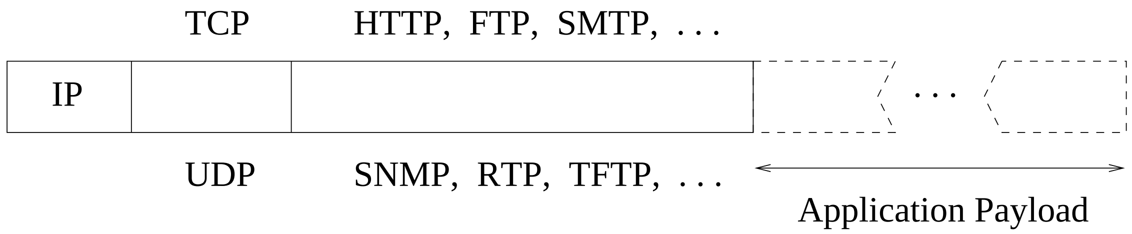
# Traffic Characteristics and Requirements

## 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

## Two Unifying Criteria:

- QoS requirements
- transmission profile

$\longrightarrow$ networking perspective

Ex.: video can be "high quality," "low quality," or in-between within acceptable range

$\longrightarrow$ move away from application specificity

## QoS Requirements:

- average delay, packet loss rate

  $\longrightarrow$ first-order performance measures

- delay variation, packet loss variance: "jitter"

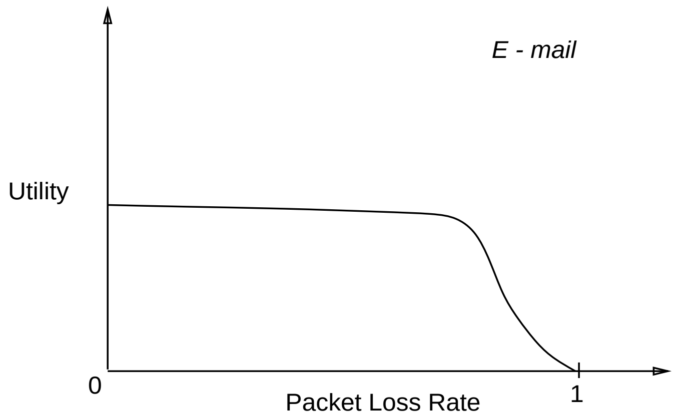
  $\longrightarrow$ second-order performance measures

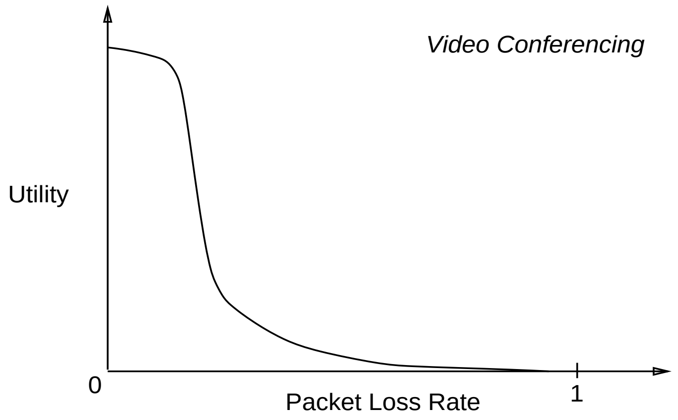
- bandwidth
- transfer or completion time (e.g., files)
- bounds on first- and second-order QoS measures
- continuous or relaxed forms of QoS requirements
- cost (e.g., pricing)

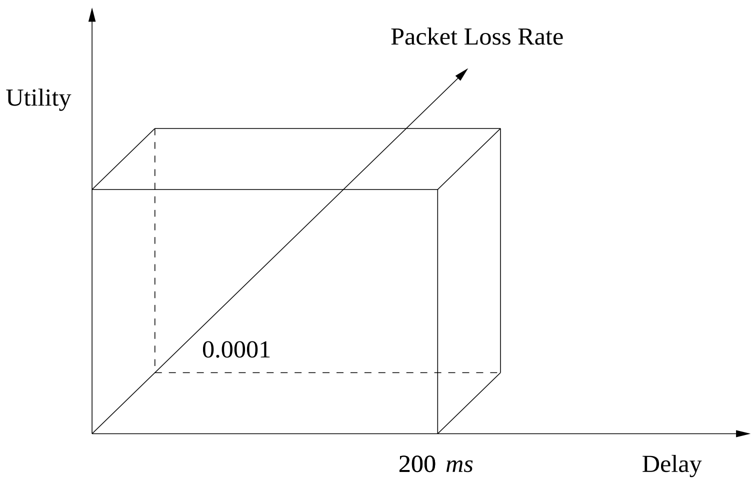
  $\longrightarrow$ user dependence (vs. application dependence)

# Utility Function

Expresses user "satisfaction" as a function of system variables and parameters






More generally,

$$U(\lambda, \boldsymbol{x}, p)$$

where

- $\lambda$: rate (bps) or volume (bytes)
- $\boldsymbol{x} = (x_1, x_2, \ldots, x_s)$: QoS vector (e.g., $x_1$ delay, $x_2$ packet loss rate, etc.)
- $p$: price or cost per unit flow (bps or byte); thus $p\lambda$ is total cost

and utility function $U$ satisfies

- $\partial U / \partial \lambda \geq 0$ "sending faster/more is good"
- $\partial U / \partial \boldsymbol{x} \leq 0$ "better QoS is good"
- $\partial U / \partial p \leq 0$ "less expensive is good"

Sometimes unit of utility can be price ($)

$\longrightarrow$ not the same as variable $p$

## Generalized QoS Requirements

Sometimes, to distinguish, *service requirements*, which can include

- security requirements
- fault-tolerance requirements
- policy requirements
- QoP (quality of presentation) requirements
- others

$\longrightarrow$ use "QoS requirements," clear from context

## 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

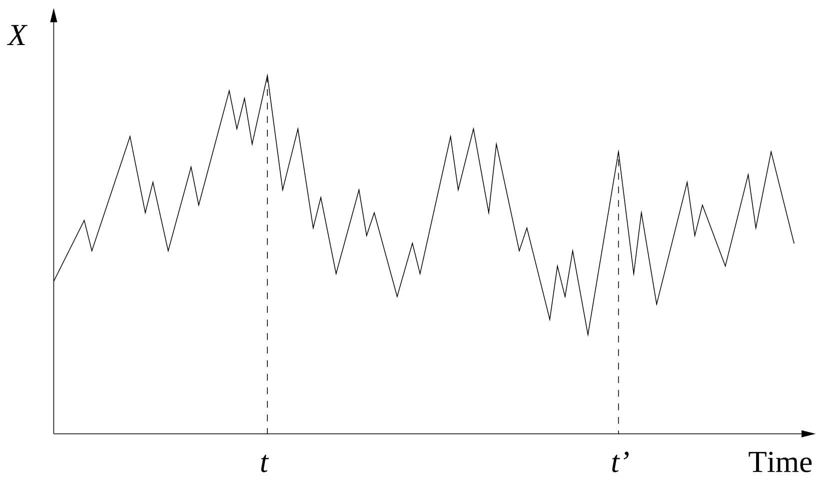
  - related: process and task life times

## Dynamic:

<u>Stochastic Transmission Profile</u>

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

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

- correlation coefficient

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

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

$\longrightarrow$ positive mutual information

$\longrightarrow$ form of predictability

Typically

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

Form of decay as a function of time lag $\Delta t$

- "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"

