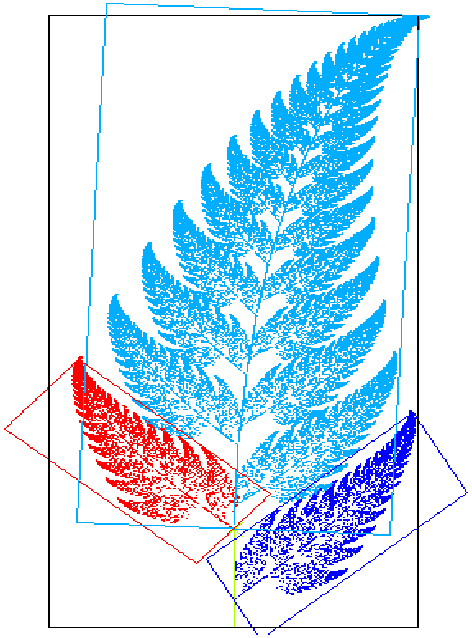
Self-similarity

Self-similarity

a self-similar object is exactly or approximately similar to a part of itself

Romanesco broccoli



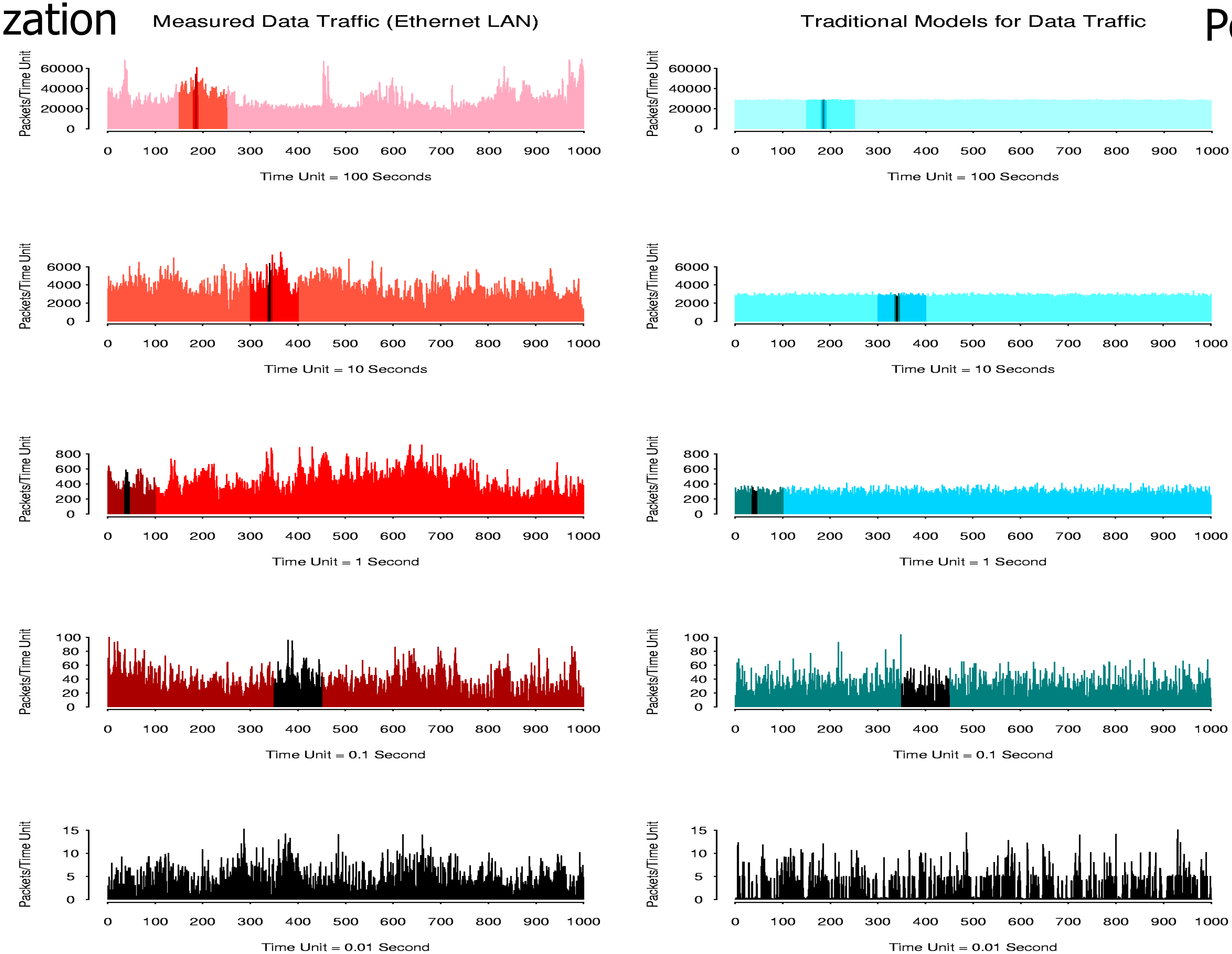
Fern

Snowflake

Self-similarity and network traffic

Realization

Poisson



Self-similarity vs short-range dependency

Aggregate process

Let $X=(X_t : t= 0,1,2,...)$ be a stationary stochastic process, then

$$X^{(m)} = (X_k^{(m)} : k=1,2,...)$$

where

$$X_k^{(m)} = 1/m(X_{km-m+1} + ... + X_{km}), k>0$$

Autocorrelation Function (ACF)

