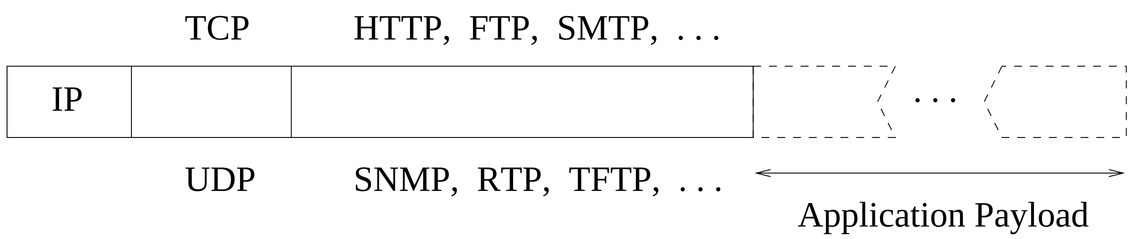
Traffic Characteristics and Its Implications

Traffic and Payload Properties

Classification Dimensions:

- content: data, voice, audio, image, video, signalling/control message, HTML, transaction message, etc.

  $\longrightarrow$ encapsulation & tunneling



- protocol type: TCP, UDP, ICMP, HTTP, TELNET, DNS, NFS, etc.

- representation: ASCII, binary, MPEG, H.261, JPEG, MP3, GSM, etc.
- size
- reliable/unreliable
- real-time/non-real-time
- broadcast, multicast, point-to-point
- others

QoS Requirements:

- average delay, packet loss rate

  $\longrightarrow$ first-order performance measures

- delay variation, packet loss variance: "jitter"

  $\longrightarrow$ second-order performance measures

- bandwidth
- transfer or completion time (e.g., files)
- bounds on first- and second-order QoS measures
- continuous or relaxed forms of QoS requirements
- cost (e.g., pricing)

  $\longrightarrow$ user dependence (vs. application dependence)

Transmission Profile:

- dynamic (in time)
  - constant bit rate (CBR)
  - variable bit rate (VBR)
  - deterministic: periodic/aperiodic
  - stochastic (or probabilistic)
  - connection arrival behavior (meta)

    $\longrightarrow$ CBR, stochastic, periodic, connection

    $\longrightarrow$ burstiness
- static
  - file size
  - object size
  - life time (e.g., connection duration, burst duration)

    $\longrightarrow$ small/large objects

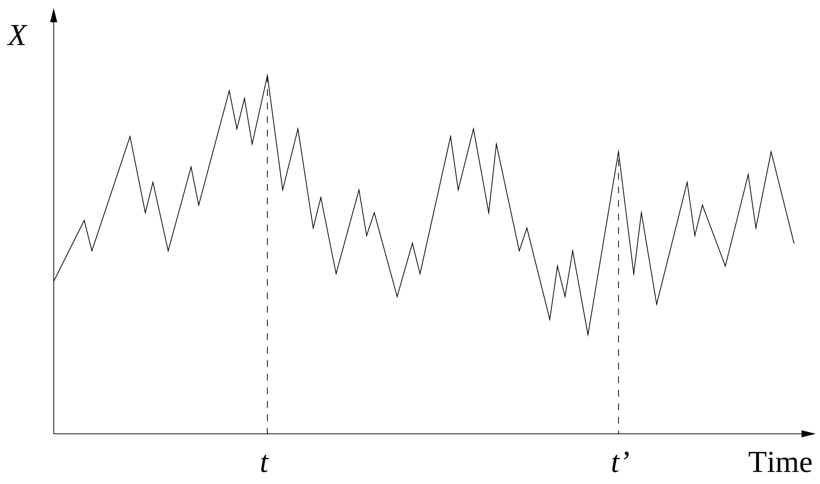
    $\longrightarrow$ short-/long-lived flows
  - related: process and task life times

Dynamic:

Stochastic Transmission Profile

Structure or regularity within bursty or variable traffic profile

Let $X(t)$ be the traffic intensity (or volume) at time t. Consider two time instances $t, t' \geq 0$ with $t \neq t'$ $(t' > t)$.



Since stochastic, $X(t)$ and $X(t')$ are random variables.

Two cases:

- independent

$$\longrightarrow \quad \Pr\{X(t), X(t')\} = \Pr\{X(t)\} \Pr\{X(t')\}$$

- dependent

$$\longrightarrow \quad \text{in what way?}$$

Probability distributions can be characterized by their moments (e.g., first-order, second-order).

- covariance

$$\mathrm{cov}(X(t), X(t')) = \mathrm{E}\{(X(t) - \overline{X(t)})(X(t') - \overline{X(t')})\}$$

- correlation coefficient

$$\mathrm{corr}(X(t), X(t')) = \frac{\mathrm{cov}(X(t), X(t'))}{\mathrm{var}(X(t))\mathrm{var}(X(t'))}$$

$\mathrm{corr}(X(t), X(t')) > 0$ impies information about $X(t)$ yields information about $X(t')$

$\longrightarrow$ positive mutual information

$\longrightarrow$ form of predictability

Typically

- $\mathrm{corr}(X(t), X(t')) \to 0$ as $\Delta t = t' - t \to \infty$
- when not the case for network traffic?

