Internet Traffic and QoS

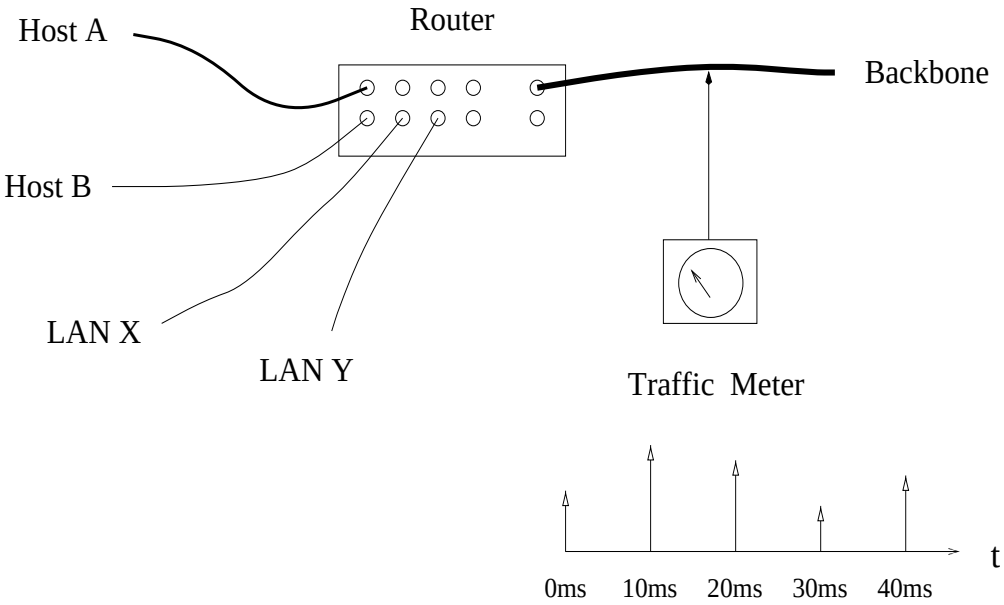
Simplest of all: constant bit rate (CBR)

- $\longrightarrow$ flat is good
- $\longrightarrow$ because predictable
- $\longrightarrow$ e.g., telephone call, real-time MP3 audio

Internet data traffic: variable bit rate (VBR)

- $\longrightarrow$ session arrivals: random
- $\longrightarrow$ exponential interarrival time
- $\longrightarrow$ swimming with the fishes: Poisson process
- $\longrightarrow$ e.g., telephone calls, TCP connections, fast food

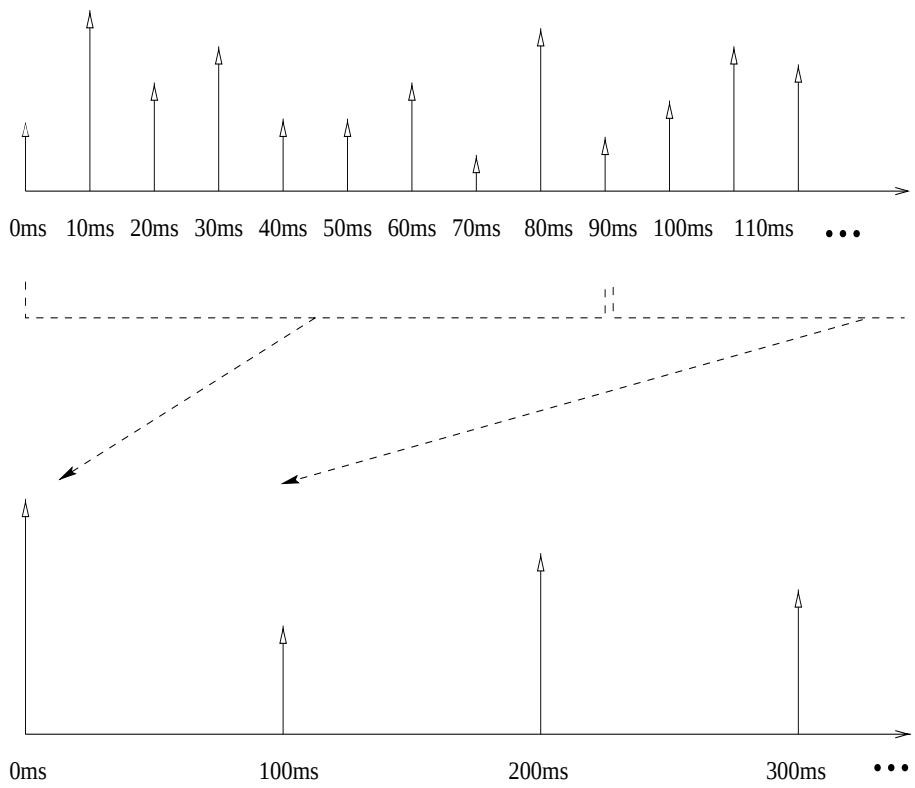
What does Internet traffic look like?



