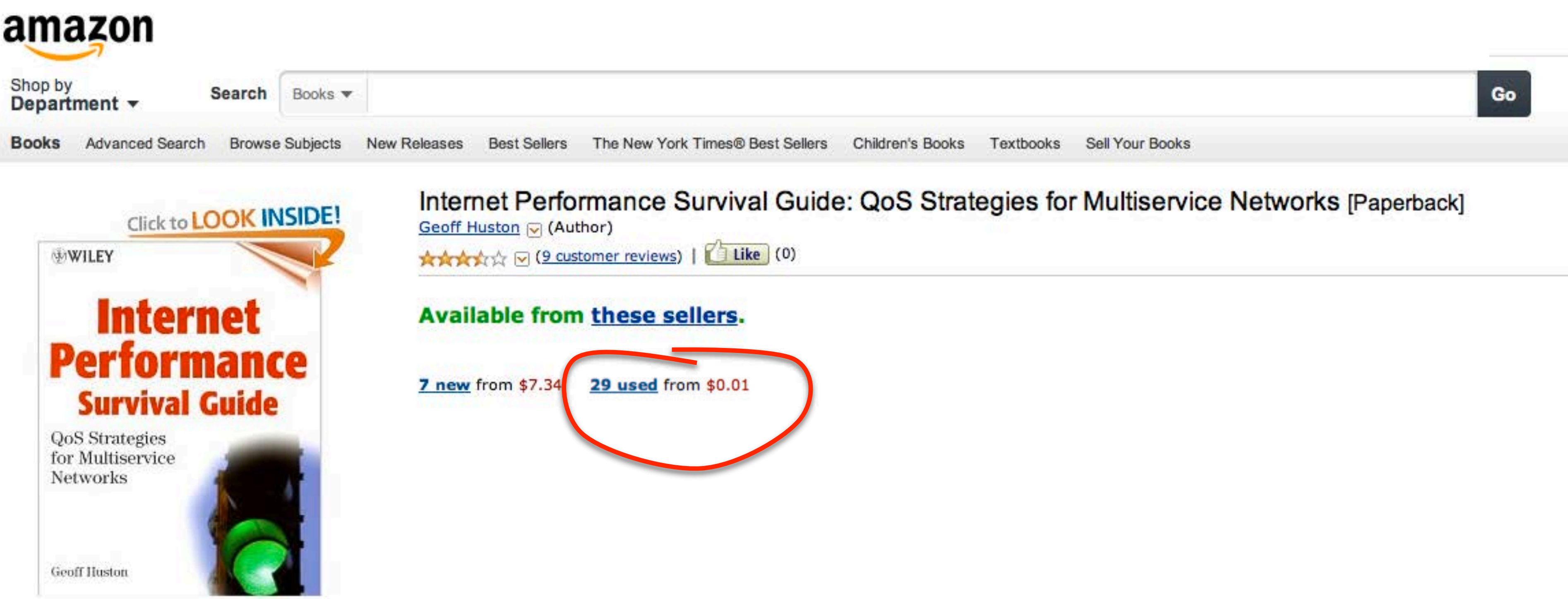
"Regardless of whether you are trying to implement QoS in a private network, or within a segment of the global Internet, QoS comes at a cost. There is no magic here."

A little while ago…



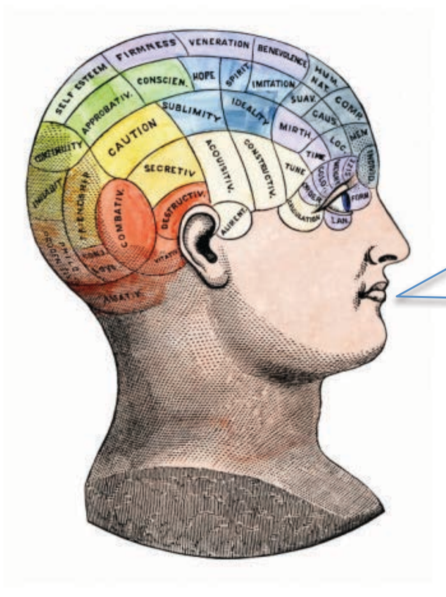
"Regardless of whether you are trying to implement QoS in a private network, or within a segment of the global Internet, QoS comes at a cost. There is no magic here."

Round 2:



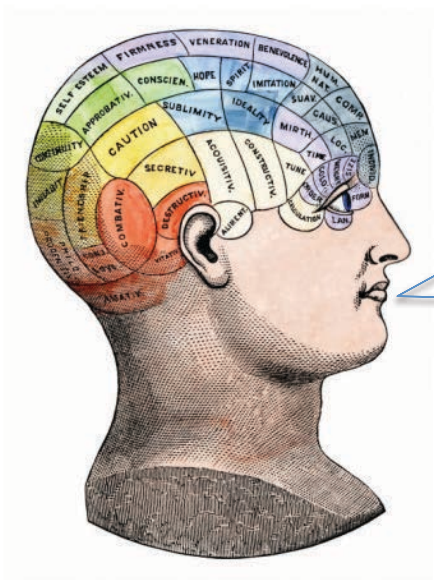
QoS: “Caveat Emptor”

Voice Networks



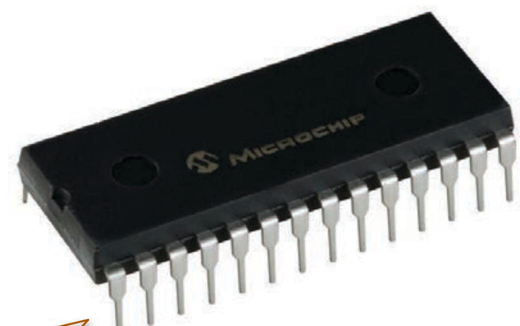
300 - 3500 Hz
Most of the energy is below 1Khz
Dynamic range of 50db