Form of decay as a function of time lag Δt

- "fast" (e.g., exponential)

  $\longrightarrow \approx e^{-\Delta t}$

  $\longrightarrow$ "short-range dependent"

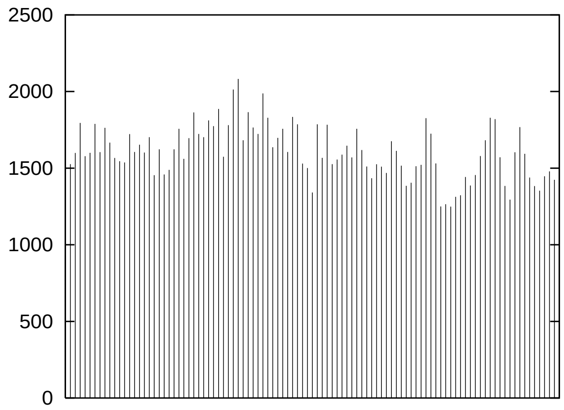
- "slow" (e.g., polynomial)

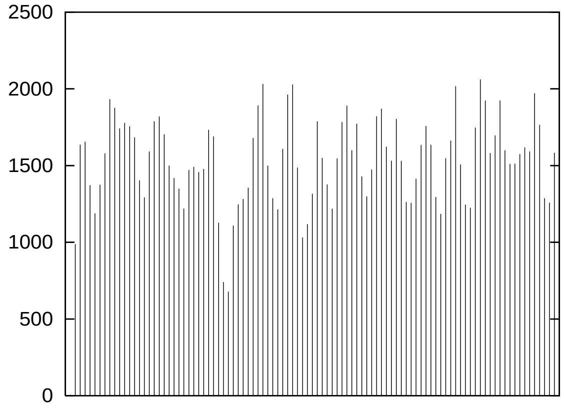
  $\longrightarrow \approx \Delta t^{-\beta}$ with $0 < \beta < 1$

  $\longrightarrow$ "long-range dependent"

What does correlation look like?

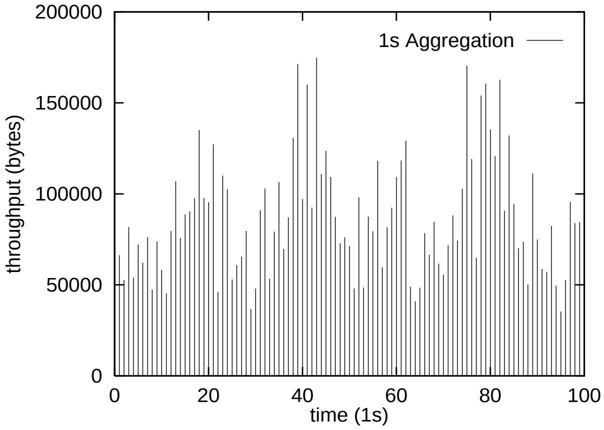
Ex.: Which time series is more correlated.

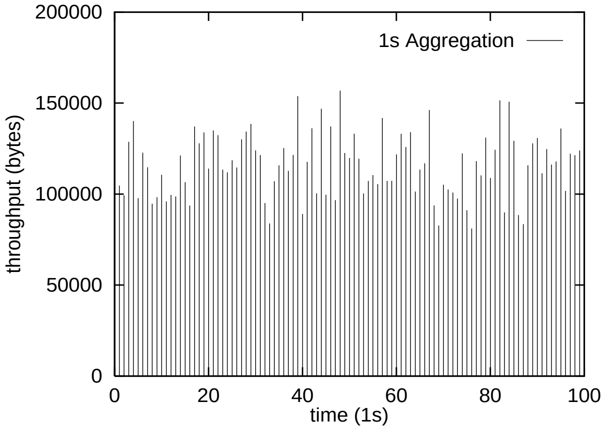




$\longrightarrow$ basis for effective "visual test"?

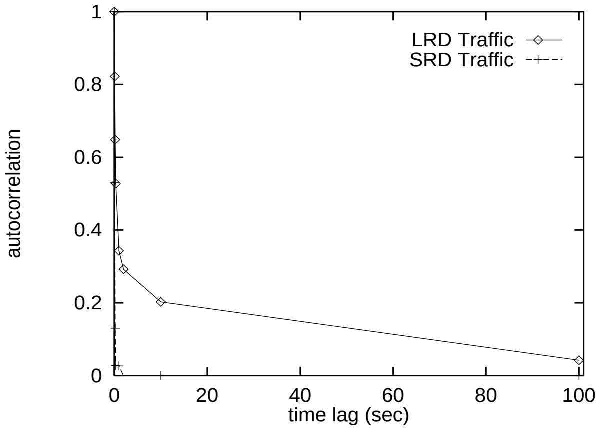
Ex.: Another example of traffic series plot.