## 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

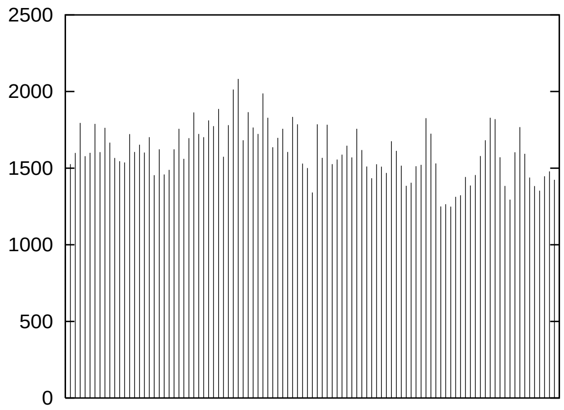
    $\longrightarrow$ exploit for traffic control

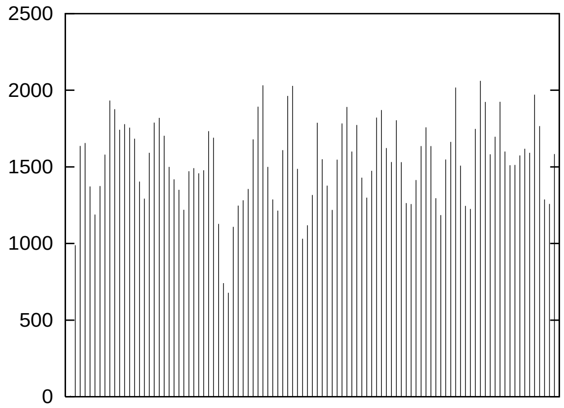
    $\longrightarrow$ technical challenges

. . . "good news"

What does correlation look like?

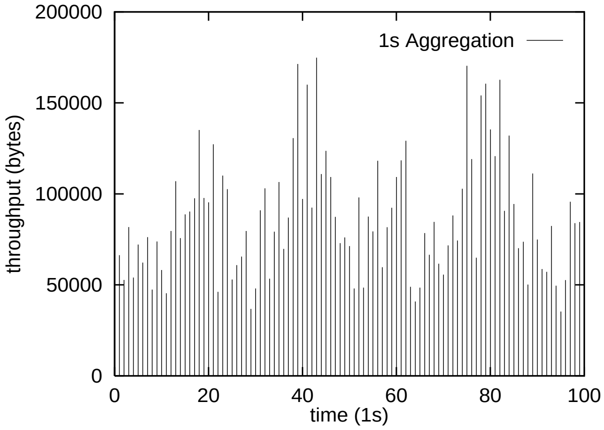
Ex.: Which time series is more correlated.

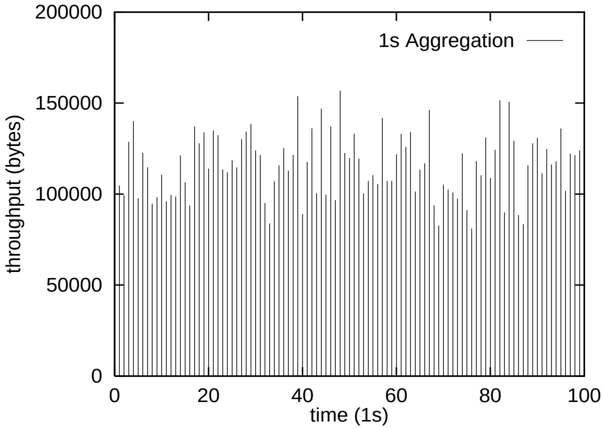



$\longrightarrow$ basis for effective "visual test"?

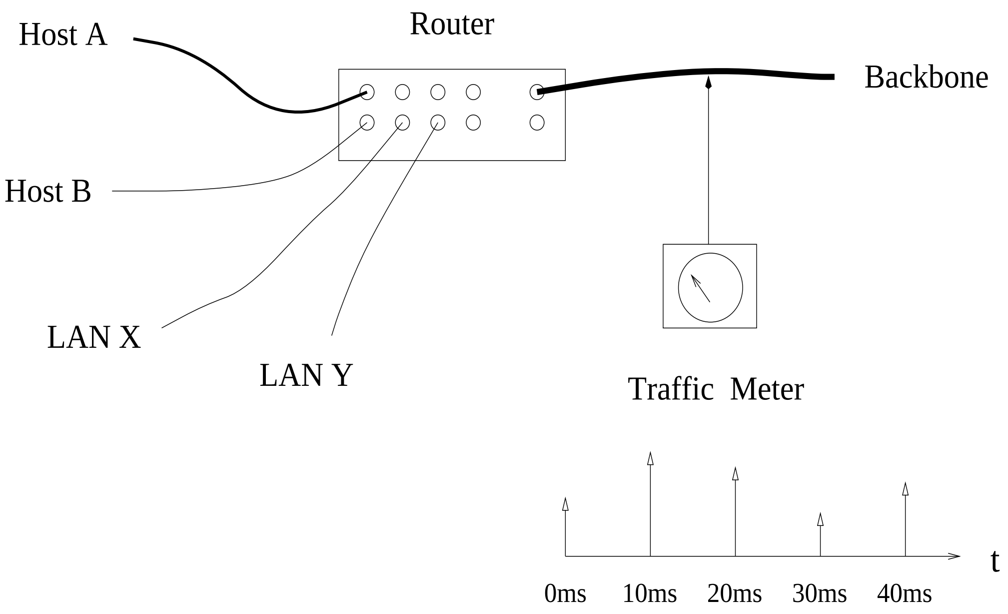
Ex.: Another example of traffic series plot.