Voice Networks

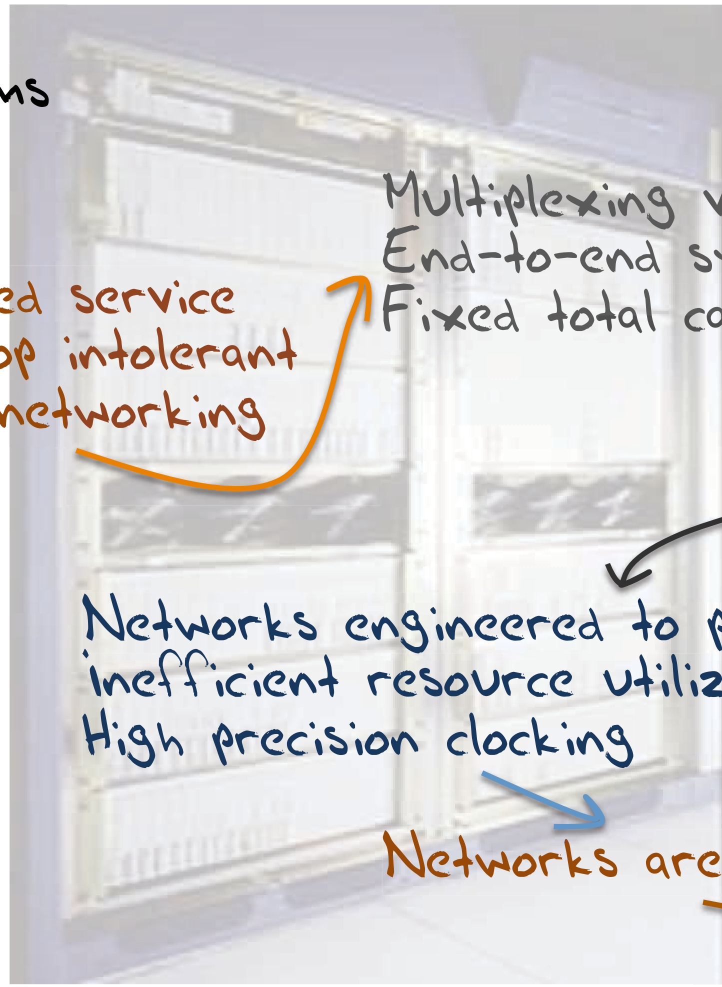


300 - 3500 Hz
Most of the energy is below 1Khz
Dynamic range of 50db

Digitization:
8000 samples / second
65,000 discrete levels
A-law encoding reduces this to 256 levels
64Kbps real time bitstream



Voice Networks



64K bitstreams

Tightly defined service
Jitter and drop intolerant
Synchronous networking

Multiplexing via strict time switching
End-to-end synchronous virtual circuits
Fixed total capacity

Networks engineered to peak load profile
inefficient resource utilization
High precision clocking

Networks are costly to run

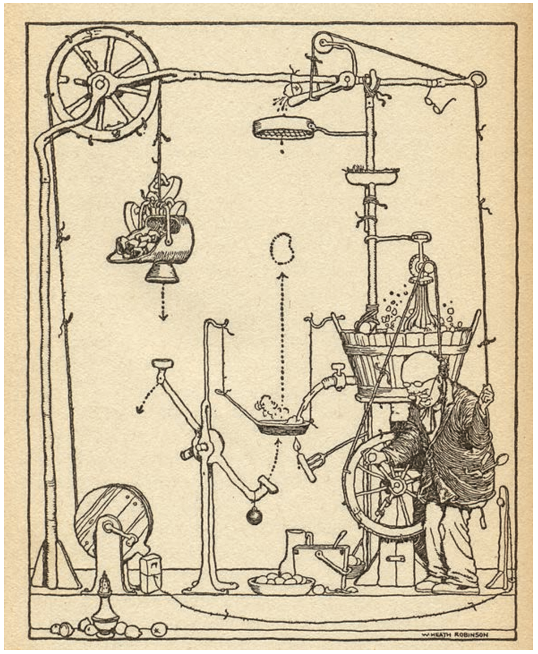
Services are expensive!

Engineering Voice Networks

it's challenging to add capacity to operational circuit switching networks - so it was common practice to overprovision the networks and wait for demand to grow!

Data Networks

Due to marginal levels of demand data networks were originally provisioned on the margins of oversupply of voice networks



Early data protocols borrowed many concepts from the voice network's functions:

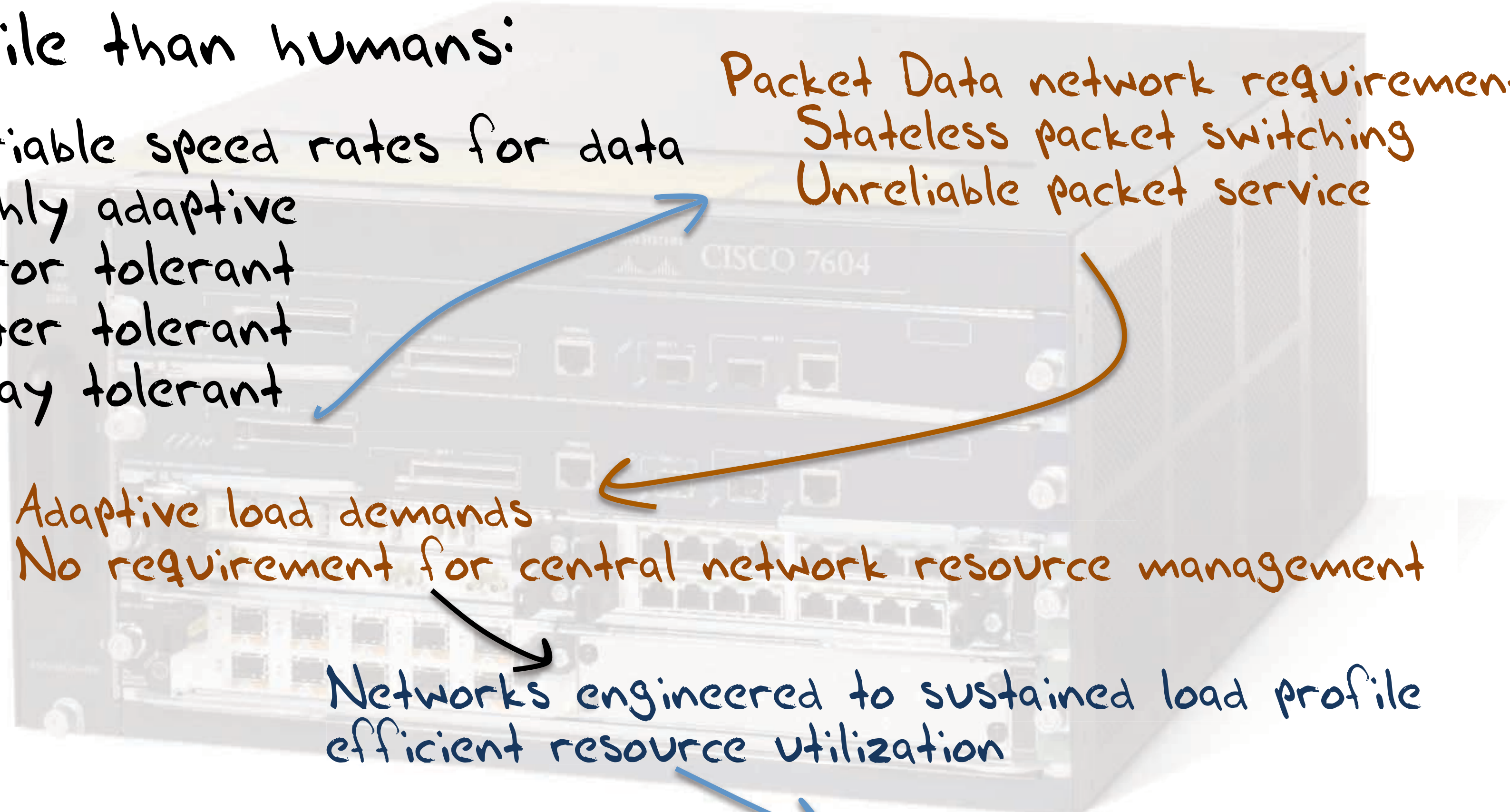
- Point-to-point Virtual Circuits
- Network defined capacity
- Synchronous bitstream services

