QoS Control and Internet Traffic

QoS (quality of service) control

- feedback traffic control
  - → closed-loop/adaptive control
  - → small time scale: msec
  - → end-to-end and/or with router support
- resource provisioning
  - → open-loop control: allocate/reserve resources in advance
  - → large time scale: seconds, minutes, and longer
  - → router-based control with admission control at edge

Internet traffic

→ simplest of all: constant bit rate (CBR)

⟶ flat is good

⟶ predictable

Internet data traffic: variable bit rate (VBR)

⟶ session arrivals: random

⟶ exponential interarrival time

⟶ swimming with the fishes: Poisson process

⟶ e.g., telephone calls, TCP connections, fast food

Packet scheduling:

→ at routers/switches

→ default: FIFO with drop-tail packet drop

Granularity

- per-flow scheduling

  → e.g., single TCP session, UDP streaming session

- aggregate-flow scheduling

  → classes of traffic: e.g., platinum, gold, silver, bronze

Scheduling algorithms:

- priority scheduling

  → each flow or class assigned priority

  → simple but can lead to starvation

- weighted fair queueing

  → assign each flow/class a weight

  → generalizes round-robin

  → prevents starvation

  → requires traffic rate control to maintain service differentiation

- real-time scheduling

  → e.g., EDF (earliest deadline first)

QoS control:

- per-user (or flow) reservation for super-quality service

  → guaranteed service

- shared service classes (platinum, gold, silver, bronze) for good service

  → differentiated service

Internet standards:

- IETF IntServ

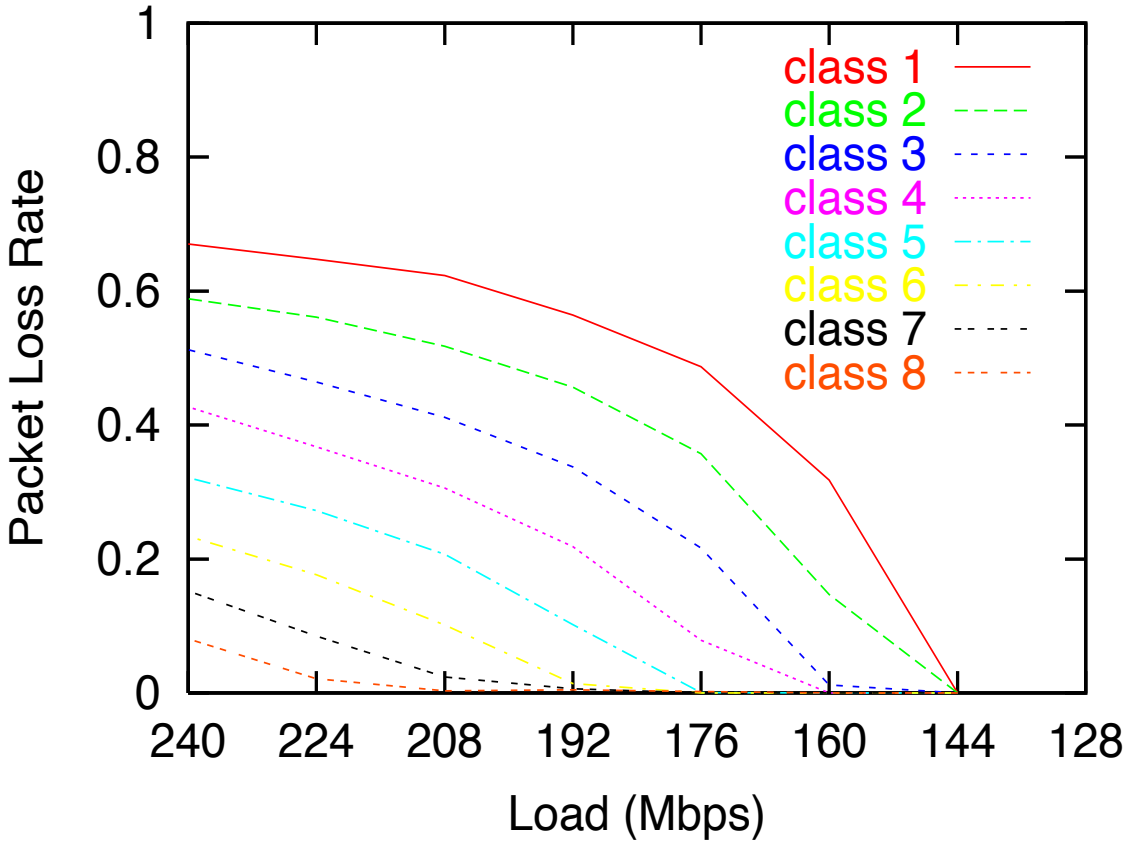
  → RSVP protocol

  → analogous to leasing a line

- IETF DiffServ

  → different types of router behavior

Cisco 7206VXR router: packet loss rate



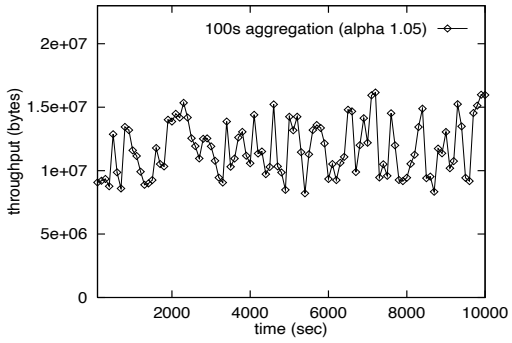
$\longrightarrow$ 8 classes

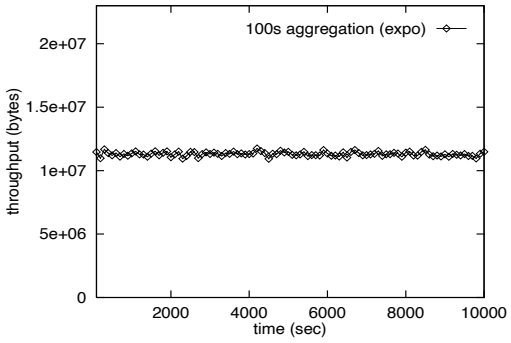
$\longrightarrow$ OC-3 link

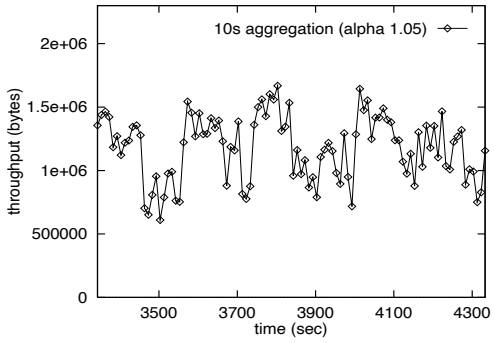
$\longrightarrow$ varying offered load

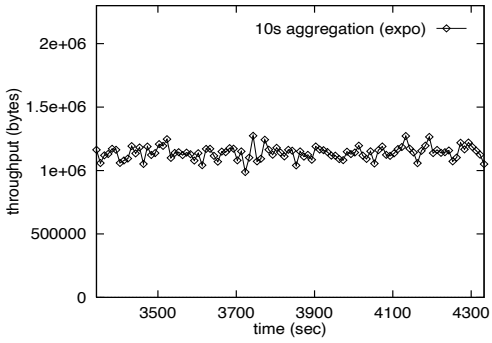
Shape of Internet traffic:

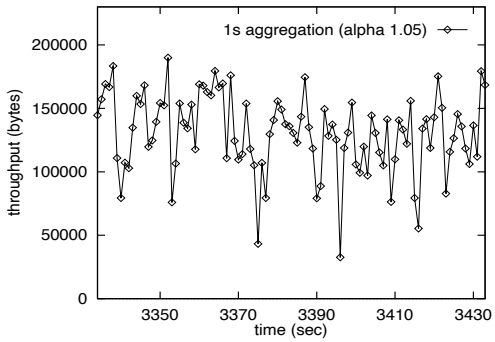
- $\longrightarrow$ what does it look like?
- $\longrightarrow$ doesn't become flat with time aggregation
- $\longrightarrow$ stays bursty!
- $\longrightarrow$ in a peculiar fashion: self-similar or fractal