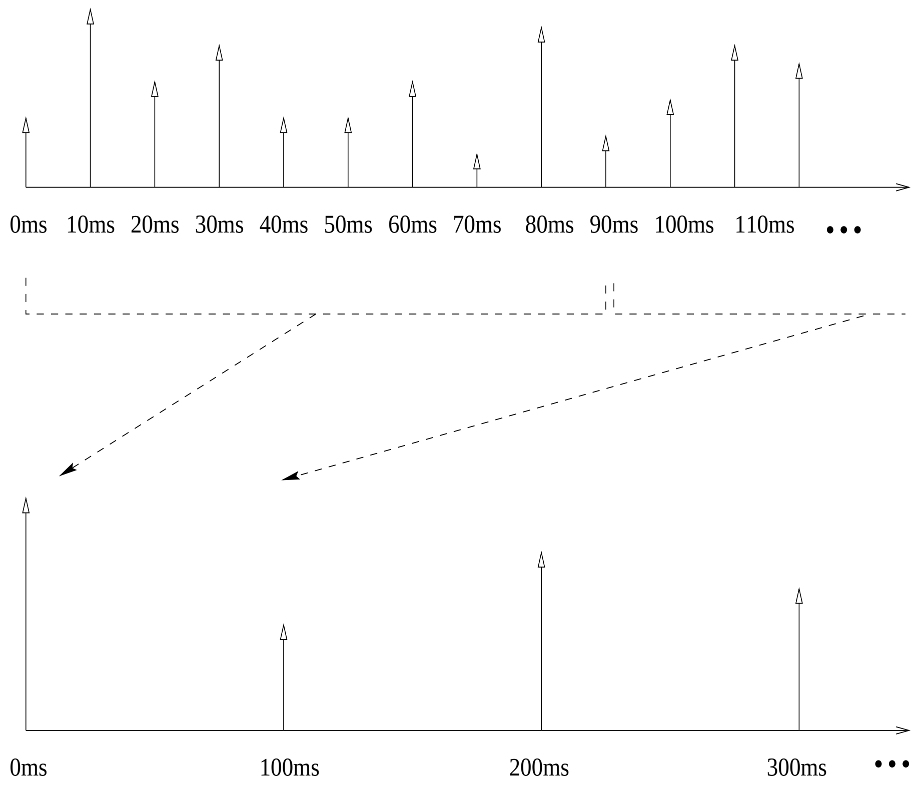

$\longrightarrow$ can be ambiguous

# Traffic Measurement and Time Aggregation




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

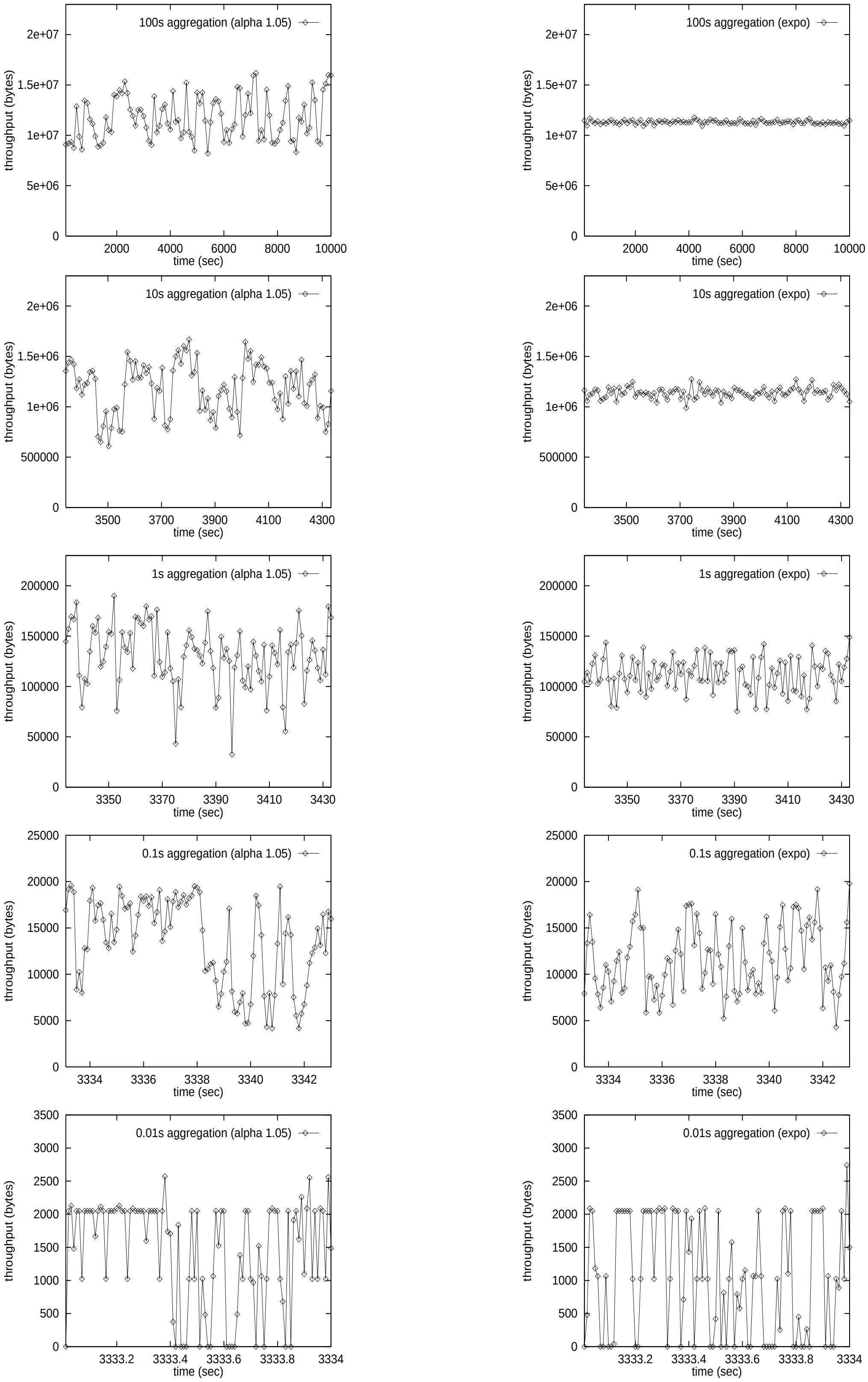
Aggregation (time):

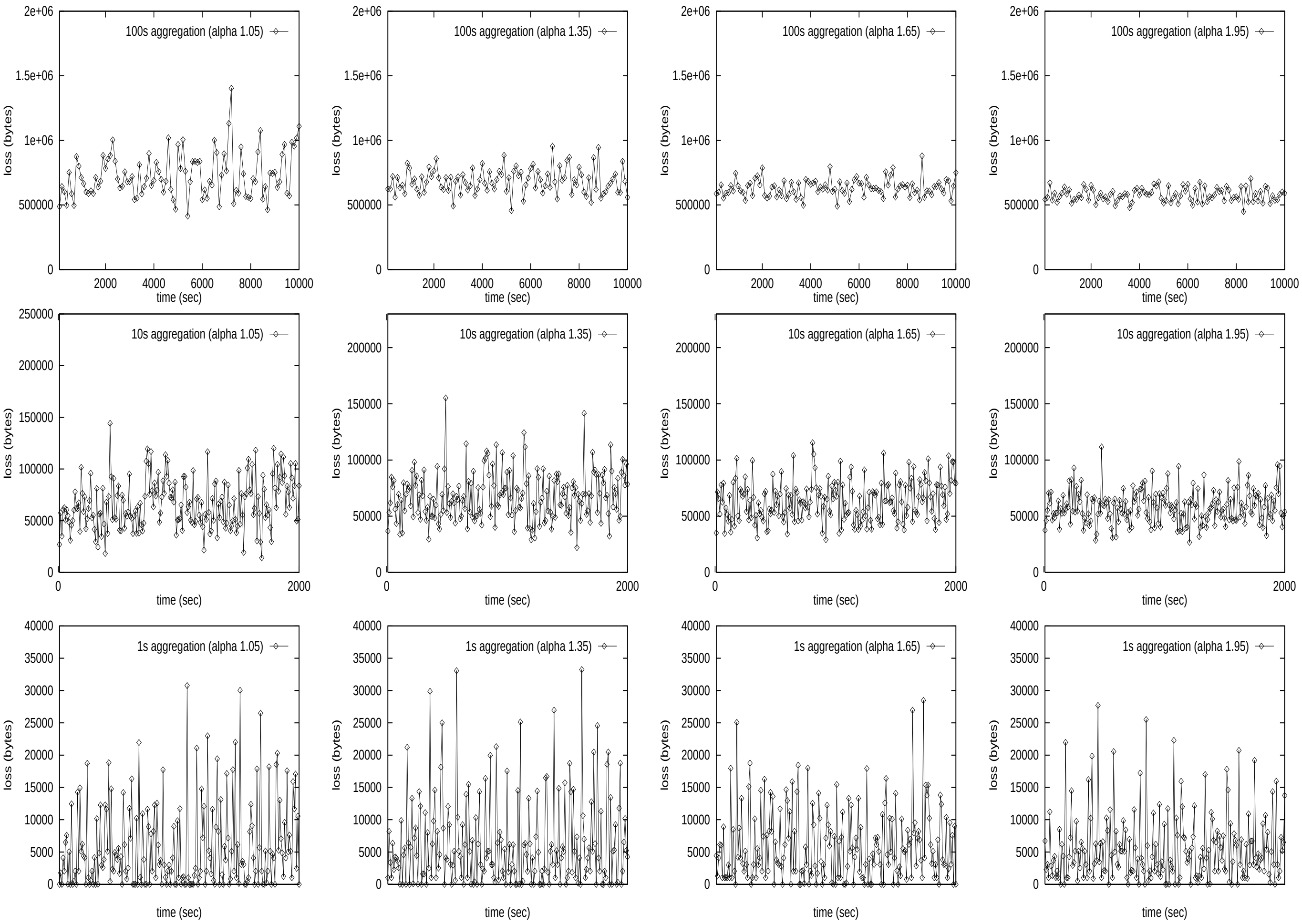



$\longrightarrow$ analogous to computing sample mean

$\longrightarrow$ 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)