Packet Networks

Computers are far more versatile than humans:

Variable speed rates for data
Highly adaptive
Error tolerant
Jitter tolerant
Delay tolerant

Packet Data network requirements:
Stateless packet switching
Unreliable packet service



Adaptive load demands
No requirement for central network resource management

Networks engineered to sustained load profile
efficient resource utilization

Networks are cheaper to run

Services are inexpensive!

The Evolution of the Common Network Platform Model

Data

Voice

Services are cheap!

Services are expensive!

Costs!

Everything over

IP!

How can you efficiently mix congestion-prone and congestion intolerant applications within a single network platform?

How can you efficiently mix congestion-prone and congestion intolerant applications within a single network platform?

Add more bandwidth!

Too easy!

How can you efficiently mix congestion-prone and congestion intolerant applications within a single network platform?

Add more bandwidth!

Too easy!

Change the adaptive behaviour of the applications

Too hard!

The Goldilocks Procedure!

How can you efficiently mix congestion-prone and congestion intolerant applications within a single network platform?

Add more bandwidth!

Too easy!

Add resource management functions to the network

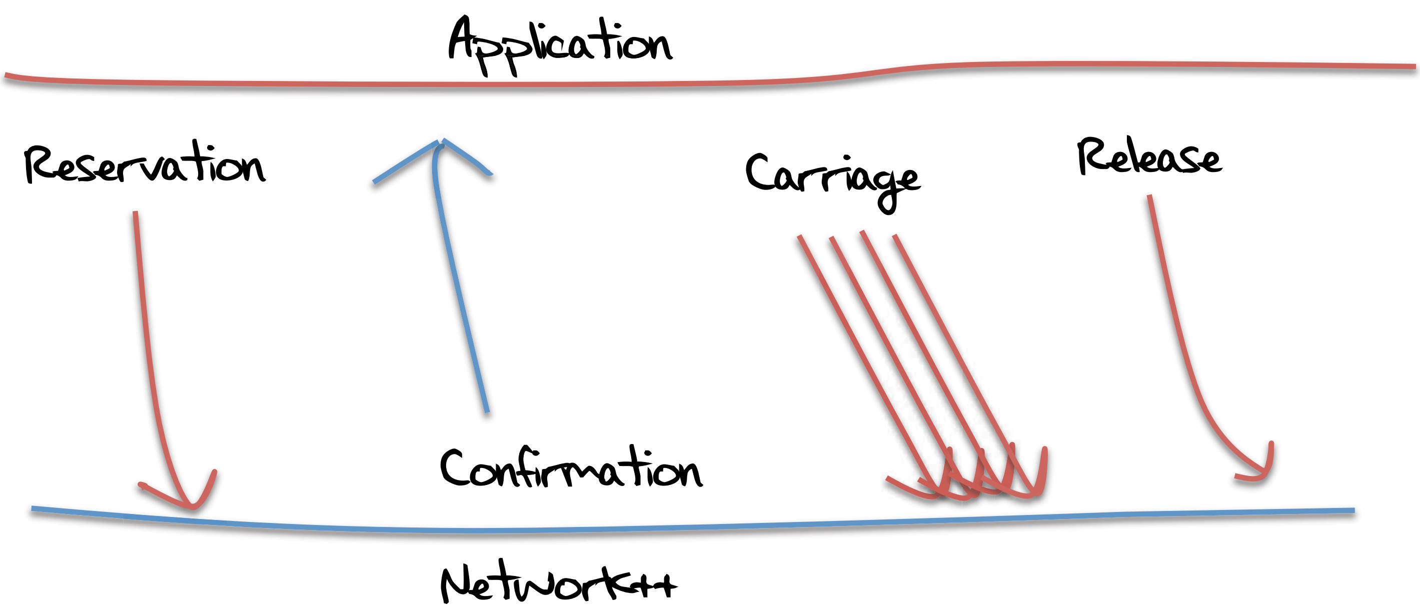
Just right! (supposedly)

Change the adaptive behaviour of the applications

Too hard!

IP QoS -- Version 1

integrated Services

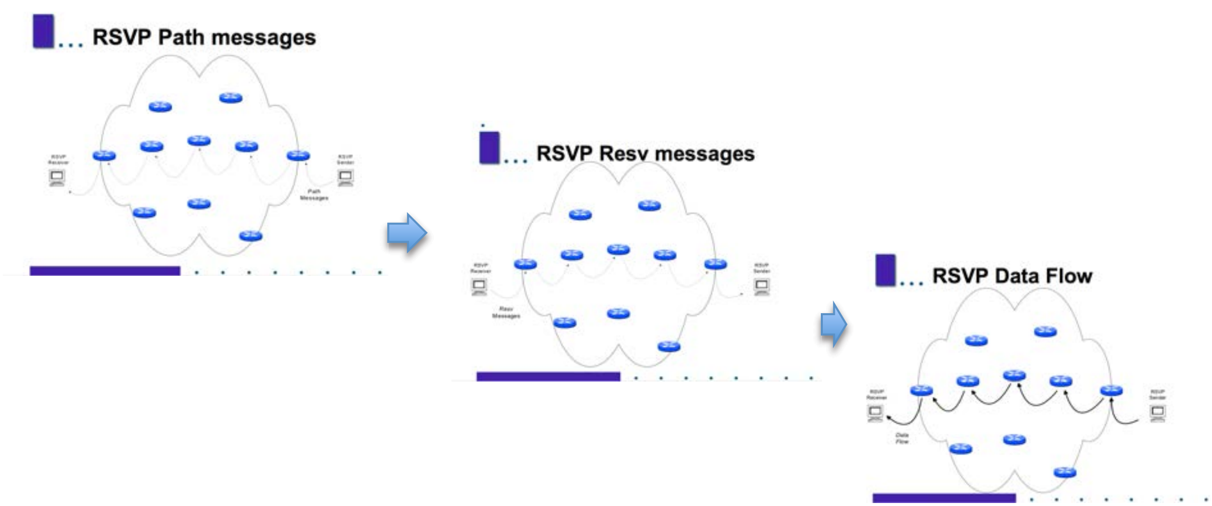


(Network equipped with admission control, virtual circuits and resource reservation capability)