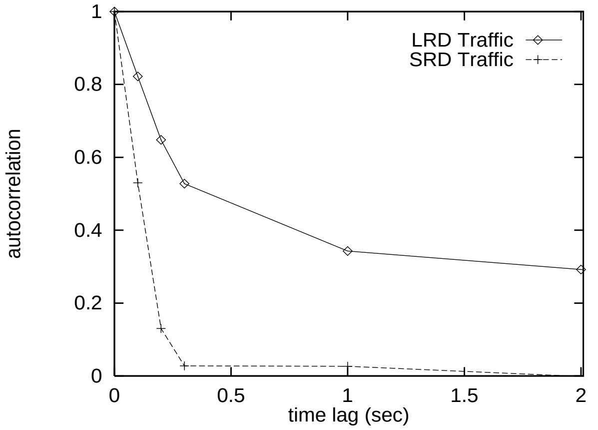




$\longrightarrow$ can be ambiguous

Ex.: Previous self-similar traffic series





Implications

What are the performance implications of dependence or correlation in time?

- performance (or queueing) perspective
  - likelihood of "bad" events increases

    $\longrightarrow$ e.g., prolonged period of packet arrivals

    $\longrightarrow$ assuming positive correlation
  - increase in packet drop rate, queueing delay

    $\longrightarrow$ first-order performance measures
  - amplified variability, i.e., jitter

    $\longrightarrow$ second-order performance measures

    $\longrightarrow$ important for multimedia

. . . "bad news"

- control perspective
  - correlation implies predictability structure

    $\longrightarrow$ e.g., CBR, periodic VBR

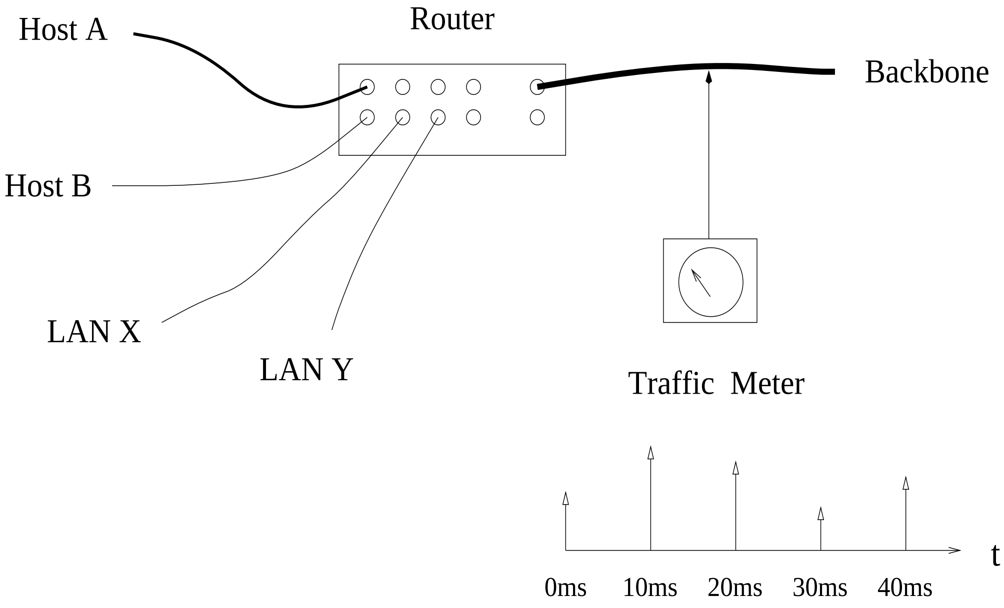
    $\longrightarrow$ even in stochastic VBR
  - traffic control actions may be taken with approximate future network state in mind