Correlation between values of the process at different times,

If ACF ~ 0 then traffic observations that are far apart are independent.

Self-similarity: definition (I)

A. A zero mean stationary process X is called self-similar (with self-similarity parameter H), if for all $m \geq 1$,

$m^H X$ has the same Distribution as $X^{(m)}$

H is called the Hurst parameter.

To be self-similar, $0.5 < H < 1$.

If $H=0.5$, X is short-range dependent.

Self-similarity, definition (II)

If process X with ACF function r(k) is self-similar, then

$$r^{(m}$$

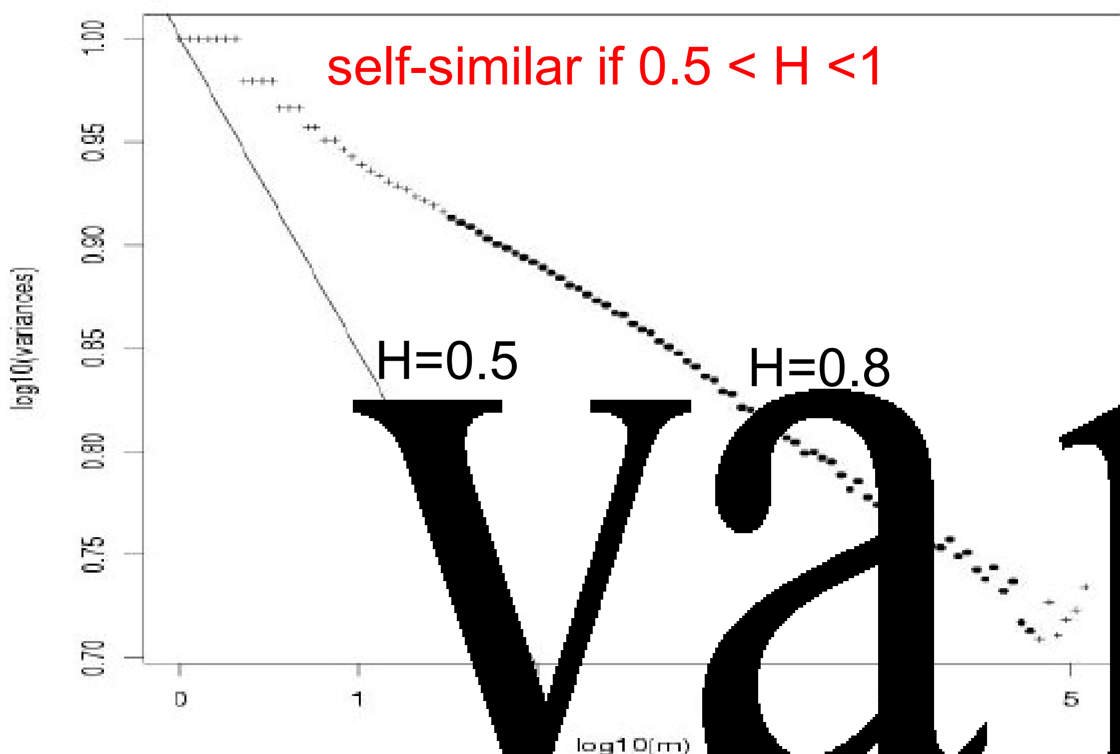
For large m, $r^{(m)}$ is ACF of $X^{(m)}$.

Self-similarity, definition (II)

Stationary process X with ACF function r(k), exhibits self-similarity if for some 0.5 < H < 1

Test of variance

The measure of how far a set of samples is spread out,
$\mathrm{var}(X)=E((X - E(X))^2)$.



var(

Short-Range Dependence (SRD)