$\longrightarrow$ logging

$\longrightarrow$ traffic time series (at 10ms granularity)

Aggregation (time):

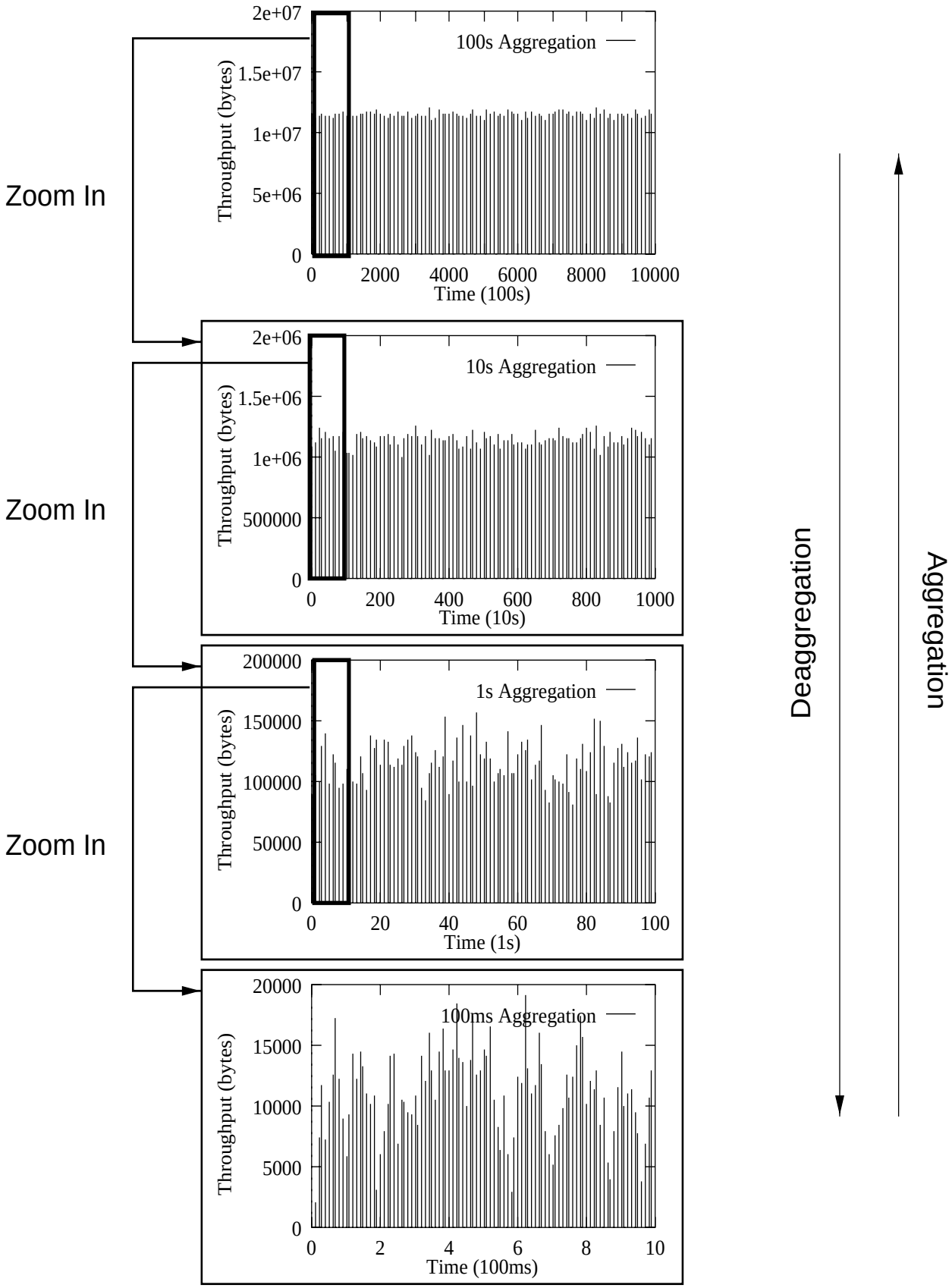


$\longrightarrow$ analogous to computing sample mean

$\longrightarrow$ aggregation over multiple time scales

$\longrightarrow$ what to expect?

$\longrightarrow$ first, telephone traffic


Zoom In
Zoom In
Zoom In
100s Aggregation
10s Aggregation
1s Aggregation
100ms Aggregation
Throughput (bytes)
Time (100s)
Time (10s)
Time (1s)
Time (100ms)
Deaggregation
Aggregation

Aggregated traffic becomes flat

- $\longrightarrow$ "flat is good" rule for QoS provisioning
- $\longrightarrow$ bandwidth dimensioning
- $\longrightarrow$ technically: law of large numbers in action
- $\longrightarrow$ not correlated in time
- $\longrightarrow$ efficient and happy customers

Also aggregation over multiple users

- $\longrightarrow$ called statistical multiplexing
- $\longrightarrow$ assuming independence between different users
- $\longrightarrow$ nice normal distribution shape
- $\longrightarrow$ allows non-peak reservation
- $\longrightarrow$ consider peak-to-mean ratio: less costly

QoS policy:

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

  → guaranteed service

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

  → differentiated service

Internet standards:

- IETF IntServ

  → RSVP protocol

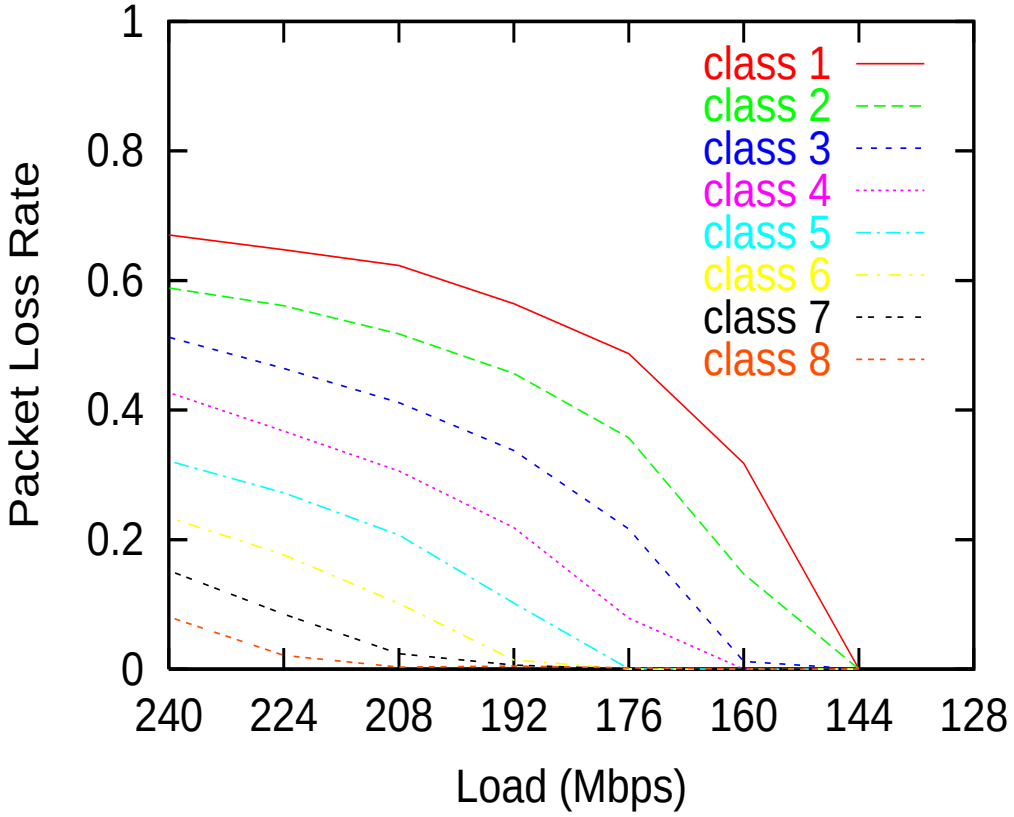
  → analogous to leasing a line

- IETF DiffServ

  → different types of router behavior

  → AF, EF, Cisco's LLQ for VoIP

Cisco 7206VXR router: packet loss rate