    $\longrightarrow$ exploit for traffic control

    $\longrightarrow$ technical challenges

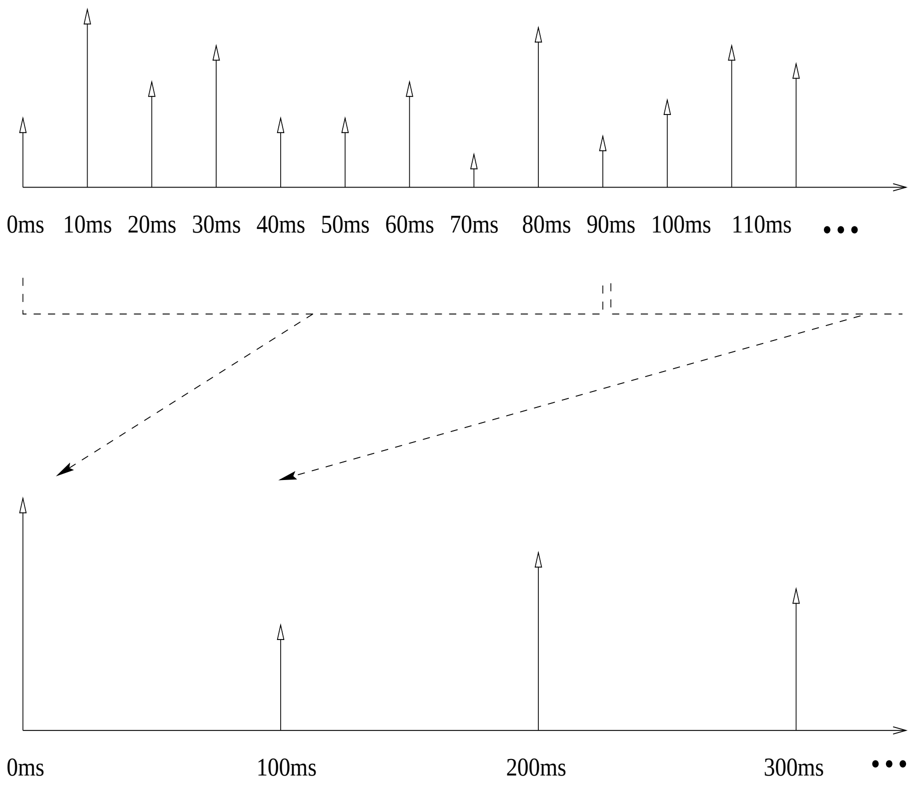
. . . "good news"

Traffic Measurement and Time Aggregation



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

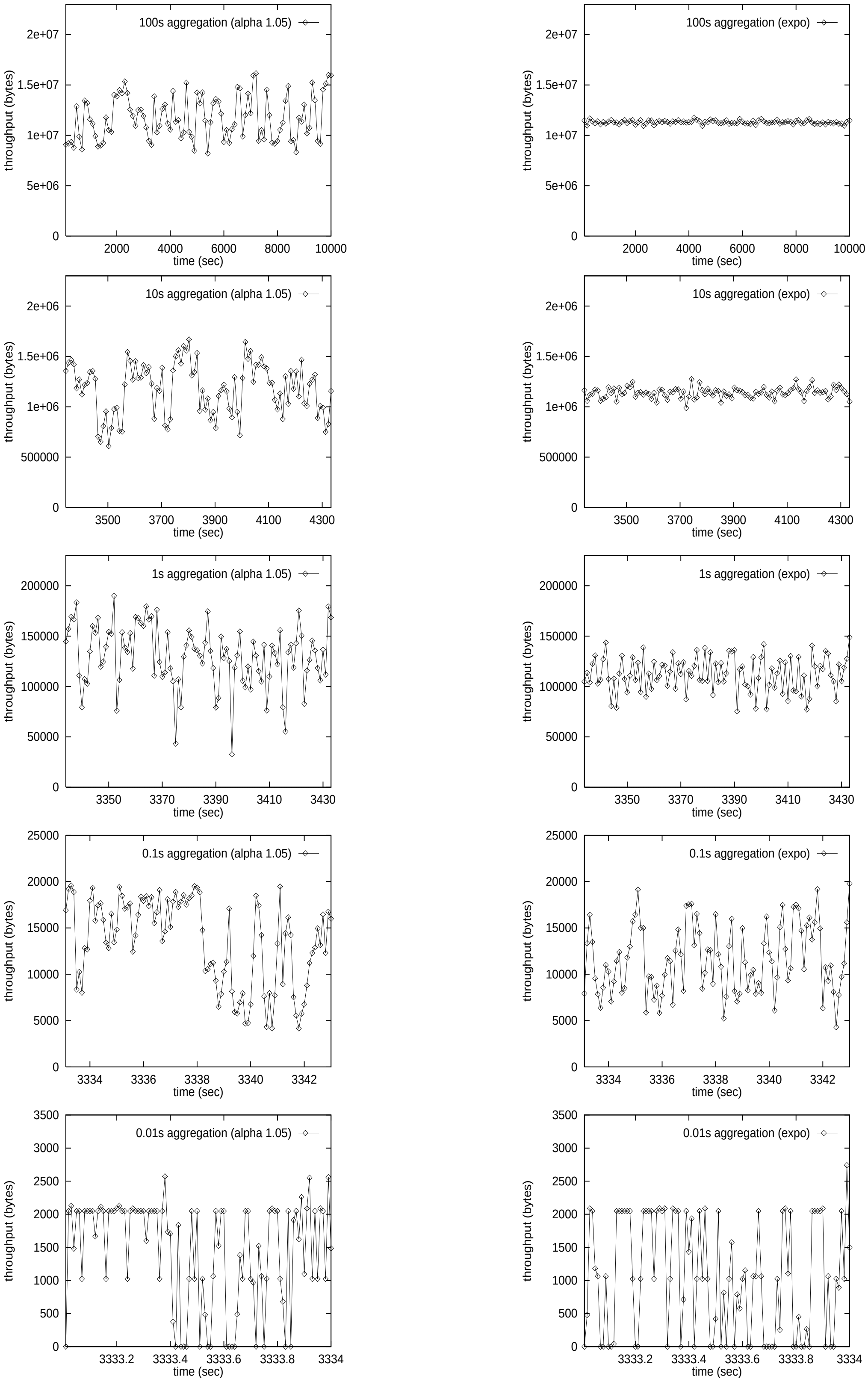
Aggregation (time):



$\longrightarrow$ analogous to computing sample mean

$\longrightarrow$ aggregation over multiple time scales

$\longrightarrow$ what to expect?