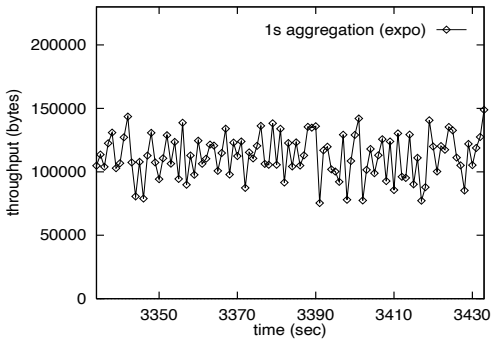

100s aggregation (alpha 1.05)
throughput (bytes)
time (sec)
0
5e+06
1e+07
1.5e+07
2e+07
2000
4000
6000
8000
10000

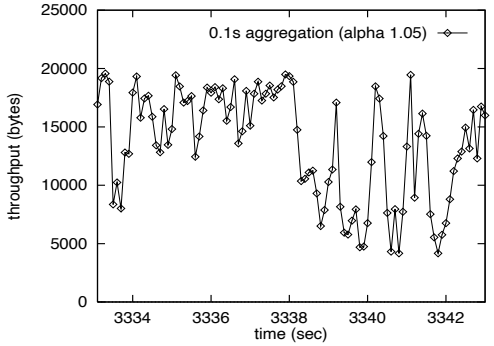

100s aggregation (expo)
throughput (bytes)
time (sec)
0
5e+06
1e+07
1.5e+07
2e+07
2000
4000
6000
8000
10000

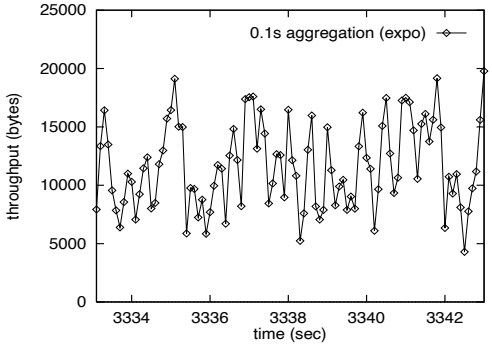

10s aggregation (alpha 1.05)
throughput (bytes)
time (sec)
0
500000
1e+06
1.5e+06
2e+06
3500
3700
3900
4100
4300

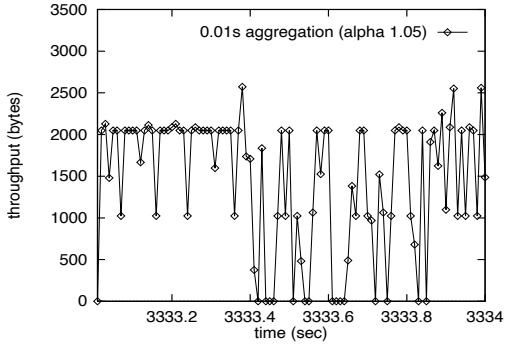

10s aggregation (expo)
throughput (bytes)
time (sec)
0
500000
1e+06
1.5e+06
2e+06
3500
3700
3900
4100
4300

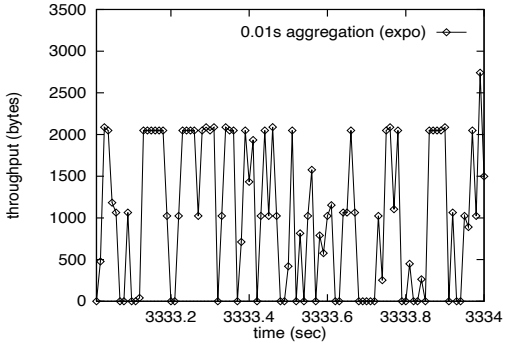

1s aggregation (alpha 1.05)
throughput (bytes)
time (sec)
0
50000
100000
150000
200000
3350
3370
3390
3410
3430