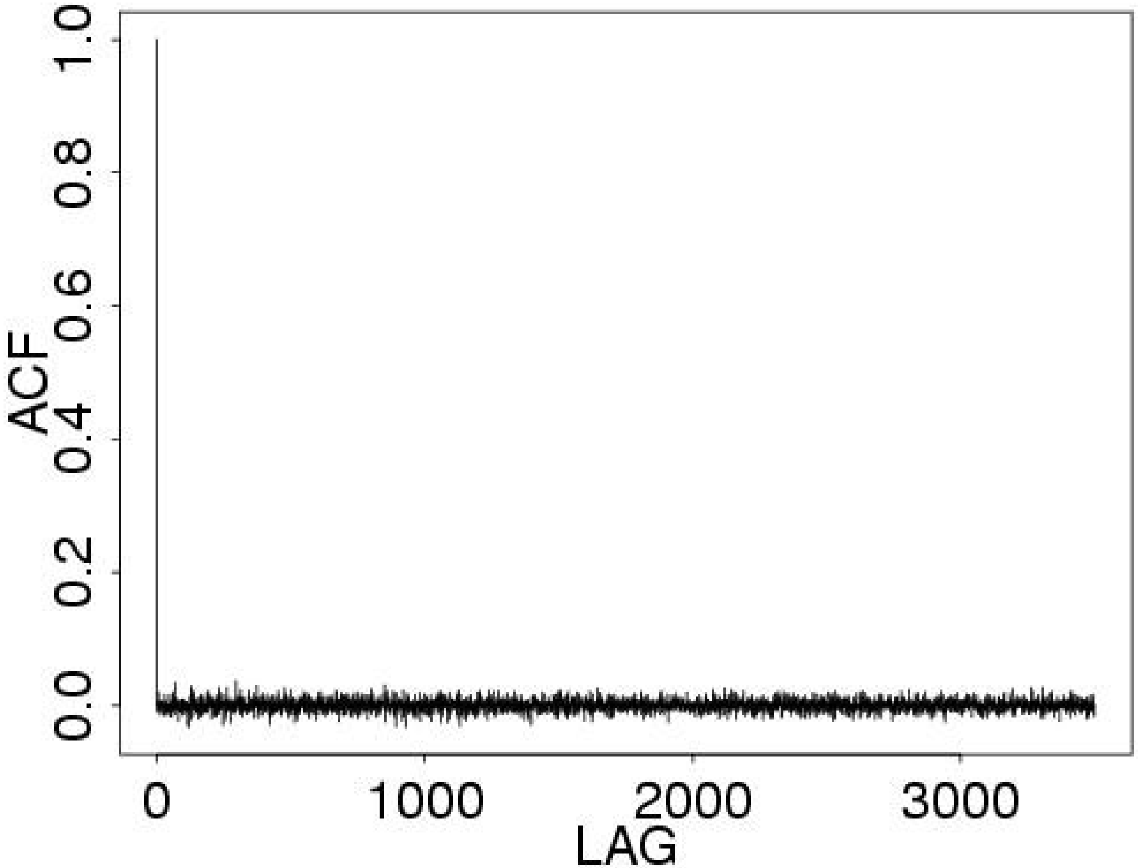
Stationary process X with ACF function r(k), exhibits SRD if there exists $0 < \rho < 1$ and $\tau > 0$ with

There are no observations "in the past" that are correlated with current observations

Important feature: Autocorrelations decay (at least) exponentially fast for large lags k