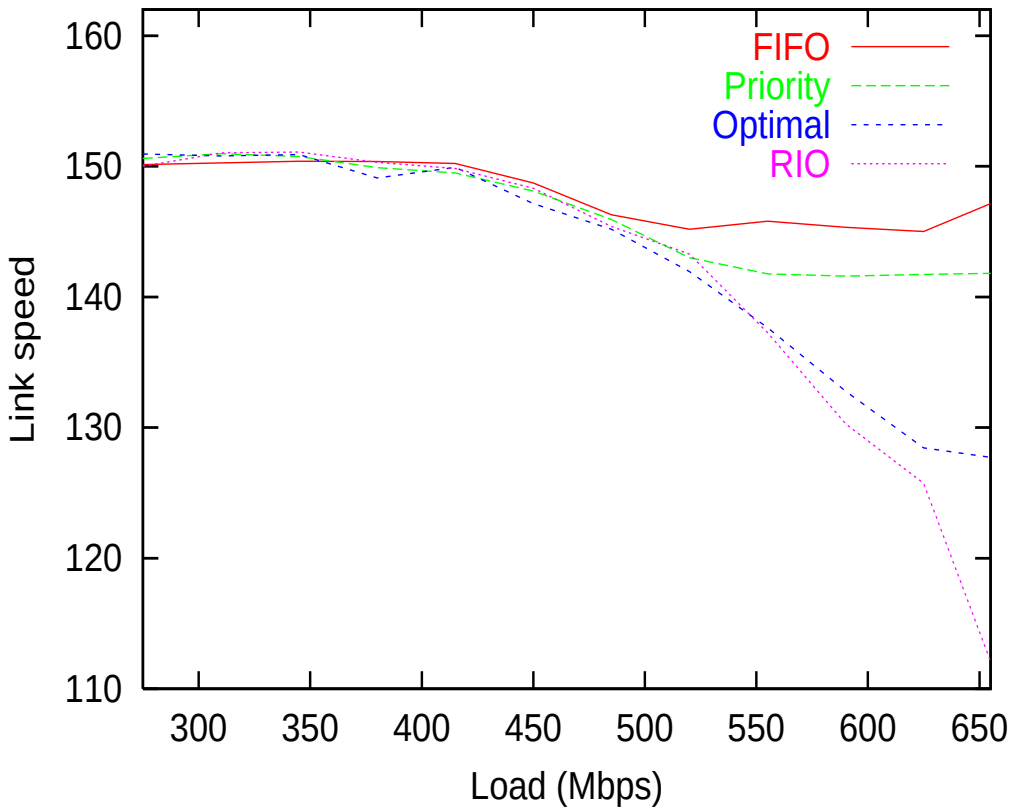


⟶ 8 classes

⟶ OC-3 link

⟶ varying offered load

Cisco 7206VXR router: processing overhead



$\longrightarrow$ offered load

$\longrightarrow$ total throughput

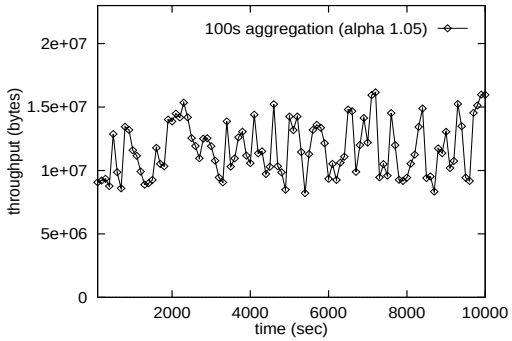
Back to Internet traffic:

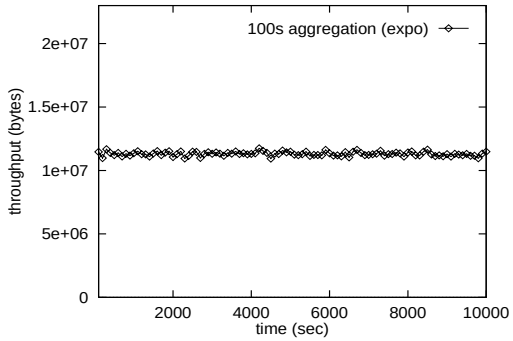
$\longrightarrow$ what does it look like?

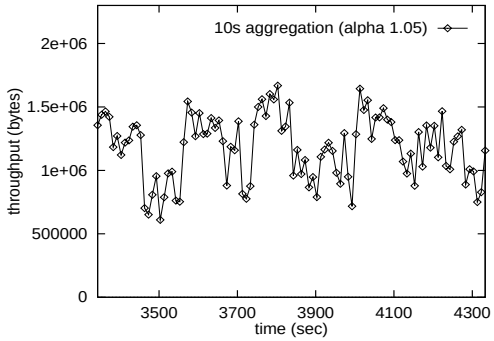
$\longrightarrow$ doesn't become flat with time aggregation

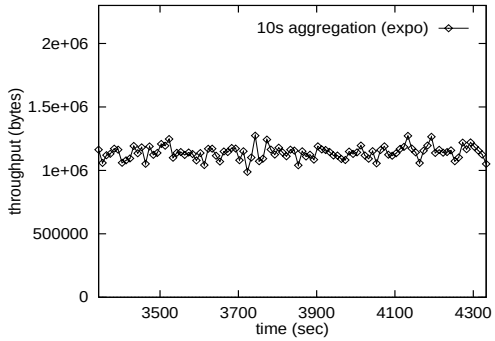
$\longrightarrow$ stays bursty!

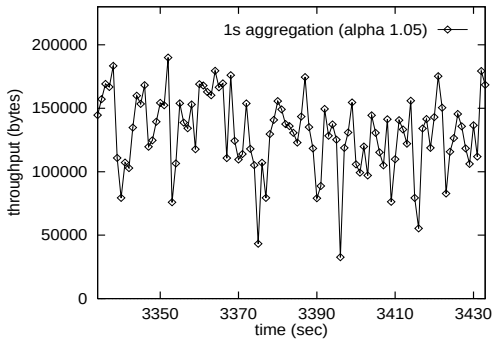
$\longrightarrow$ in a peculiar fashion: self-similar or fractal

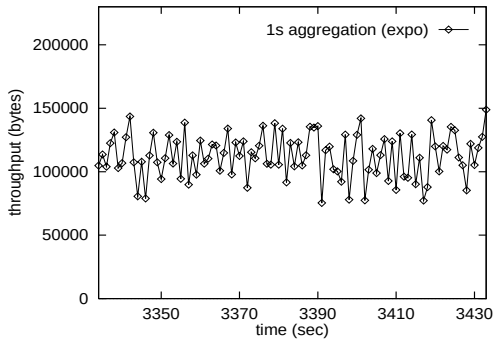

100s aggregation (alpha 1.05)
throughput (bytes)
time (sec)

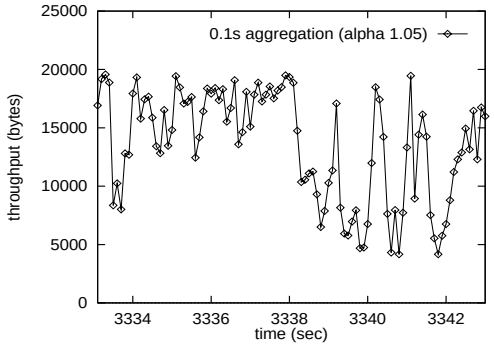

100s aggregation (expo)
throughput (bytes)
time (sec)