100s aggregation (alpha 1.05)
100s aggregation (alpha 1.35)
100s aggregation (alpha 1.65)
100s aggregation (alpha 1.95)
10s aggregation (alpha 1.05)
10s aggregation (alpha 1.35)
10s aggregation (alpha 1.65)
10s aggregation (alpha 1.95)
1s aggregation (alpha 1.05)
1s aggregation (alpha 1.35)
1s aggregation (alpha 1.65)
1s aggregation (alpha 1.95)
loss (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 $\mu$ 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

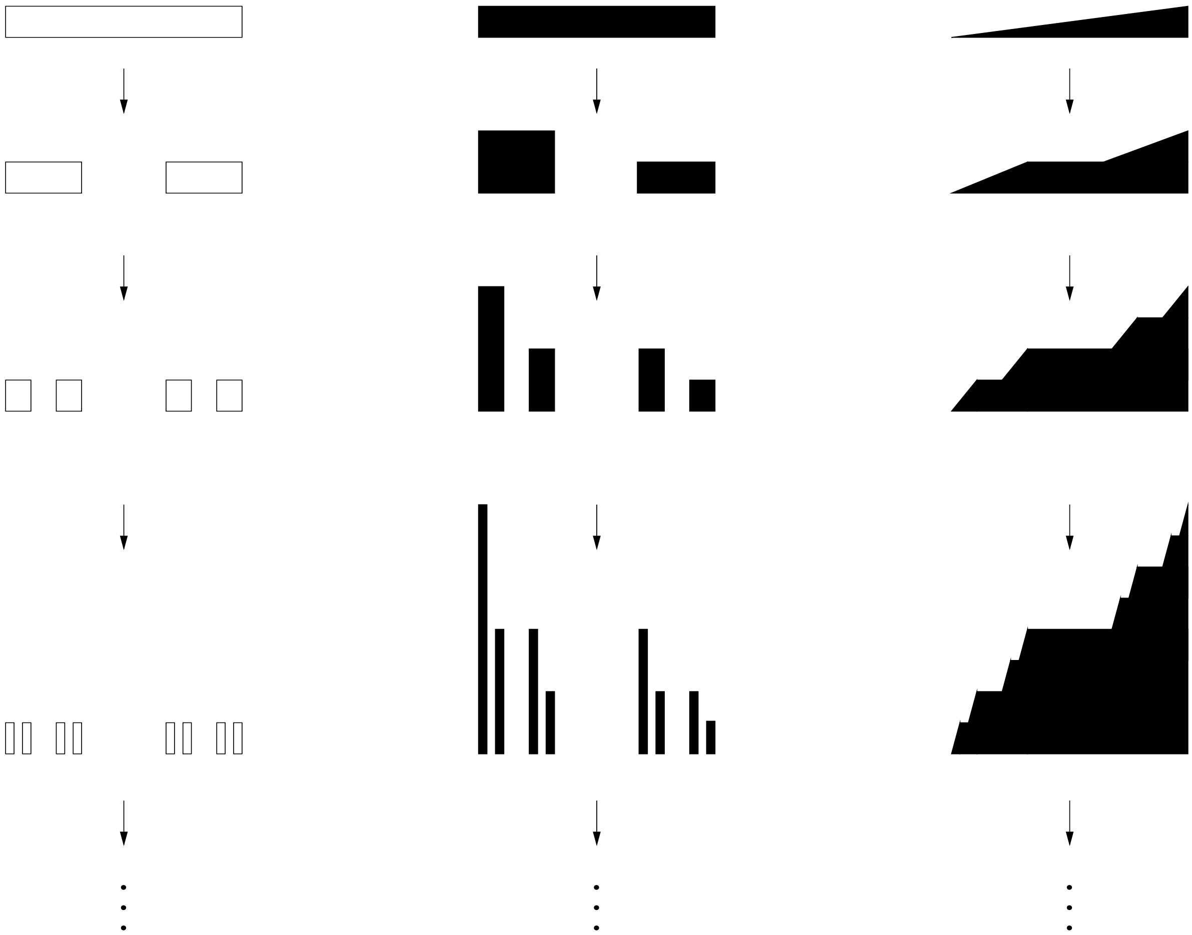
$\longrightarrow$ . . . "self-similar" (or fractal)

$\longrightarrow$ stochastic self-similarity

Deterministic self-similarity:

$\longrightarrow$ "Cantor traffic"

$\longrightarrow$ traffic intensity vs. cumulative traffic



## Quantifying Correlation across Multiple Time Scales

Correlation viewpoint:

Direct Measure for Checking "short-range" vs. "long-range" dependence structure

- autocorrelation function $\mathrm{corr}(X(t), X(t + \Delta t))$

  $\longrightarrow$ fixed $t$, as a function of $\Delta t$ (time lag)

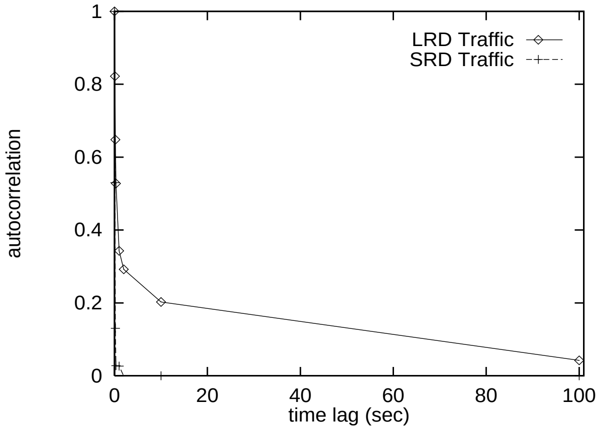
- stationarity (weak)

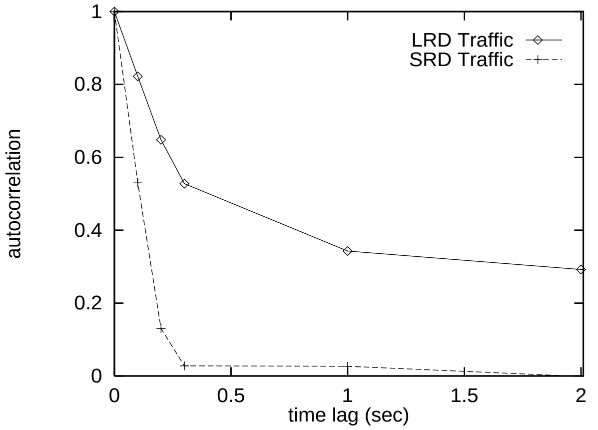
  $\longrightarrow$ $\mathrm{cov}(X(t+T), X(t'+\Delta T)) = \mathrm{cov}(X(t), X(t'))$

  $\longrightarrow$ $\mathrm{corr}(\Delta t) = \mathrm{corr}(X(0), X(\Delta t))$

  $\longrightarrow$ autocorrelation function $r(k)$; discrete time

# Ex.: Previous self-similar traffic series



Technical condition:

$$r(k) \sim c\, k^{-\beta}$$

where $c > 0$ is a constant and $0 < \beta < 1$.