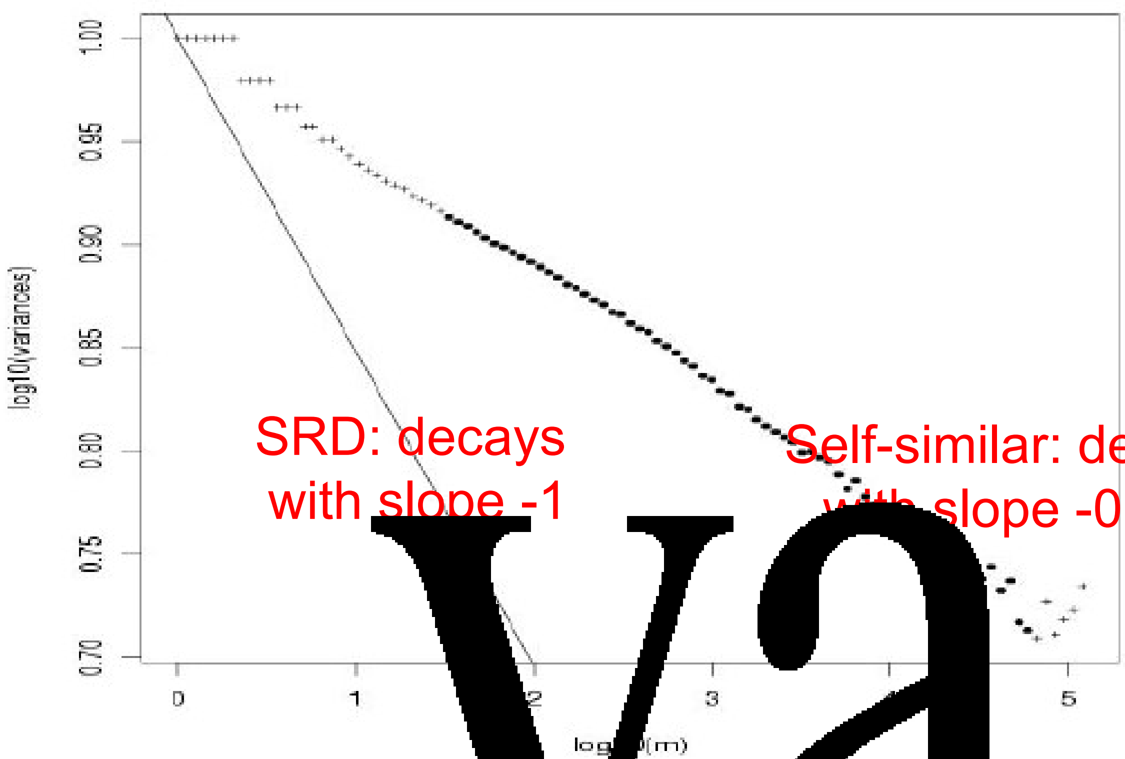
$r(k$

An example of SRD processes



Test of variance

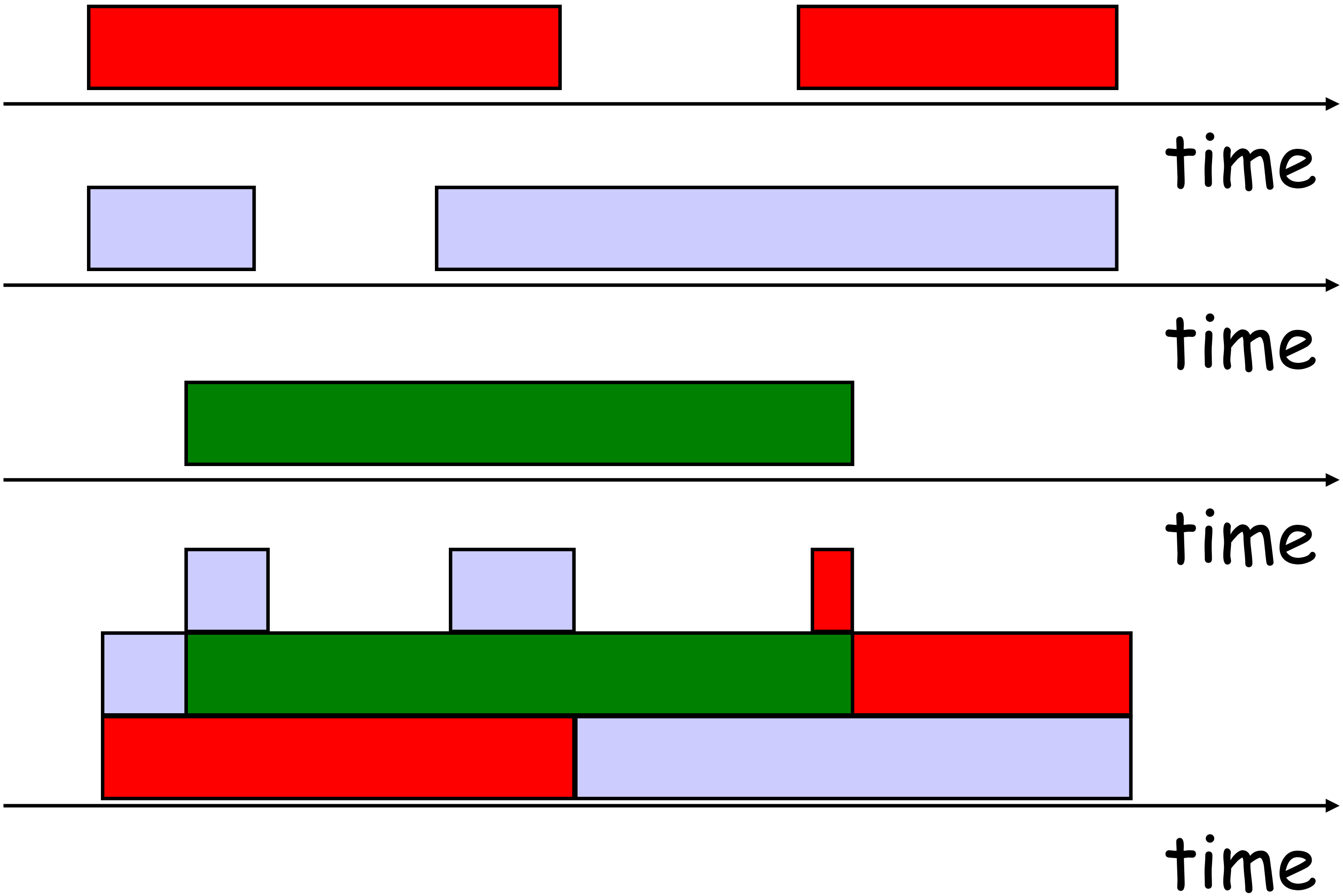
The variance of the sample mean (e.g. $X^{(m)}$), as a function of m, satisfies:



var(

Route cause of Self-similarity in the Internet

Superposition of independent sources



Self-similarity vs short-range dependency

Traffic is superposition of ON/OFF sources

Self-similar:

if durations of periods are heavy-tailed with infinite variance, superposition is self-similar