"Integrated Services"

Adds the concept of a "flow state" into the network

The network must distribute a resource reservation along a static ("pinned") flow path



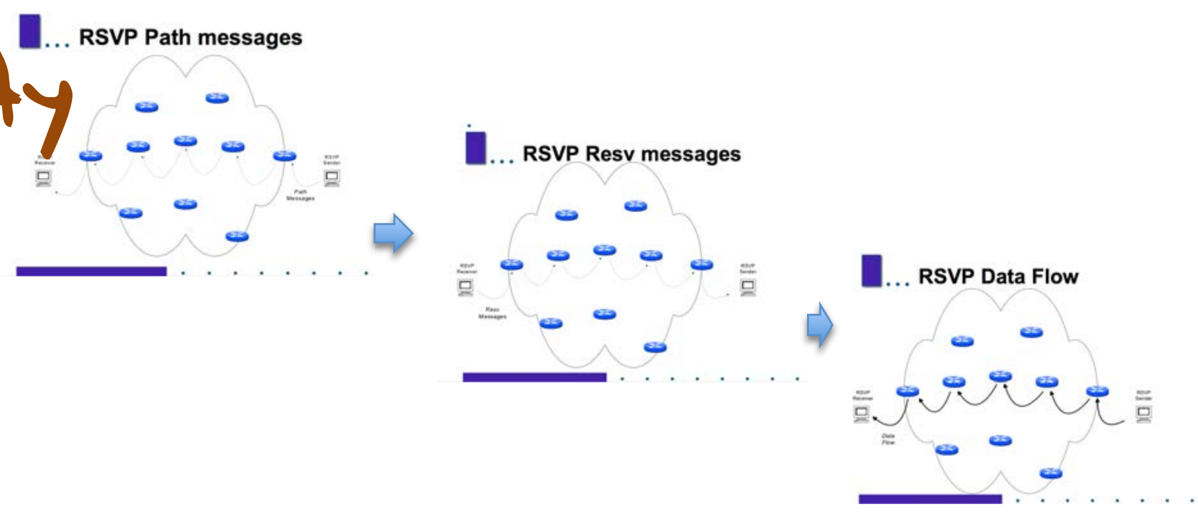
"Integrated Services"

Adds the concept of a "flow state" into the network

The network must distribute a resource reservation along a static ("pinned") flow path

This creates within the data network:

- state
- complexity
- fragility
- COST!



"Integrated Services"

Adds the concept of a "flow state" into the n...

The network must distribute a re...

static ("pinned") flow pa...

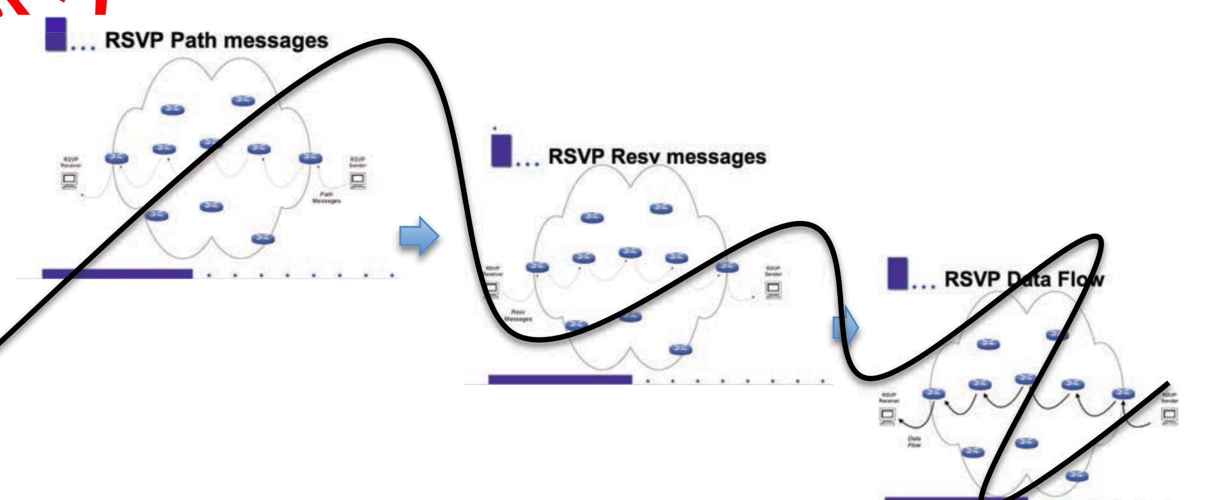
This creat...

...lexity

fragility

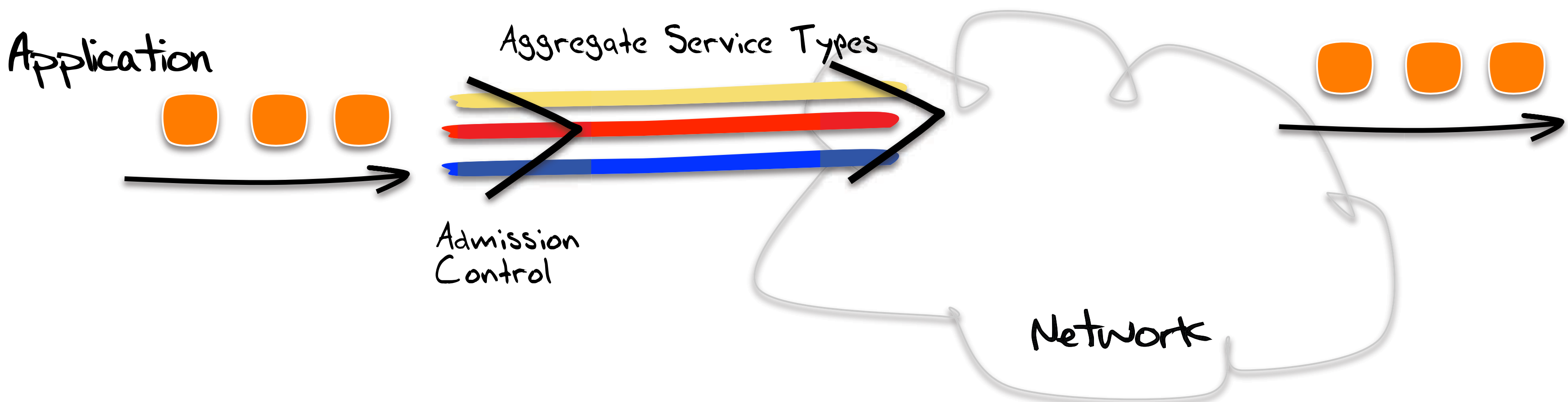
COST!