100s aggregation (alpha 1.05)
100s aggregation (expo)
10s aggregation (alpha 1.05)
10s aggregation (expo)
1s aggregation (alpha 1.05)
1s aggregation (expo)
0.1s aggregation (alpha 1.05)
0.1s aggregation (expo)
0.01s aggregation (alpha 1.05)
0.01s aggregation (expo)
throughput (bytes)
time (sec)

We observe

- burstiness preserved across time scales five orders of magnitude apart

  $\longrightarrow$ data traffic

  $\longrightarrow$ Poisson traffic smoothes out quickly

- persistent burstiness exists even in packet drop trace

  $\longrightarrow$ prolonged periods of contention & idleness

  $\longrightarrow$ especially bad for multimedia traffic

Law of Large Numbers: The sample mean $\bar{X}_n$ of a sequence of n independent random variables converges to the population mean μ as n increases:

$$\bar{X}_n \to \mu$$

Note difference between Weak LLS and Strong LLS.

$\longrightarrow$ how can it be used for visual test?

By LLN, if it does not flatten out with time aggregation—scaling in time—evidence of dependence.

$\longrightarrow$ converse not necessarily true

$\longrightarrow$ example?

"Burstiness" is preserved across multiple time scales with respect to aggregation

$\longrightarrow$ invariant burstiness structure

$\longrightarrow$ ... "self-similar" (or fractal)

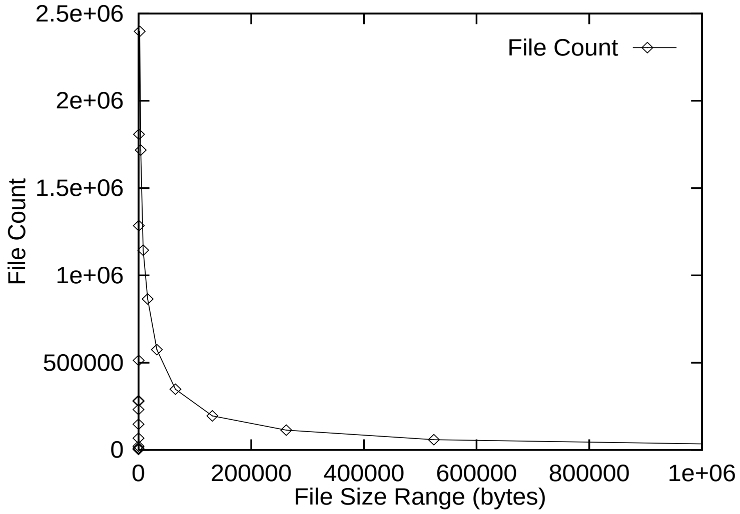
$\longrightarrow$ stochastic self-similarity

Files and their Sizes

Consider a file server—e.g., a UNIX file system—or Web server and the distribution of files w.r.t. their sizes.

$\longrightarrow$ file size distribution

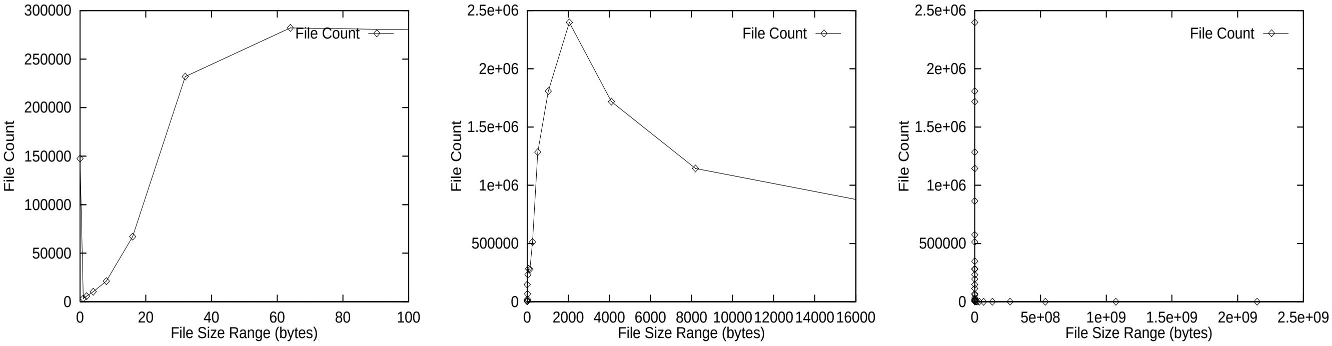
Empirical observation of UNIX file systems[1]:



$\longrightarrow$ many small ones, few large ones . . .