Short-range dependent:

if durations of periods are light-tailed, superposition is short-range dependent

Light-tailed distributions

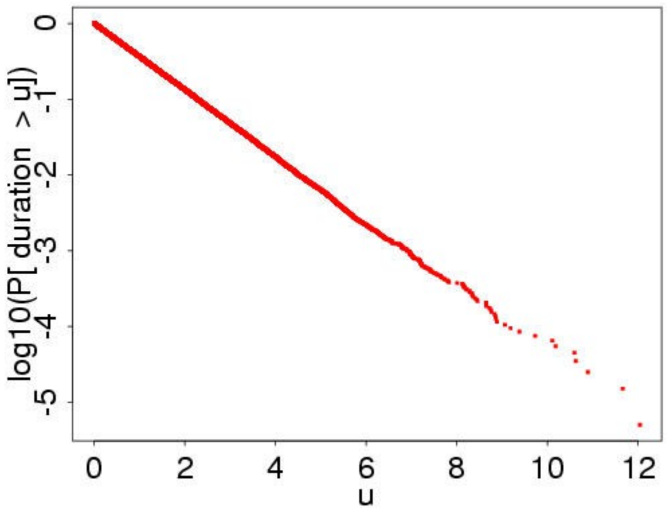
- X random variable with distribution function F.
- F is said to be light-tailed if there exists c > 0

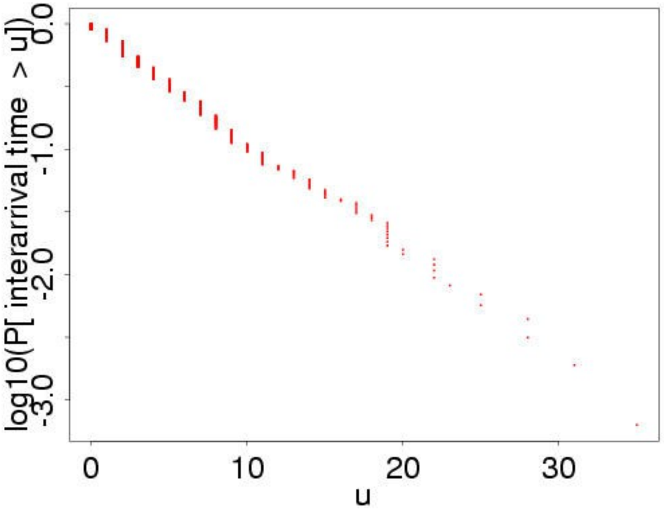
- Important feature: tails decay exponentially fast for large x; i.e.,

Where P[X>x] is the complementary cumulative distribution function (CCDF)

Light-tailed distributions

- Examples: Exponential, Normal, Poisson, Binomial
- Key features:
  - F has limited variability
  - F is tightly concentrated around its mean
  - F has finite moments
  - P[X > x] (CCDF) vs. x on log-linear scale is linear for large x





Heavy-tailed distributions

- X random variable with distribution function F
- F is said to be heavy-tailed if there exists c > 0