Note:

- $r(k)$ is characterized asymptotically as $k \to \infty$

  $\longrightarrow$ nothing claimed for "small" $k$

- range of $\beta$ implies $\sum_{k=0}^{\infty} r(k) = \infty$

  $\longrightarrow$ definition of long-range dependence

- the closer $\beta$ is to 0, the more correlated the traffic at large time scales

- could use $\beta$ to quantify long-range dependence
- compute $r(k)$, then take logarithm

$$\log r(k) \approx -\beta \log k$$

  to estimate $\beta$ using linear regression

- $\beta$ is related to Hurst parameter $H$ (introduced next)

$$\beta = 2(1 - H)$$

Self-similar burstiness viewpoint:

Time aggregation of $X(t)$ at level $m$ means

$$X^{(m)}(i) = \frac{1}{m} \sum_{t=m(i-1)+1}^{mi} X(t).$$

Since $X(t)$ are random variables, $X^{(m)}(i)$ in time series is analogous to computing the sample mean.

The visual phenomenon of "burstiness preservation" corresponds to

$$\mathrm{var}(X^{(m)}(i)) \approx \mathrm{var}(X(t))$$

for a range of time scales $m$.

If the $X(t)$'s were *independent*, then

$$\text{var}(X^{(m)}(i)) = \sigma^2 m^{-1}$$

where $\sigma^2$ is assumed to be common variance (for simplicity) of the $X(t)$'s.

$\longrightarrow$ elementary fact

Consider rewriting expression with parameter $H$ as

$$\sigma^2 m^{-2(1-H)}$$

where $1/2 \leq H < 1$.

If $H = 1/2$, then we have previous expression $\sigma^2/m$.

$\longrightarrow$ $\sigma^2$ decays at rate $m^{-1}$

If $1/2 < H < 1$, then rate of decay is slower.

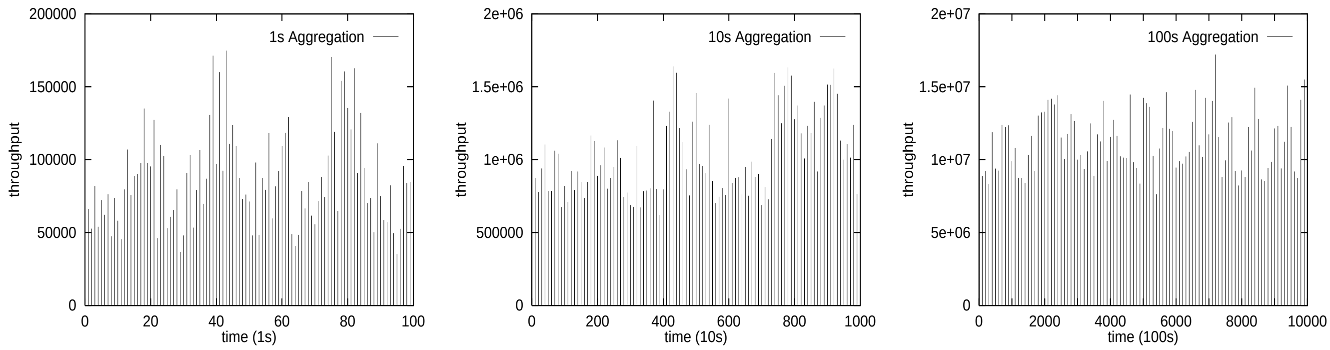
$\longrightarrow$ $m^{-\beta}$ where $0 < \beta < 1$

Thus, if $H \approx 1$, then can expect

$$\mathrm{var}(X^{(m)}(i)) \approx \mathrm{var}(X(t))$$

$\longrightarrow$ burstiness dies out very slowly w.r.t. scaling

$\longrightarrow$ consistent with visual observation



Recall definition of long-range dependence (LRD) with respect to autocorrelation decay $k^{-\beta}$

$\longrightarrow$ $H$, $\beta$ are related by $H = 1 - \beta/2$

$\longrightarrow$ connection between self-similarity & LRD

In a "canonical" model of self-similar traffic—*second-order self-similarity*—the two viewpoints turn out to be equivalent.

- self-similar burstiness implies LRD
- LRD implies self-similar burstiness

$\longrightarrow$ why?

**Static**: Network traffic is influenced by what is being sent, i.e., *source structure*.

Examples:

- files (or objects): stored data, image, video, audio, programs

  $\longrightarrow$ TCP (HTTP, FTP, TELNET)

- real-time streams: MPEG video/audio, voice, interactive games, e-commerce

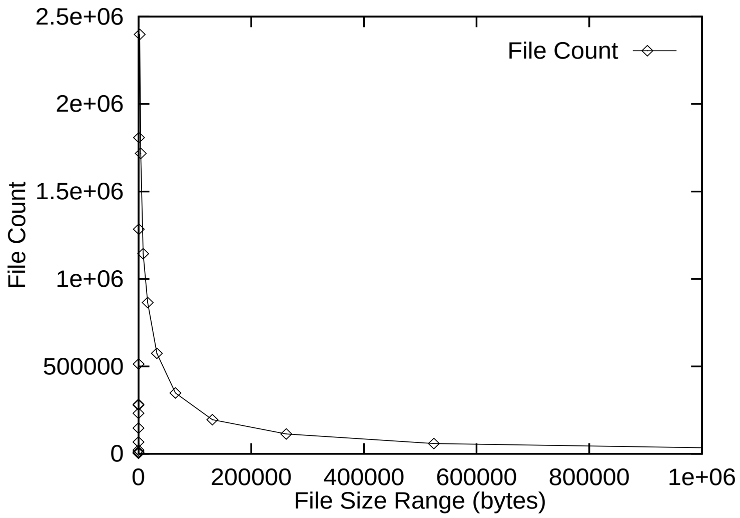
  $\longrightarrow$ UDP (RTP, AFEC, NFS/RPC)

## 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

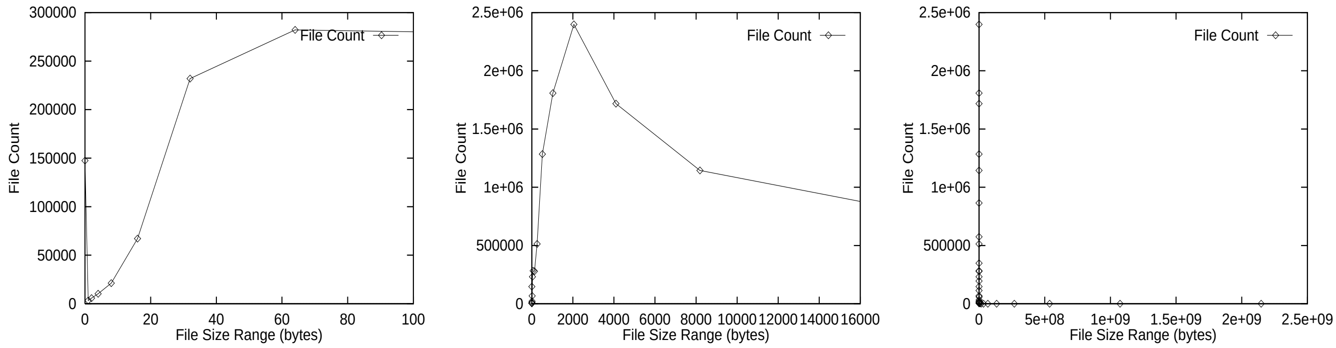
Empirical observation of UNIX file systems[1]:




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

[1] Gordon Irlam, 1993.