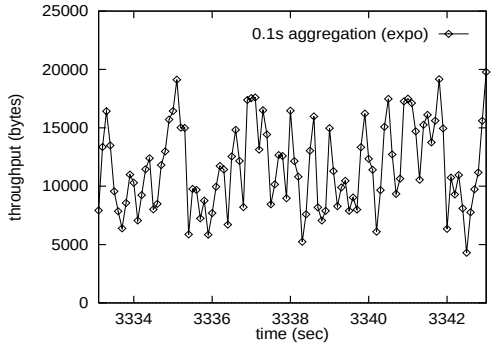

10s aggregation (alpha 1.05)
throughput (bytes)
time (sec)

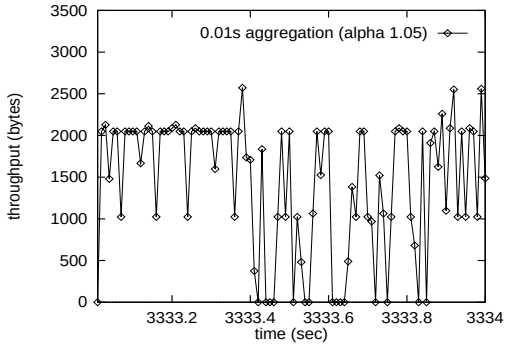

10s aggregation (expo)
throughput (bytes)
time (sec)

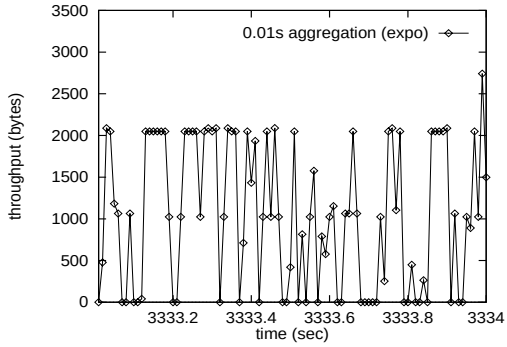

1s aggregation (alpha 1.05)
throughput (bytes)
time (sec)


1s aggregation (expo)
throughput (bytes)
time (sec)


0.1s aggregation (alpha 1.05)
throughput (bytes)
time (sec)


0.1s aggregation (expo)
throughput (bytes)
time (sec)


0.01s aggregation (alpha 1.05)
throughput (bytes)
time (sec)

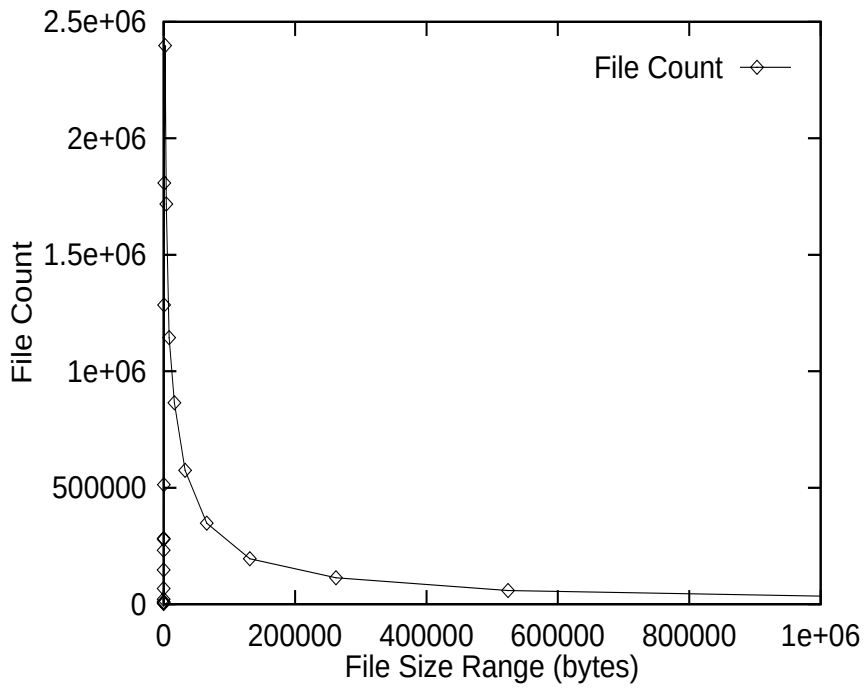

0.01s aggregation (expo)
throughput (bytes)
time (sec)