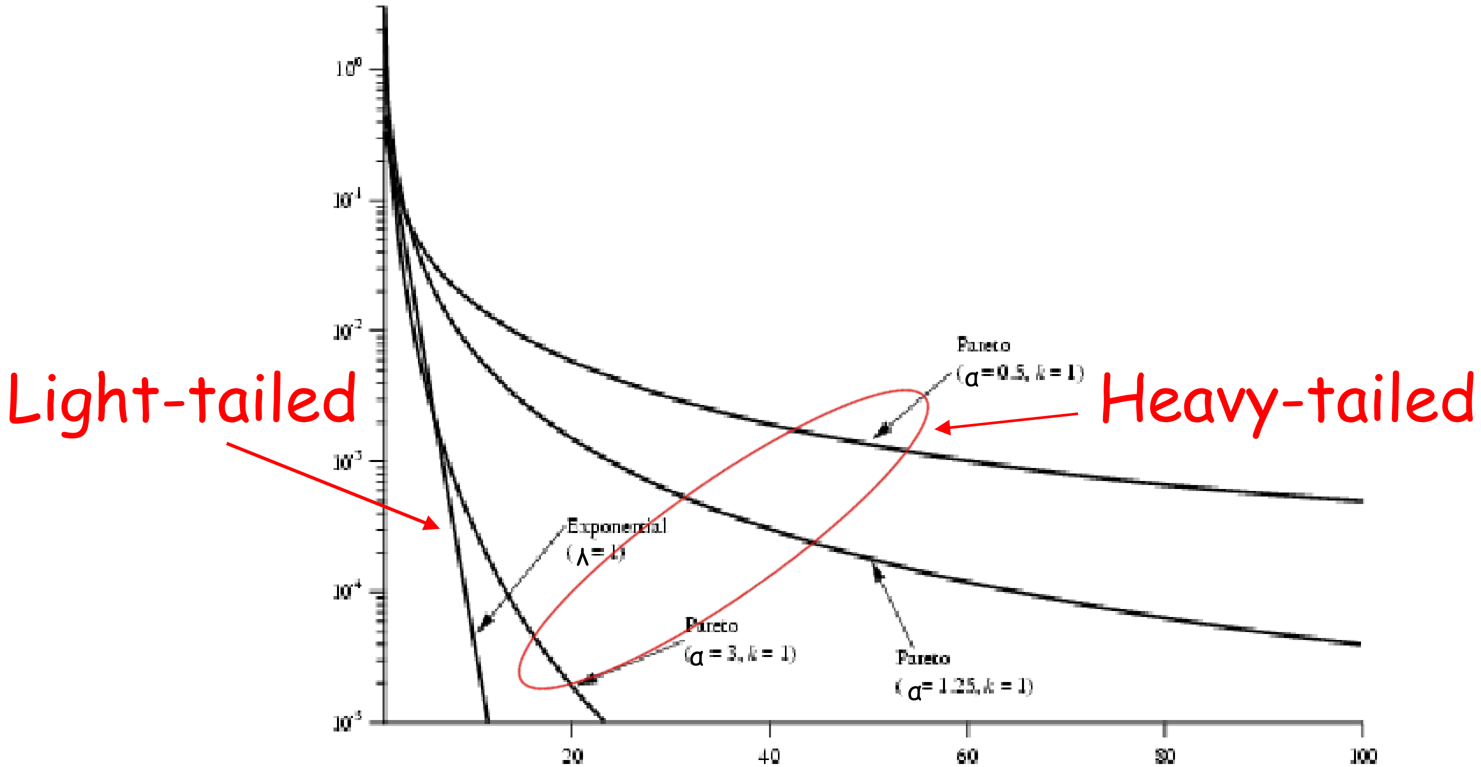
$1 < \alpha < 2$, X has finite mean but infinite variance.

→parsimonious model (small number of parameters)

- **Note:** Stationary models such as [illegible]f-similari[illegible]eavy-tailed distributions and LRD are appropr[illegible]te models [illegible]ime-scale o[illegible]econds to minutes to hours. With them [illegible] model in ([illegible]) "invariants[illegible]

Heavy-tailed distribution

Ex. Pareto distribution: $p(x) = \alpha k^{\alpha} x^{-\alpha-1}$



if α≤2, the variance is infinit

if α≤1, the mean is also infinit

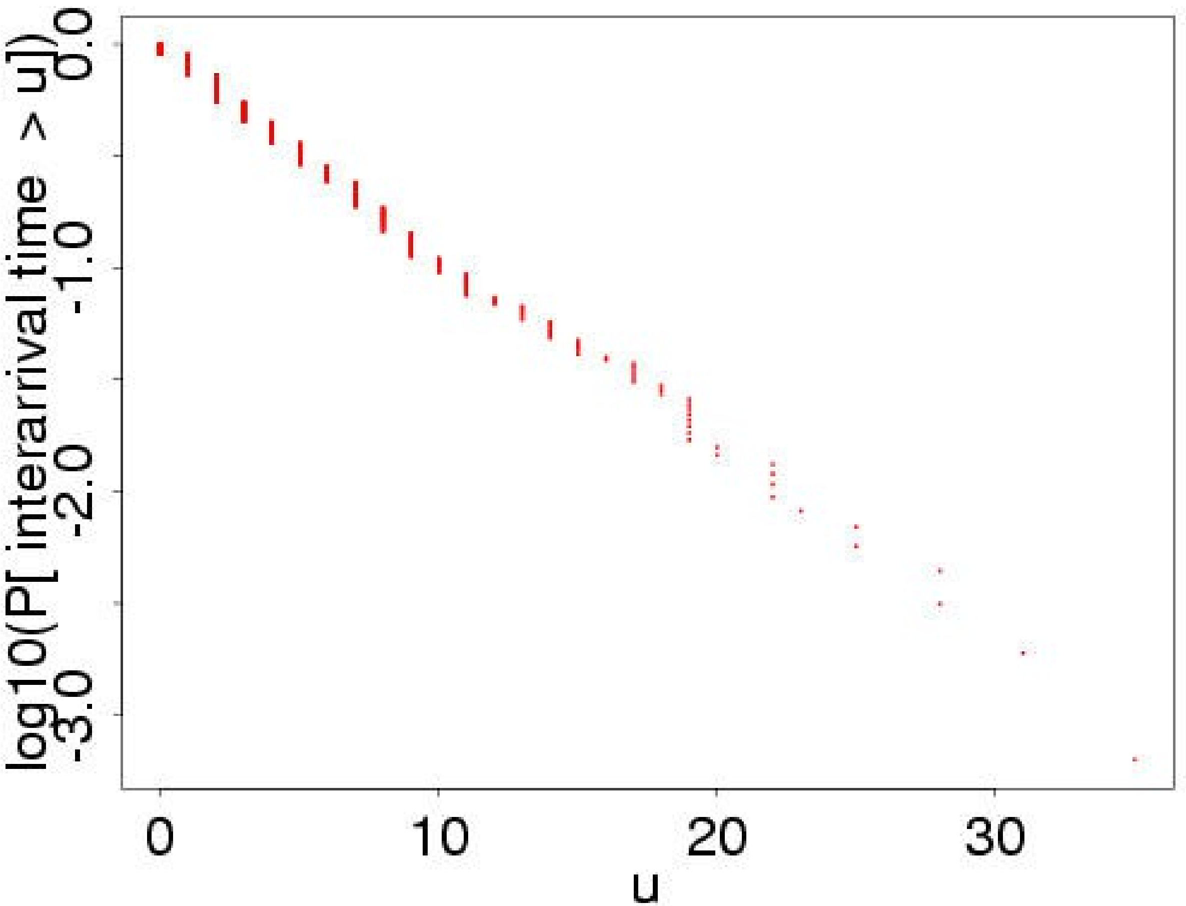
Long-range dependence (LRD) & Heavy-tail Distributions