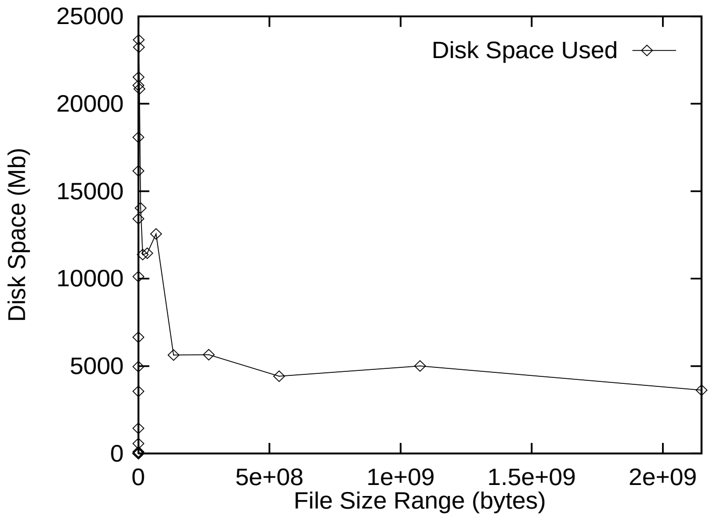
[1] Gordon Irlam, 1993.

How "large" is large and "small" is small?



- $\longrightarrow$ from 0 bytes to $\sim$2 GB
- $\longrightarrow$ mean around $\sim$2 kB
- $\longrightarrow$ variation in small size range

What about disk space used?



$\longrightarrow$ few large files take up disproportionate space

$\longrightarrow$ $\sim$10 % of files consume $\sim$90 % of space

$\longrightarrow$ few large ones exert significant impact

Phenomenon of "few but resource-intensive and dominant" depends on tail of size distribution

$\longrightarrow$ heavy vs. light tail

Consider Pareto distribution. A random variable Z is Pareto if its distribution function

$$\Pr\{Z \leq x\} = 1 - \left(\frac{k}{x}\right)^{\alpha}$$

where

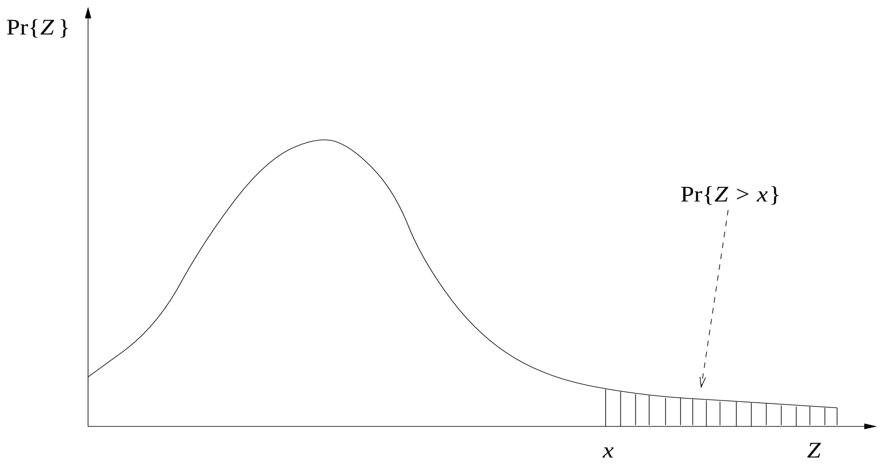
$0 < \alpha < 2$ is the shape parameter;

$k \leq x$ is the location parameter

Of interest:

$\longrightarrow$ shape of tail probability $\Pr\{Z > x\}$

$\longrightarrow$ determined by shape parameter α

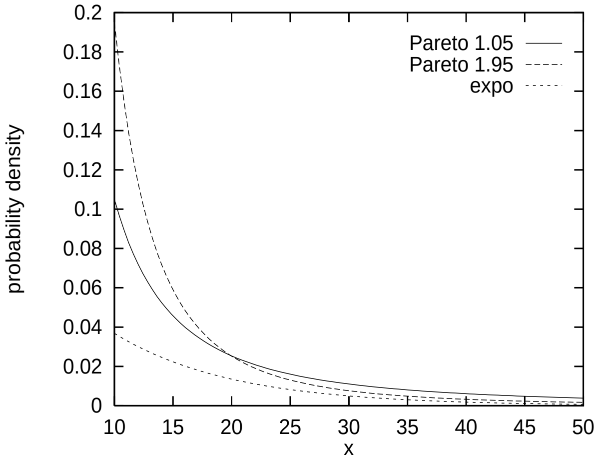


Ex.: For exponential distribution $\lambda e^{-\lambda x}$,

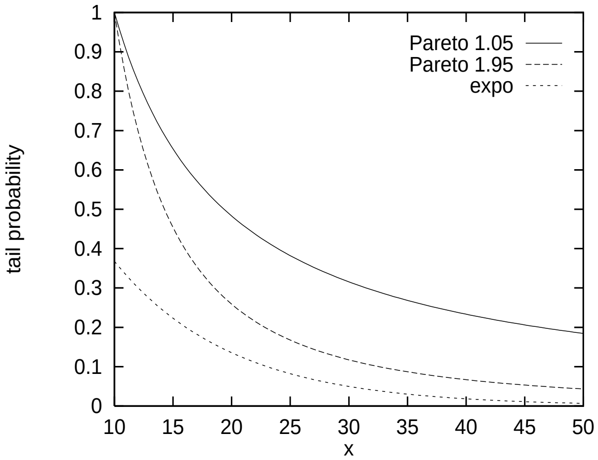
$$\Pr\{Z > x\} = e^{-\lambda x}$$

$\longrightarrow$ total prob. mass in tail exponentially small

Probability densities: Pareto with $\alpha = 1.05, 1.95$ ($k = 10$); exponential distribution with parameter 0.1.



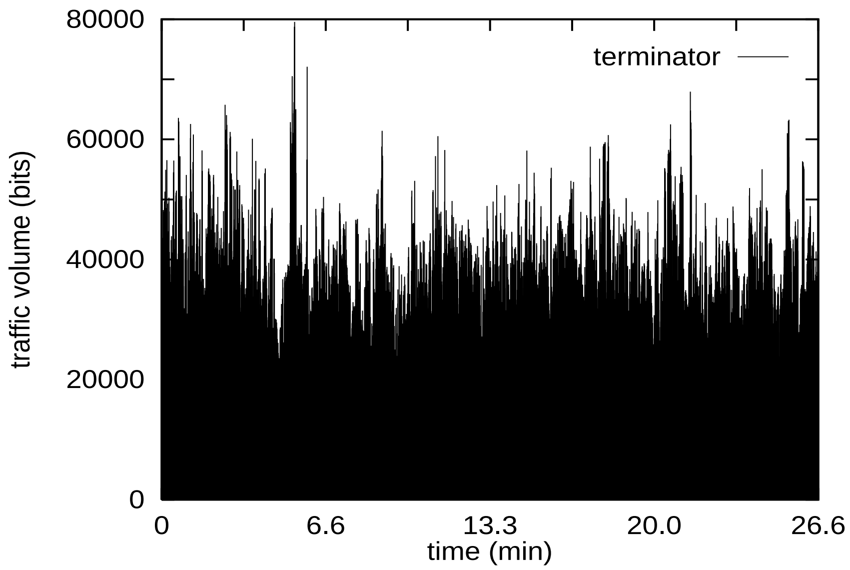
Corresponding tail probabilities.



$\longrightarrow$ probability mass difference in tail pronounced

Real-time VBR source profile

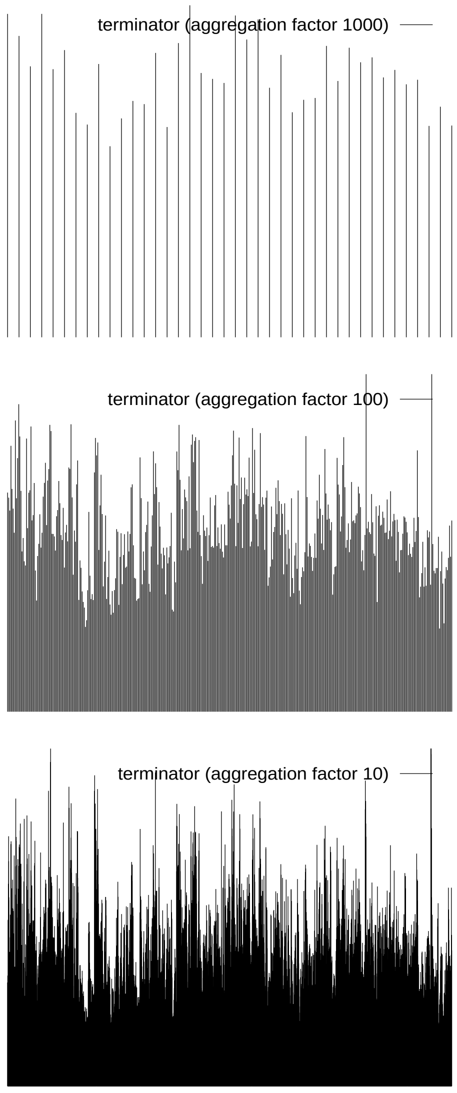
Consider traffic profile of compressed video (e.g., MPEG)