This form of QoS architecture simply does not scale!



IP QoS -- Version 2

Differentiated Services



"Differentiated Services"

This is a pretty simple rerun of the TOS packet painting approach

It's stateless, so it has more potential to scale to larger networks

... Differentiated Services

- Active differentiation of packet-based network traffic to provide a *better than best effort* performance for a defined traffic flow, as measured by one of more of:
  - Packet jitter
  - Packet loss
  - Packet delay
  - Available peak flow rate
- Implementable within a large network.
- Relatively difficult to measure success in providing service differentiation.

"Differentiated Services"

This is a pretty simple rerun of the TOS packet painting approach

It's stateless, so it has more potential to scale to larger networks

But DiffServe service outcomes are relative, not absolute

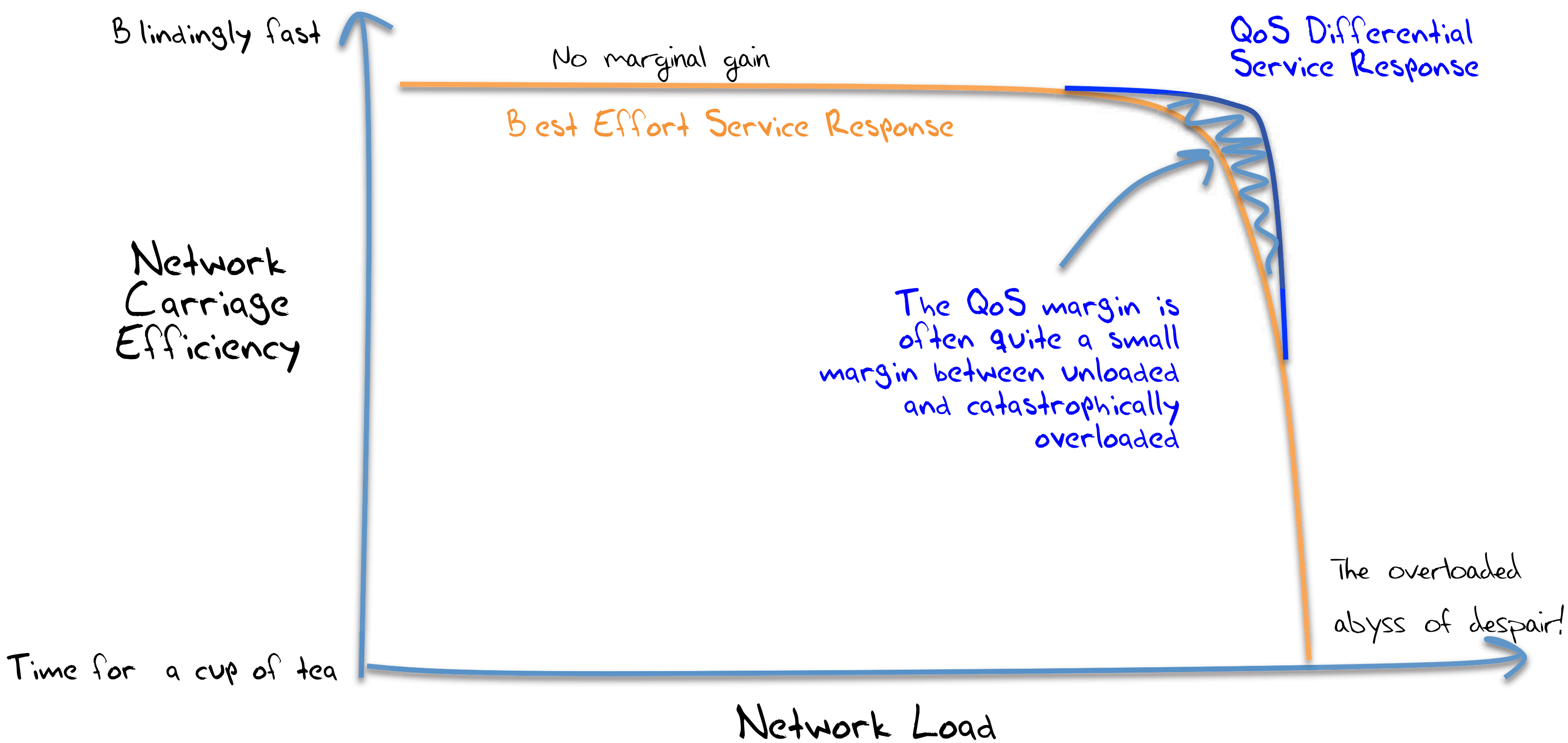
And there is no effective form of feedback control to monitor the outcomes that the network is providing

... Differentiated Services

- Active differentiation of packet-based network traffic to provide a *better than best effort* performance for a defined traffic flow, as measured by one of more of:
  - Packet jitter
  - Packet loss
  - Packet delay
  - Available peak flow rate
- Implementable within a large network.
- Relatively difficult to measure success in providing service differentiation.

"Differentiated Services"

What is DiffServe attempting to tinker with?



"Differentiated Services"

This is a pretty simple rerun of the TOS packet painting approach

It's stateless, so it has more potential to scale to larger networks

But ...

rel...

And ... ...rm of

feedb... ...ol to monitor the outcomes that the network is providing

. . . Differentiated Se... ...es

- Active differenti... ...work traffic t...

...within a large network.

...atively difficult to measure success in providing service differentiation.

Can't perform per-flow resource reservations

Can't deliver assured outcomes

Can't guarantee fixed service response

Can't see it and can't measure it!

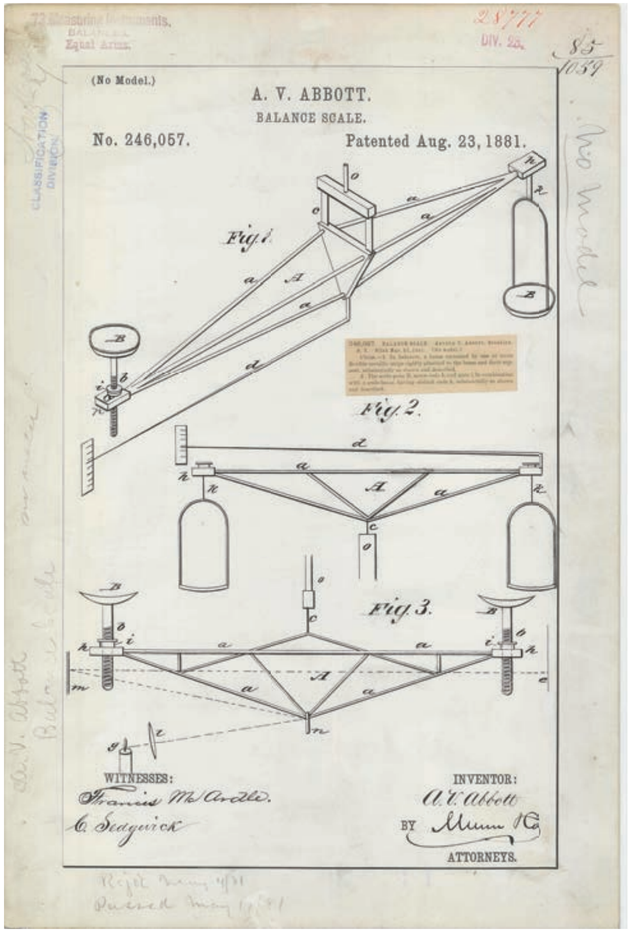
And so on and so on...