- There are observations in the past that are correlated with current observations
- Parsimonious models available
- It changes the way we design systems

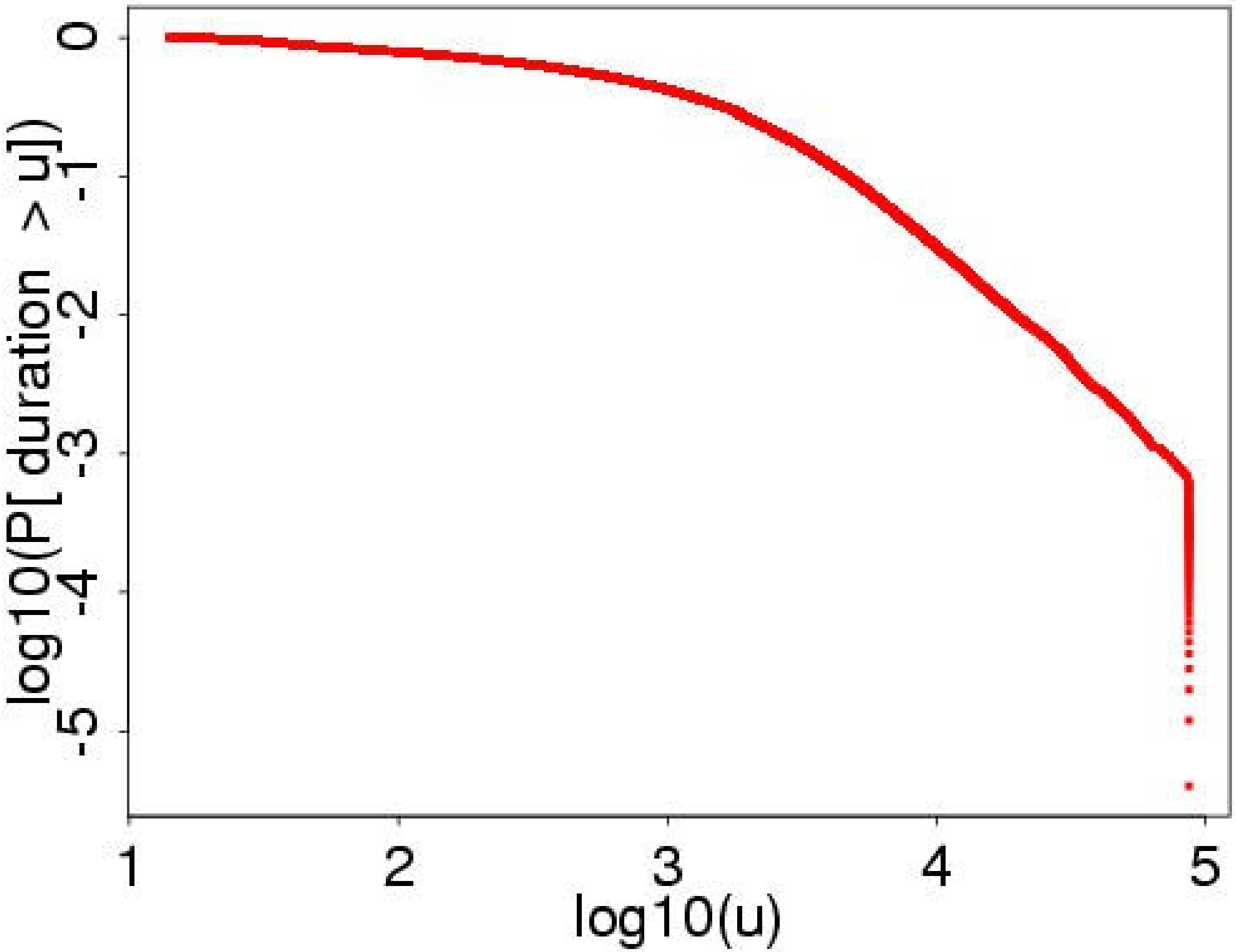
(e.g., how to deal with bursts, effect of queuing, protocol design)