$\longrightarrow$ periodic real-time application

$\longrightarrow$ period $T = 1/f$ (f: frame rate)

Burstiness structure:



$\longrightarrow$ burstiness persists across multiple time scales

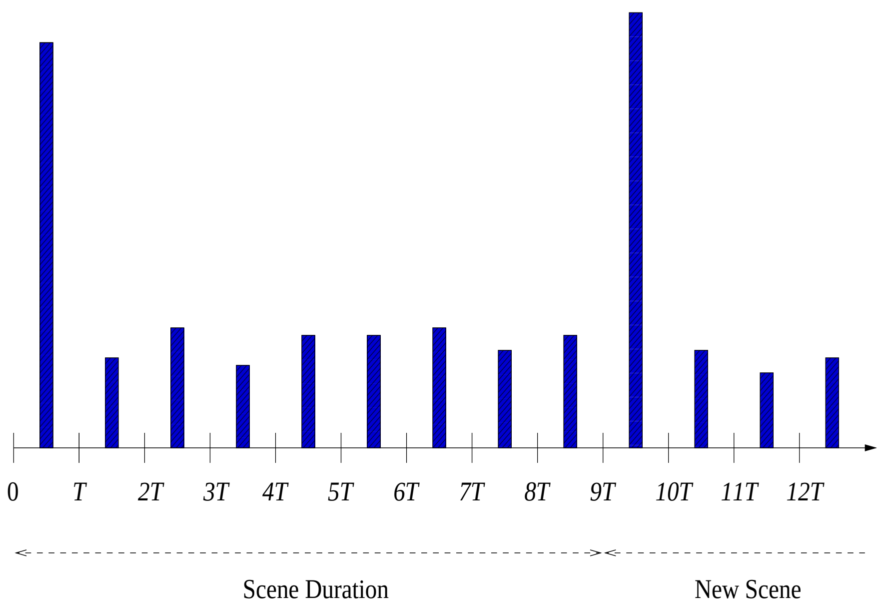
Possible causes:

- Heavy-tailed scene durations

  $\longrightarrow$ facilitates inter-frame compression

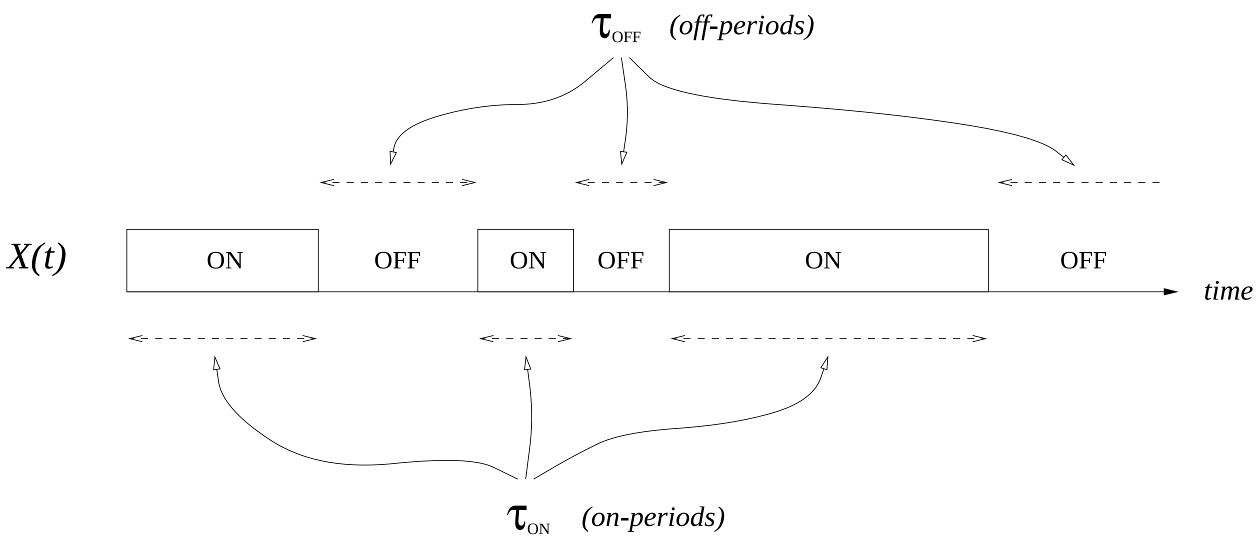
- Repeating GOP pattern

  $\longrightarrow$ e.g., I B B P B B



On/Off Traffic Model

$\longrightarrow$ packet train view



Interested in special case:

- on-periods are i.i.d.; off-periods are i.i.d.
- on-periods are heavy-tailed or light-tailed

  $\longrightarrow$ random variable τ_{on}

- similarly for off-periods

Aggregation of on/off sources:

Ex.: $N = 3$ on/off sources $X_1(t), X_2(t)$, and $X_3(t)$, and their aggregation $S_3(t) = X_1(t) + X_2(t) + X_3(t)$.