1s aggregation (expo)
throughput (bytes)
time (sec)
0
50000
100000
150000
200000
3350
3370
3390
3410
3430


0.1s aggregation (alpha 1.05)
throughput (bytes)
time (sec)
0
5000
10000
15000
20000
25000
3334
3336
3338
3340
3342


0.1s aggregation (expo)
throughput (bytes)
time (sec)
0
5000
10000
15000
20000
25000
3334
3336
3338
3340
3342


0.01s aggregation (alpha 1.05)
throughput (bytes)
time (sec)
0
500
1000
1500
2000
2500
3000
3500
3333.2
3333.4
3333.6
3333.8
3334

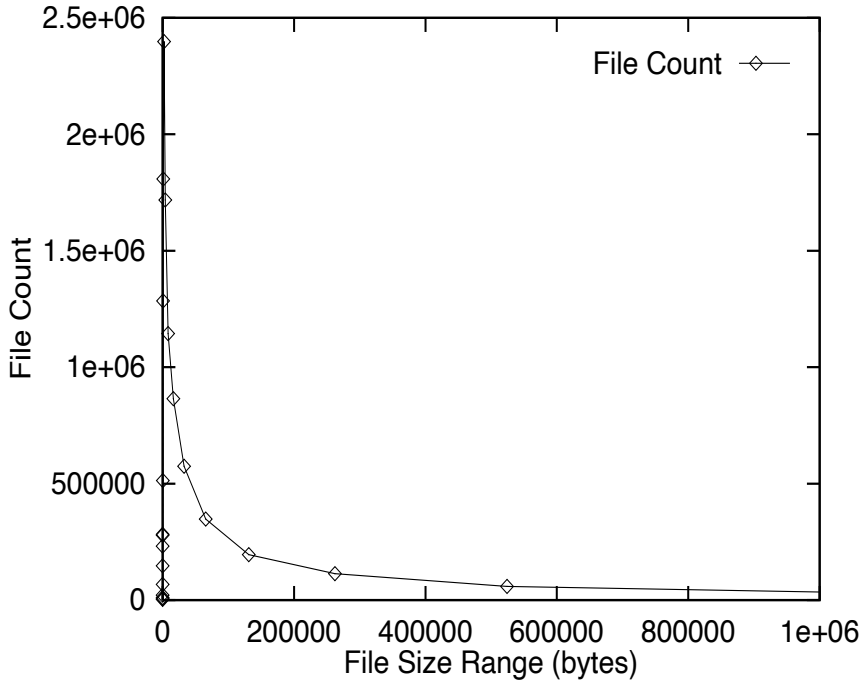

0.01s aggregation (expo)
throughput (bytes)
time (sec)
0
500
1000
1500
2000
2500
3000
3500
3333.2
3333.4
3333.6
3333.8
3334

Consequences:

- $\longrightarrow$ cannot use "flat is good" method anymore
- $\longrightarrow$ intrinsic trade-off between QoS and efficiency
- $\longrightarrow$ bad news for QoS provisioning
- $\longrightarrow$ traffic must be correlated in time (why?)

Tale of elephants and mice:

$\longrightarrow$ UNIX and WWW file systems



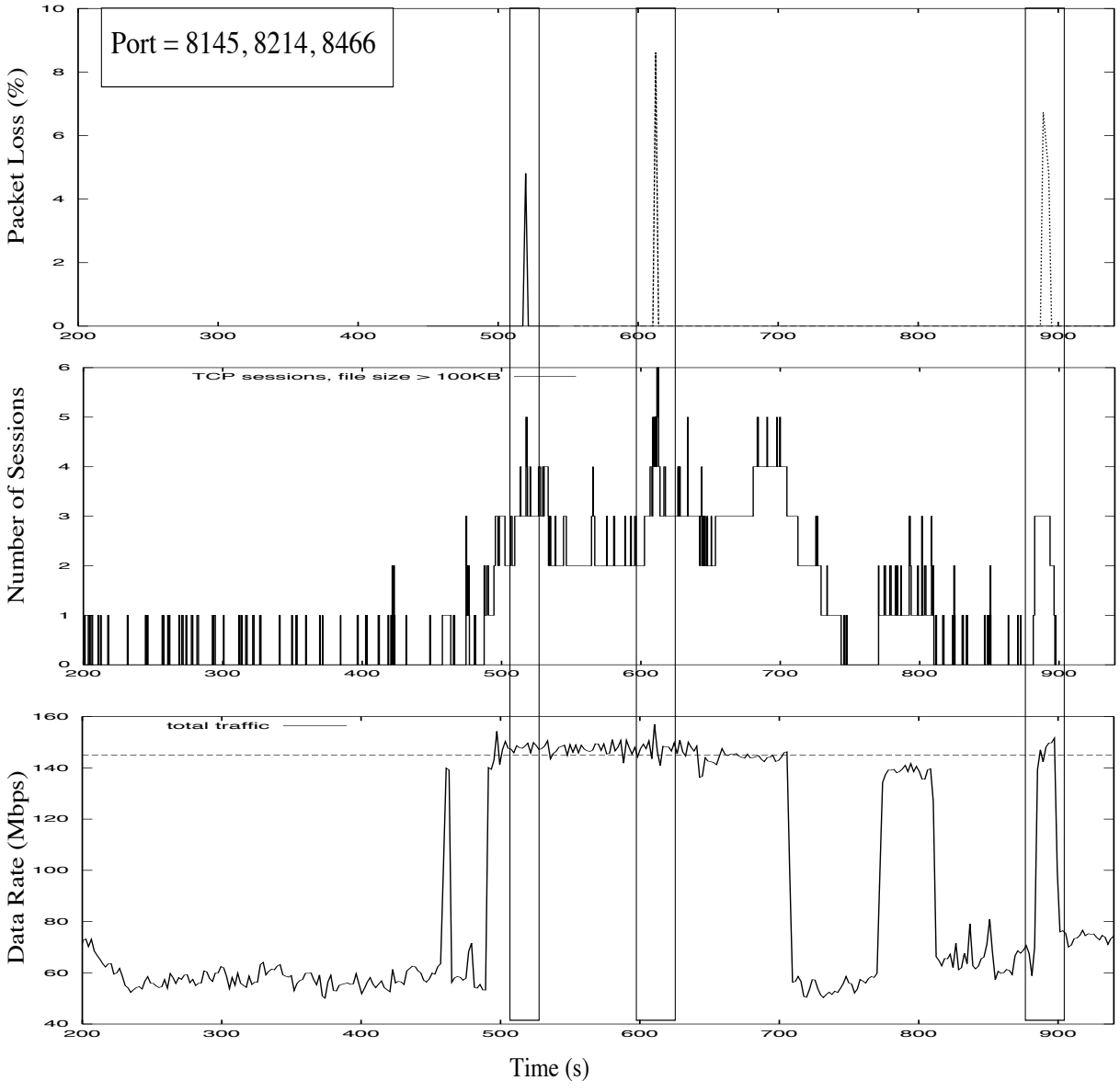
$\longrightarrow$ many small ones (mice)

$\longrightarrow$ a few very large ones (elephants)

$\longrightarrow$ 10% consumes 90% of bandwidth

Elephants in action:

$\longrightarrow$ at backbone router



Can the problem be solved?

$\longrightarrow$ no: as long as elephants and mice holds

Turns out to be a wide-spread phenomenon is sociology, networks, and elsewhere

$\longrightarrow$ size (population) of cities

$\longrightarrow$ popularity

$\longrightarrow$ frequency of words in books

$\longrightarrow$ etc.

In the real world:

$\longrightarrow$ norm: skewed distribution of sizes

$\longrightarrow$ power-law