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

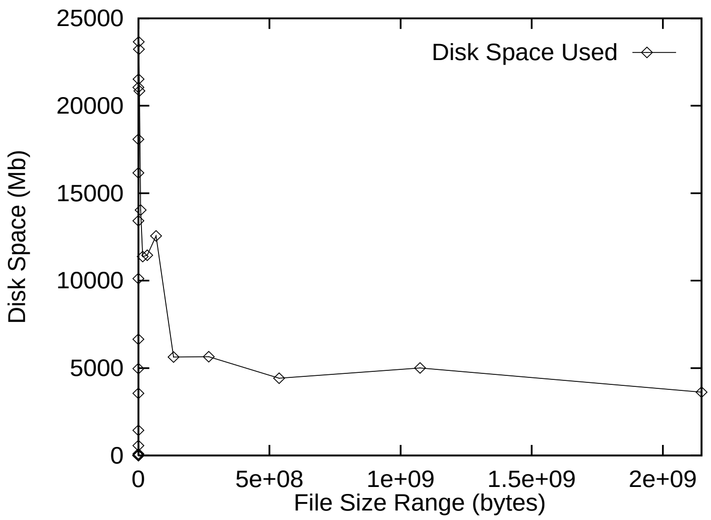



→ from 0 bytes to ∼2 GB

→ mean around ∼2 kB

→ 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

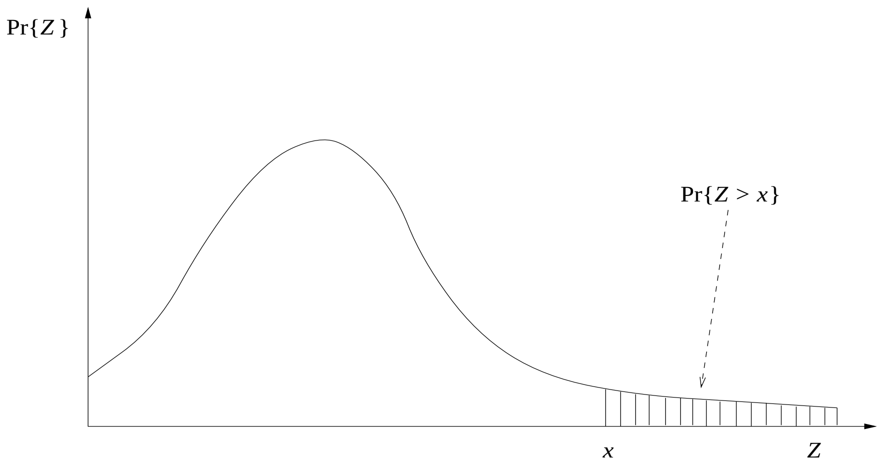
$0 < \alpha < 2$ is the shape parameter;

$k \leq x$ is the location parameter

Of interest:

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

$\longrightarrow$ determined by shape parameter $\alpha$

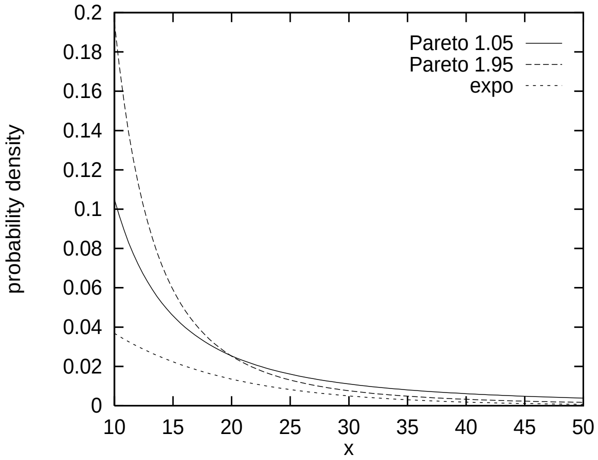



Ex.: For exponential distribution $\lambda e^{-\lambda x}$,

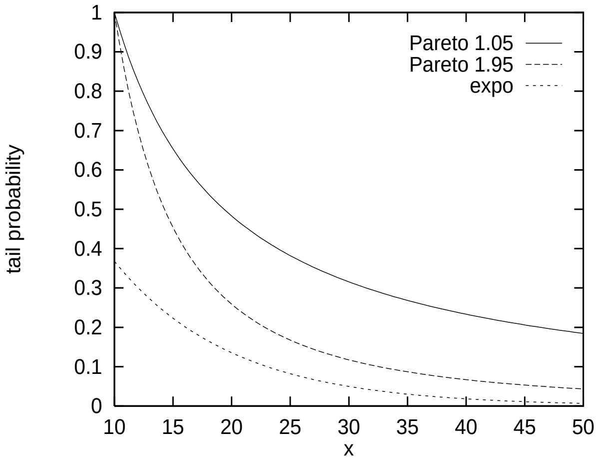
$$\Pr\{Z > x\} = e^{-\lambda x}$$

$\longrightarrow$ 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

Heavy-tailed distribution:

A random variable $Z$ is heavy-tailed if its tail probability

$$\Pr\{Z > x\} \sim c\,x^{-\alpha}, \qquad x \to \infty$$

where $0 < \alpha < 2$ is shape parameter, and $c > 0$ constant.

Note:

- only care about tail when $Z$ (i.e., $x$) is large
- $\mathrm{var}(Z)$ is unbounded for $0 < \alpha < 2$

  $\longrightarrow$ distinguishing mark (extreme variability)
- $\mathrm{E}(Z)$ is unbounded when $0 < \alpha < 1$

  $\longrightarrow$ of interest $1 < \alpha < 2$

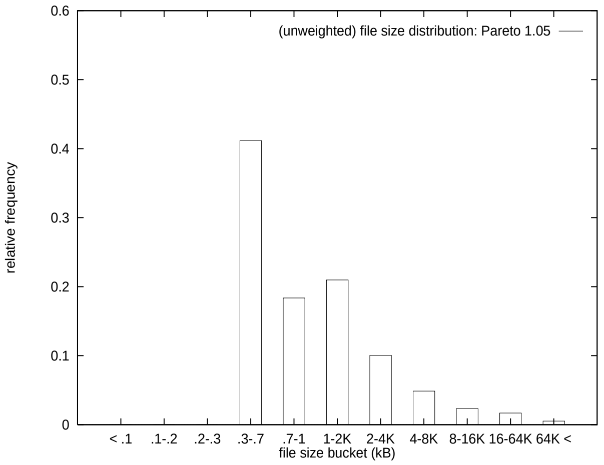
A heavy-tailed random variable:

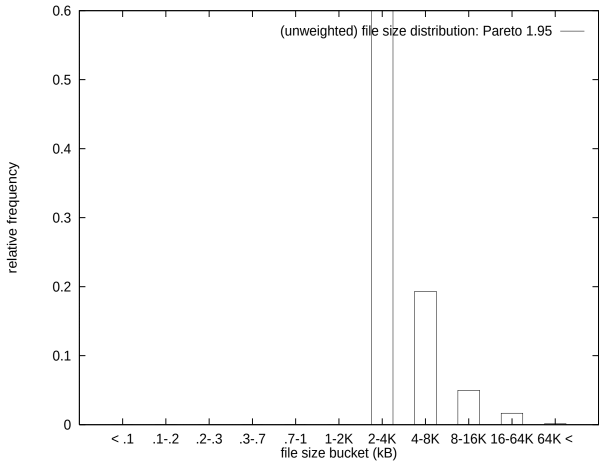
- exhibits extreme variability

  $\longrightarrow$ takes on large values with nonnegligible prob.