- NSIS effort to standardise the signalling protocol between the application and the network for diffserve
- MPLS as the elastic QoS band aid!
- "Aggregated QoS" as an amalgam of Intserv and Diffserve, achieving none of either!

IP QoS

Balancing Cost and Benefit:

- Simple QoS mechanisms can be supported in small scale environments
- But as you try to scale up the QoS approach the cost rapidly increases and the relative benefits decrease
- it becomes a skewed exercise of spending 95% of your engineering budget to secure less than 1% of your revenue!



Why is IP QoS a Failure?

QoS does not create more network resources or a faster network

it just attempts to redistribute damage!

Why is IP QoS a Failure?

QoS does not create more network resources or a faster network

it cannot fix:

- over subscription
- buffer bloat and congestion
- poor network design
- poor business plans
- continental drift
- the speed of light

Why QoS?

Why QoS?

Why is ETNO so keen on QoS?

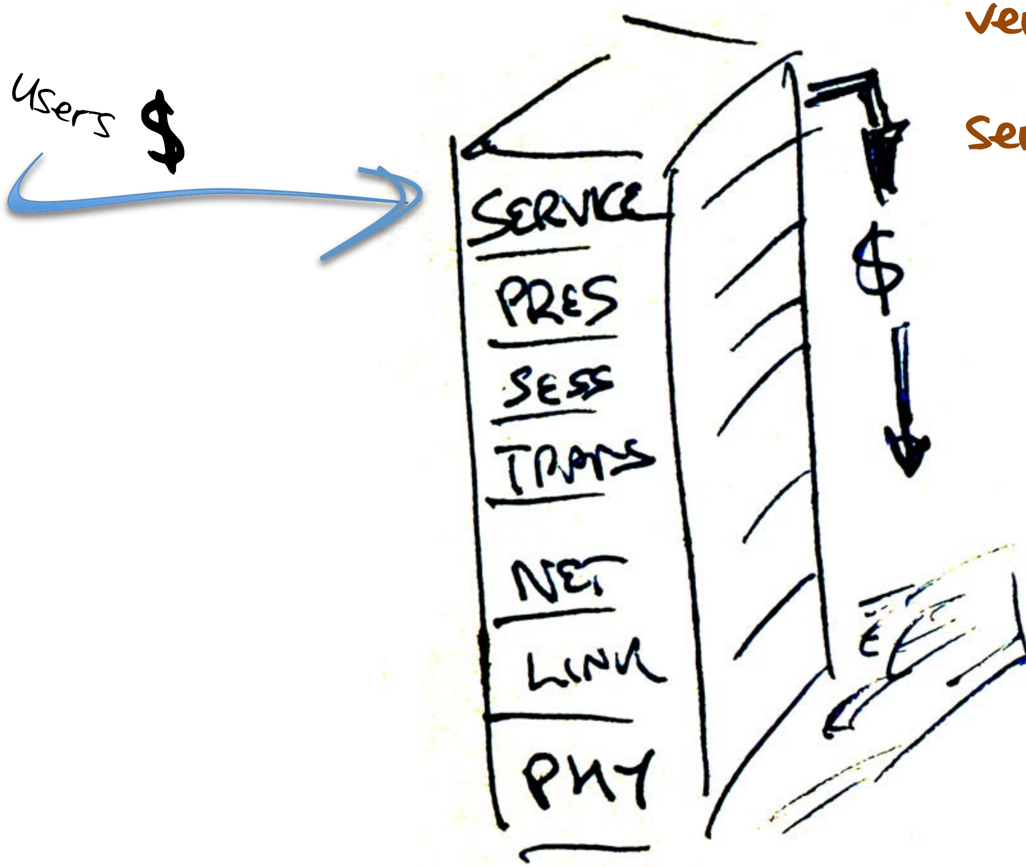
Why QoS?

Why is ETNO so keen on QoS?

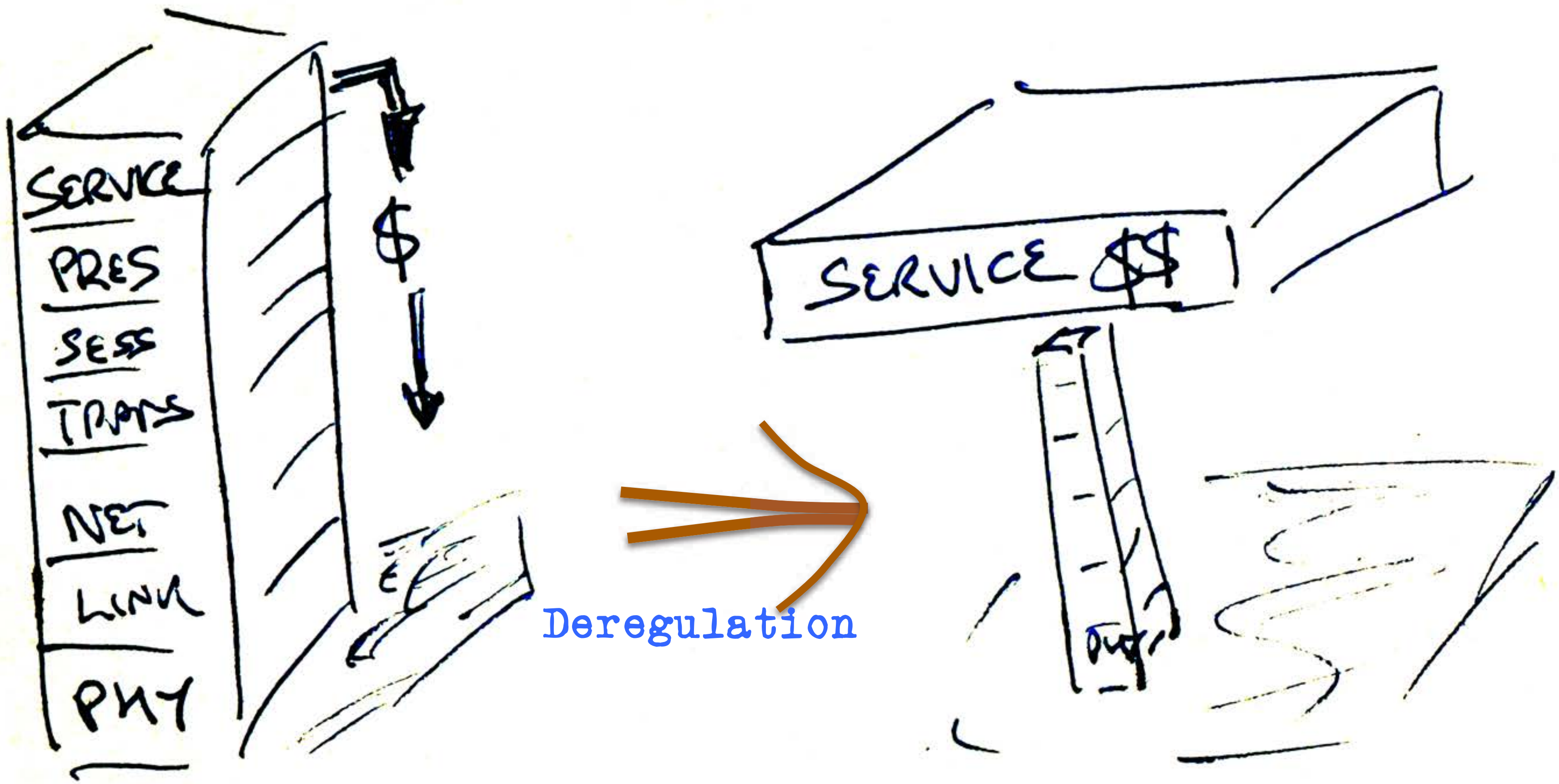
- Because QoS appears to offer network operators increased visibility and the possibility of control over traffic flows that are passed over their networks