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:

⟶ UNIX and WWW file systems



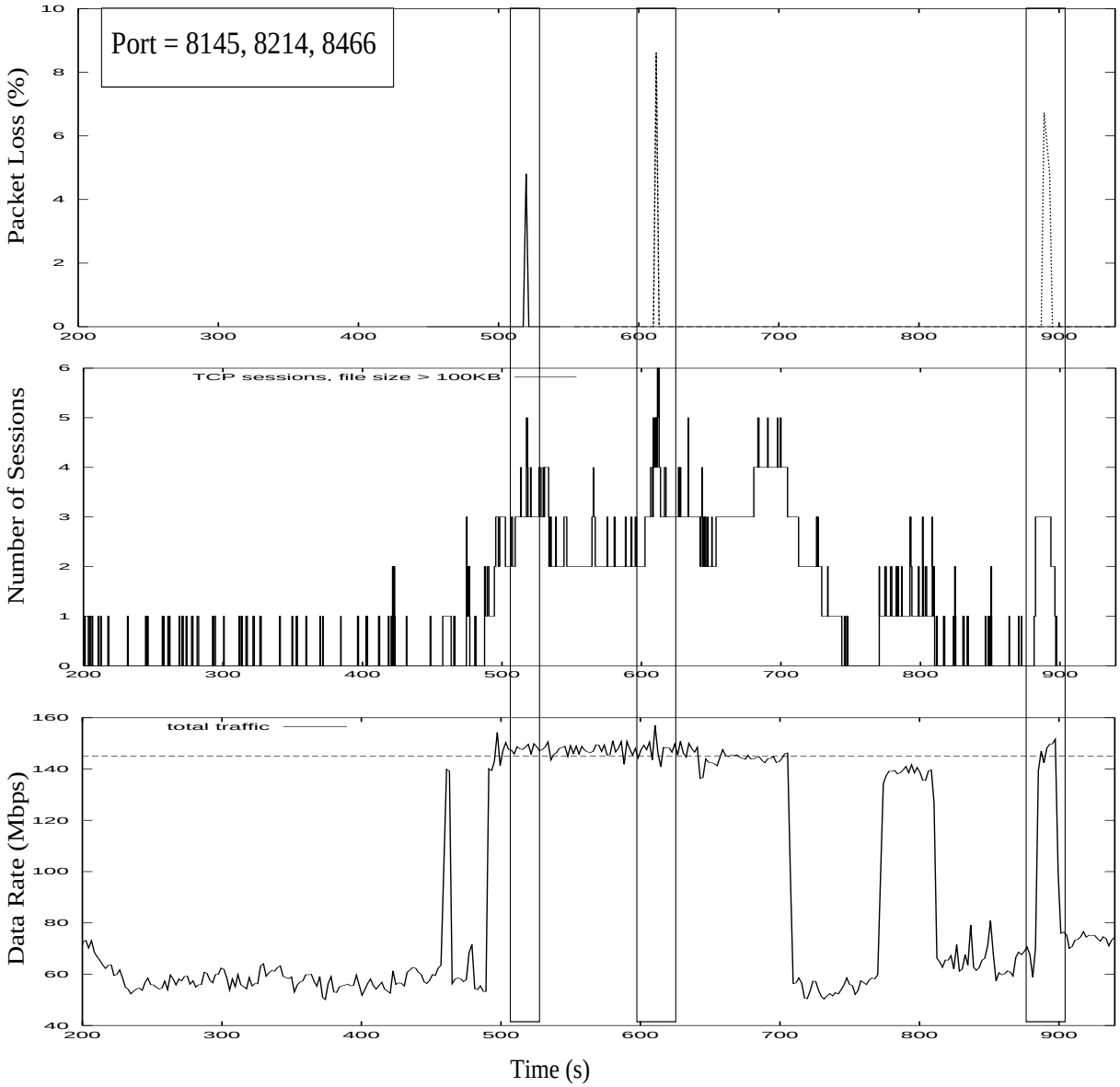
⟶ many small ones (mice)

⟶ a few very large ones (elephants)

⟶ 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