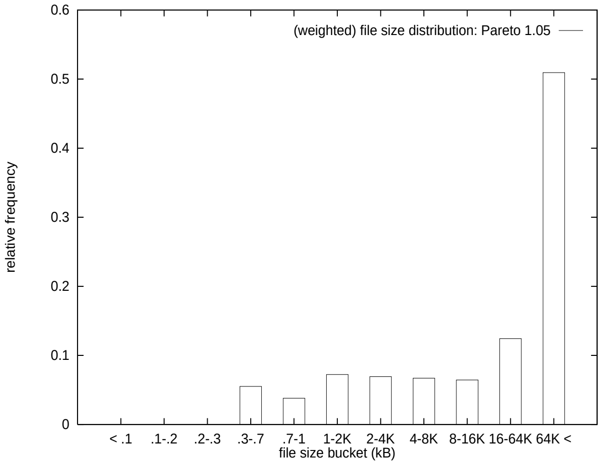
- if $\alpha \geq 2$, then finite variance
- sampling properties are pretty bad
  - slow convergence to population mean
  - underestimation typical
  - need "huge" sample size
- predictable (discussed later)
- Pareto is heavy-tailed across entire range
- typically ignore detailed structure when $Z$ is small

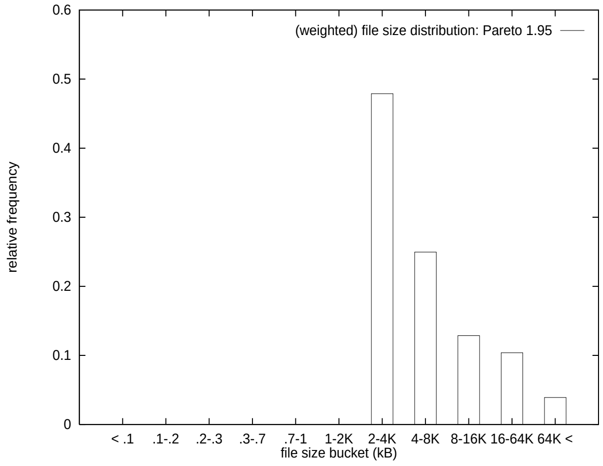
Measured file size distribution from $\alpha = 1.05, 1.95$ Pareto:




Corresponding distribution weighted by volume:

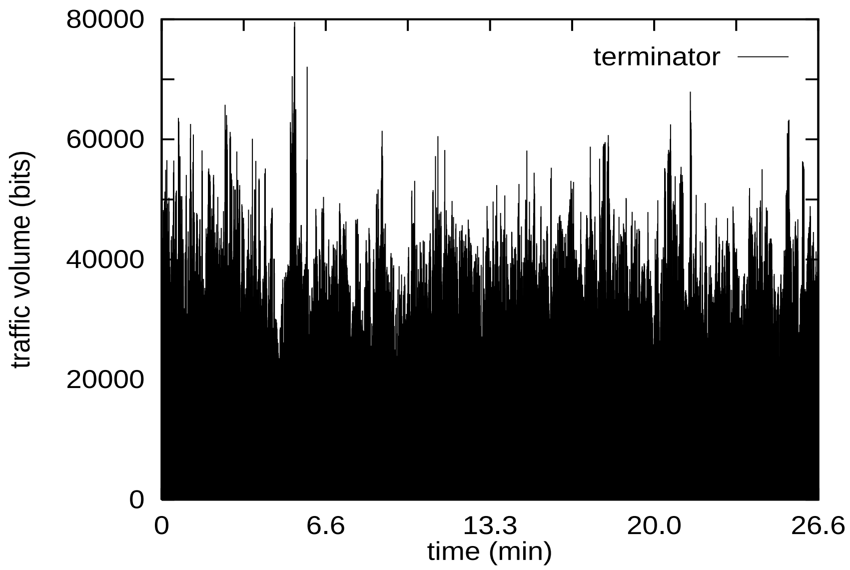



$\longrightarrow$ observe reversal for $\alpha = 1.05$

## Real-time VBR source profile

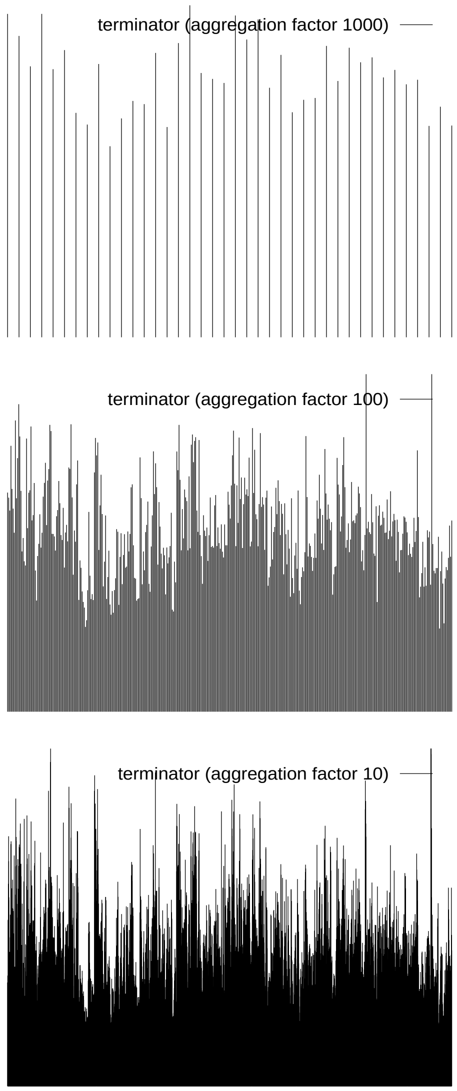
Consider traffic profile of compressed video (e.g., MPEG)




$\longrightarrow$ periodic real-time application

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

Burstiness structure:




$\longrightarrow$ burstiness persists across multiple time scales

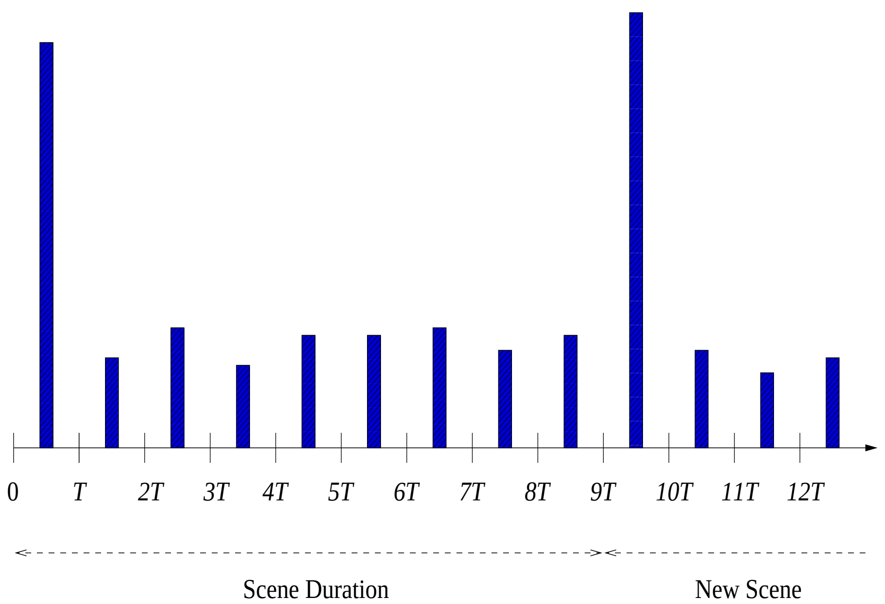
Possible causes:

- Heavy-tailed scene durations

  $\longrightarrow$ facilitates inter-frame compression

- Repeating GOP pattern

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



## Dynamic and Static:

- there exists an intimately relationship between self-similarity & heavy-tailedness
- i.e., between dynamic and static structure

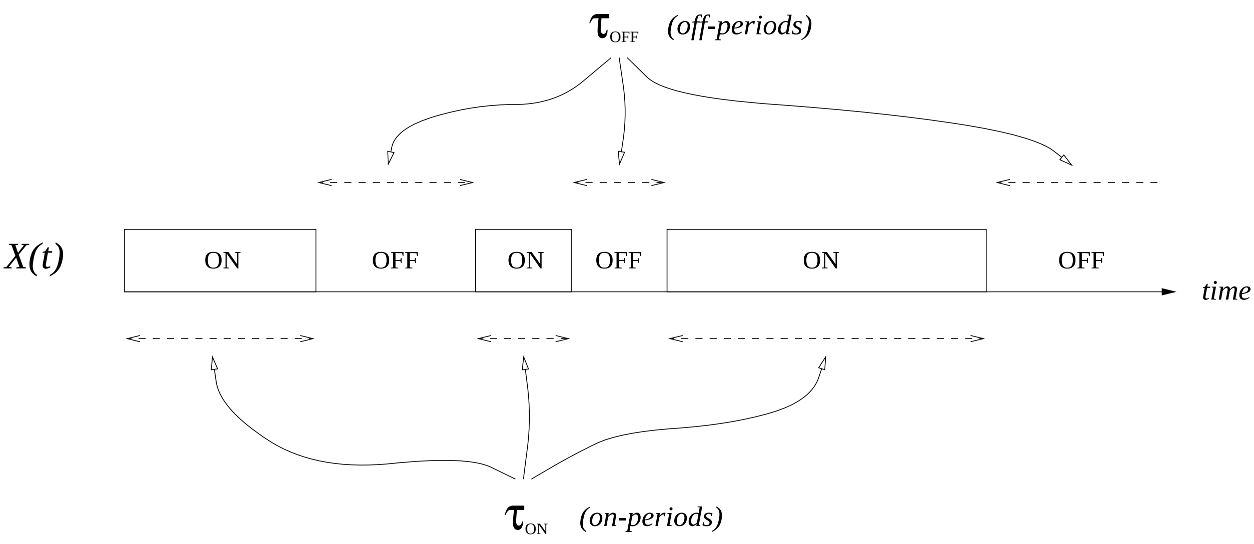
$\longrightarrow$ $\exists$ "if, and only if" relationship

$\longrightarrow$ structural causality

$\longrightarrow$ physical modeling

## On/Off Traffic Model

$\longrightarrow$ packet train view




Interested in special case:

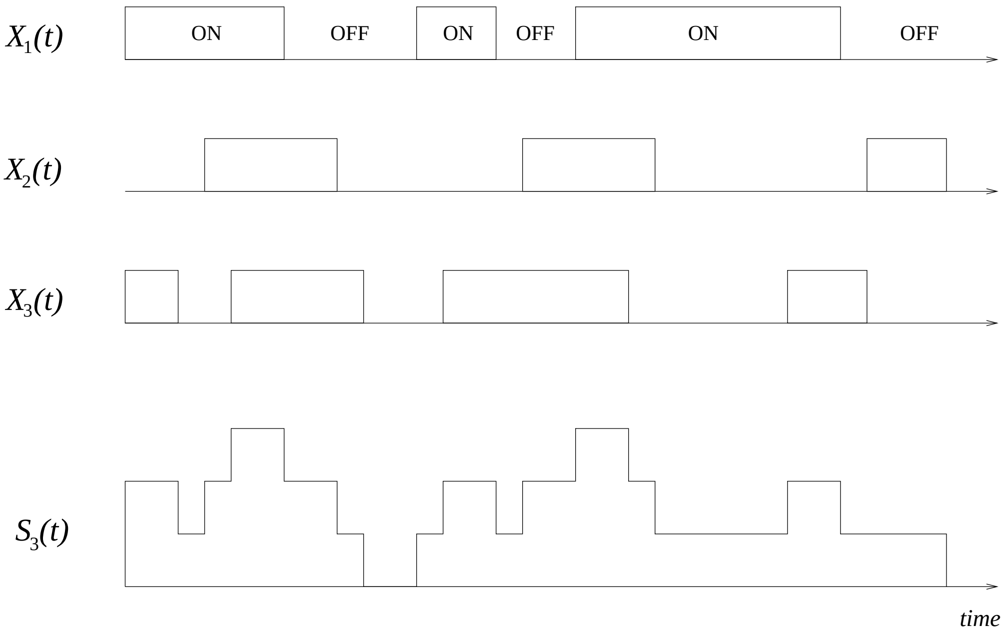
- on-periods are i.i.d.; off-periods are i.i.d.
- on-periods are heavy-tailed or light-tailed

  $\longrightarrow$ random variable $\tau_{\text{on}}$

- 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 $S_3(t) = X_1(t) + X_2(t) + X_3(t)$.



- relevance as multiplexed traffic

  $\longrightarrow$ aggregation in space (or flows), not time

- if sources are independent, then essential effect is Gaussian property by Central Limit Theorem (CLT)

Variations:

- a source is "on" only once (single connection view)

  $\longrightarrow$ Poisson arrival of new connections

  $\longrightarrow$ heavy-tailed on-period

- M/G/$\infty$ queue view of source

  $\longrightarrow$ Poisson arrivals

  $\longrightarrow$ heavy-tailed service time

  $\longrightarrow$ how many servers are busy?

Canonical self-similar traffic model:

$\longrightarrow$ second-order self-similarity

$X(t)$ is *asymptotically second-order self-similar* with Hurst parameter $H$ $(1/2 < H < 1)$ if

$$r^{(m)}(k) \sim \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H})$$

for all $k \geq 1$ as $m \to \infty$.

Here $r^{(m)}(k)$ is the autocorrelation function of time aggregated process $X^{(m)}(i)$.

Note: captures both

- correlation view (slow decay)
- burstiness preservation view

$\longrightarrow \;\; H = 1 - \beta/2$

Main result:

$X(t)$ is asymptotically second-order self-similar with parameter $H = (3 - \alpha)/2$ if, and only if, the on- or off-periods $\tau_{\text{on}}$, $\tau_{\text{off}}$ are heavy-tailed with shape parameter $1 < \alpha < 2$.

$\longrightarrow$ goes back to Mandelbrot

If, in addition, space aggregation:

$\longrightarrow$ leads to Gaussian process