Back to networking basics....

Telco nostalgia...

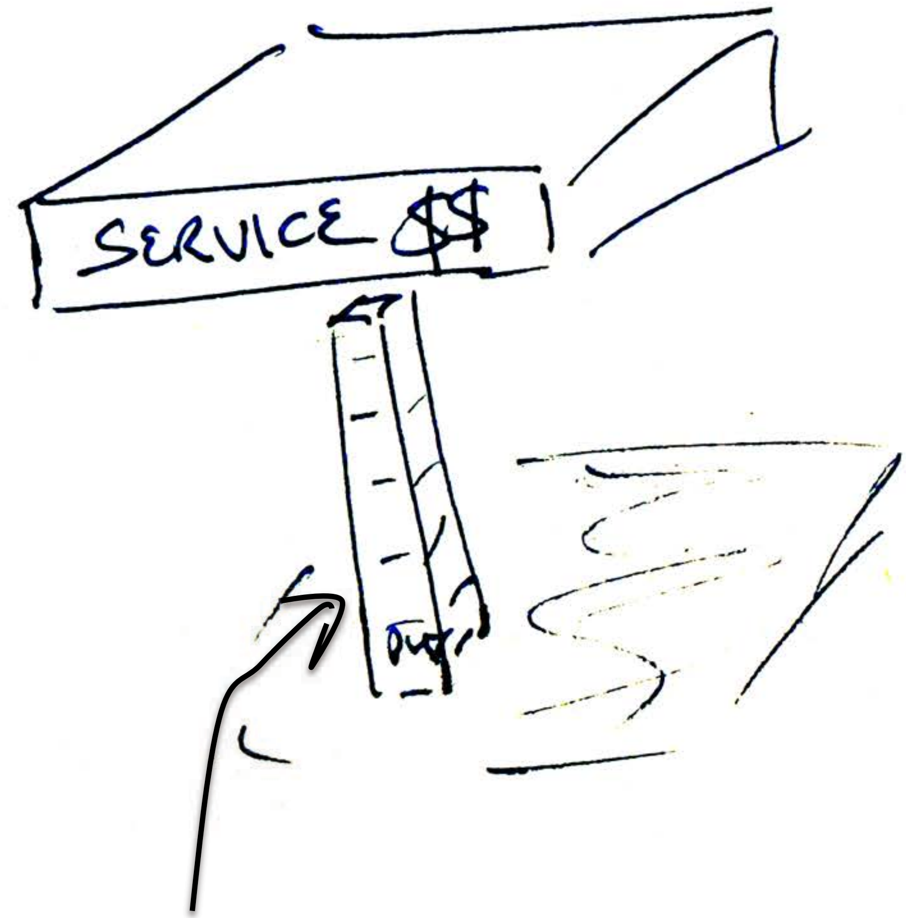


The historical vertically integrated service architecture

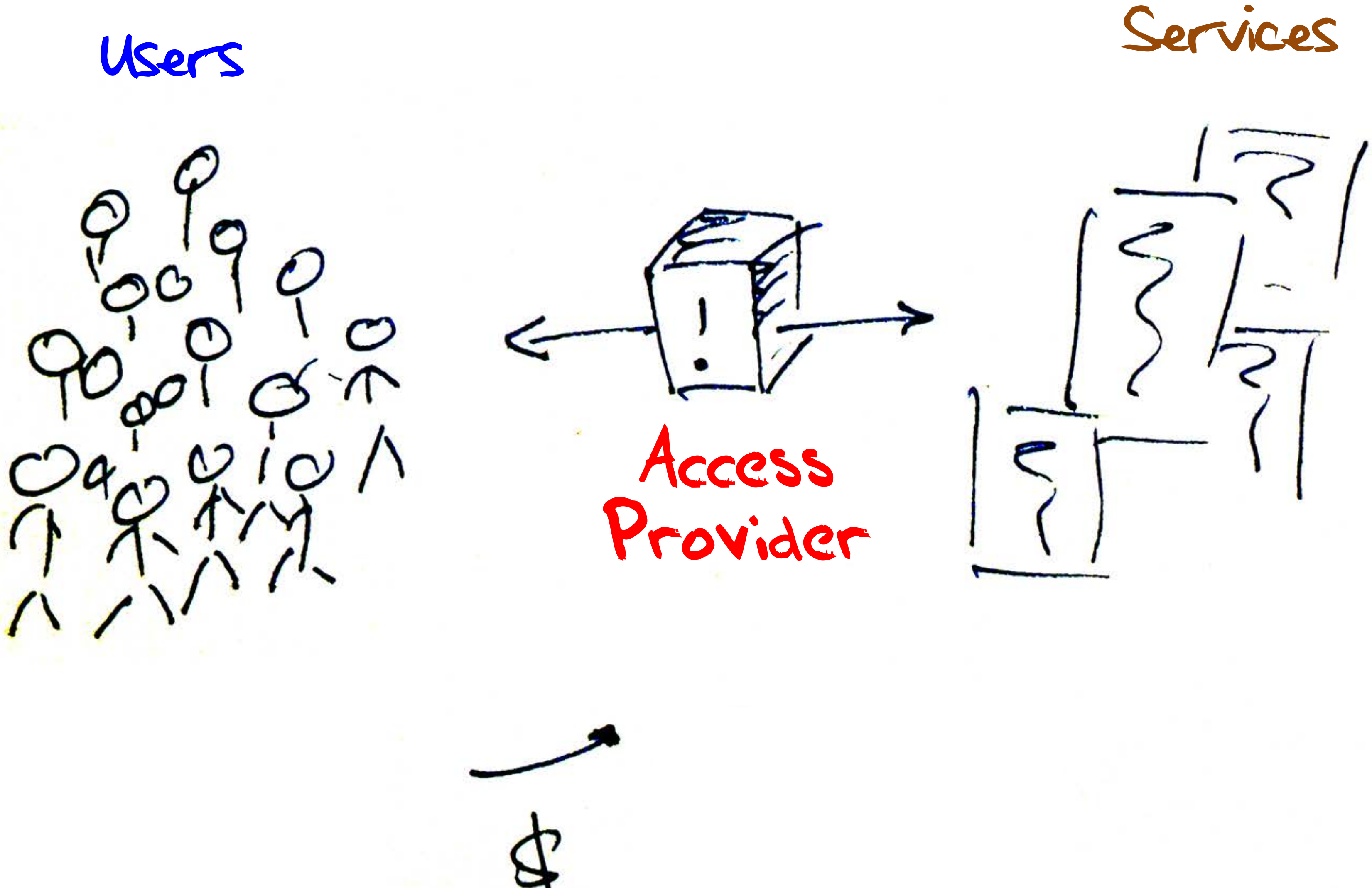


Devolution of the integrated service architecture through an open IP service architecture and deregulation

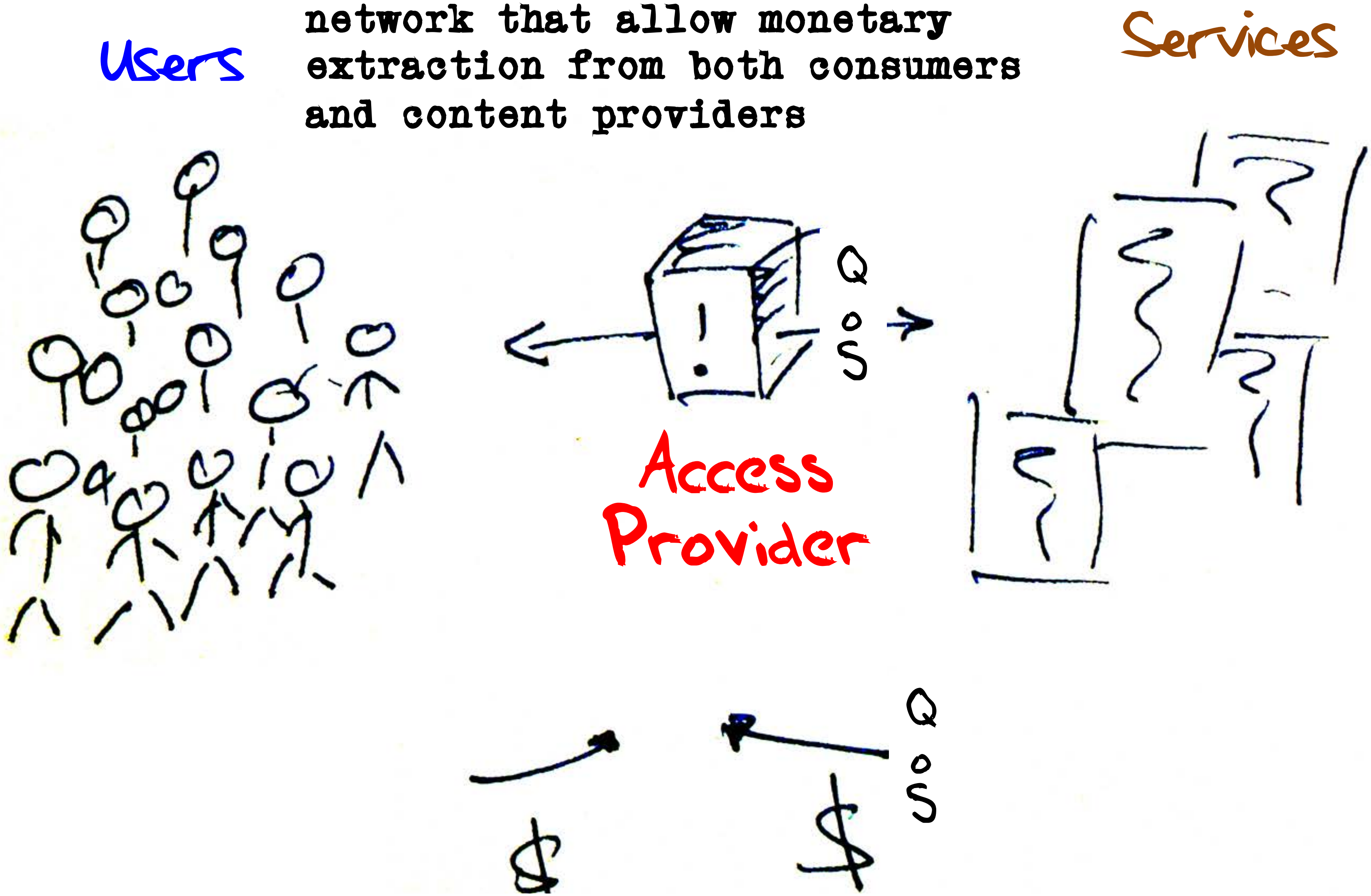
Devolution of the integrated service architecture



Where's the money to invest in new network services?


Users
Services
Access
Provider
$

Services-facing QoS provide control points in the IPv4 network that allow monetary extraction from both consumers and content providers



Why QoS?

Why is this control important?

- Because network operators believe that this will allow them to extort revenue from content service providers

Why QoS?

Why appeal to the ITU to mandate inter-provider IP QoS into the ITRs?

- Because when you are stuck with an unattractive business plan and you want to address this by generating an unnatural outcome in the market, there is nothing quite like having regulatory impost on your side!

Goldilocks was wrong!

How can you efficiently mix congestion-prone and congestion intolerant applications within a single network platform?

Add more bandwidth!

Current Operational Practice!

~~Add resource management functions to the network~~

Too hard!

Change the adaptive behaviour of the applications

Active Research Topic

Thank You!