Detour
Characteristics of modem calls (~1999)

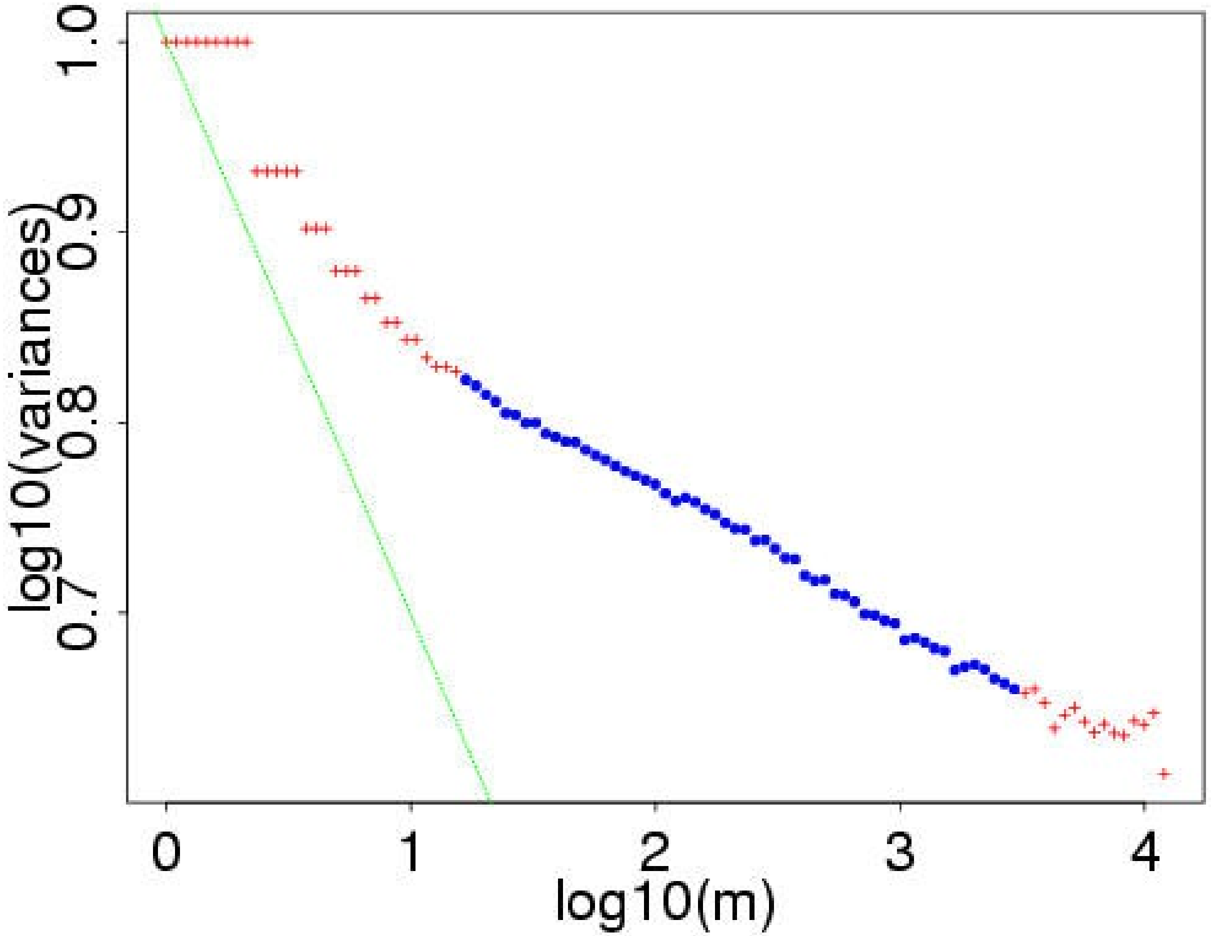
Interarrival times of modem calls



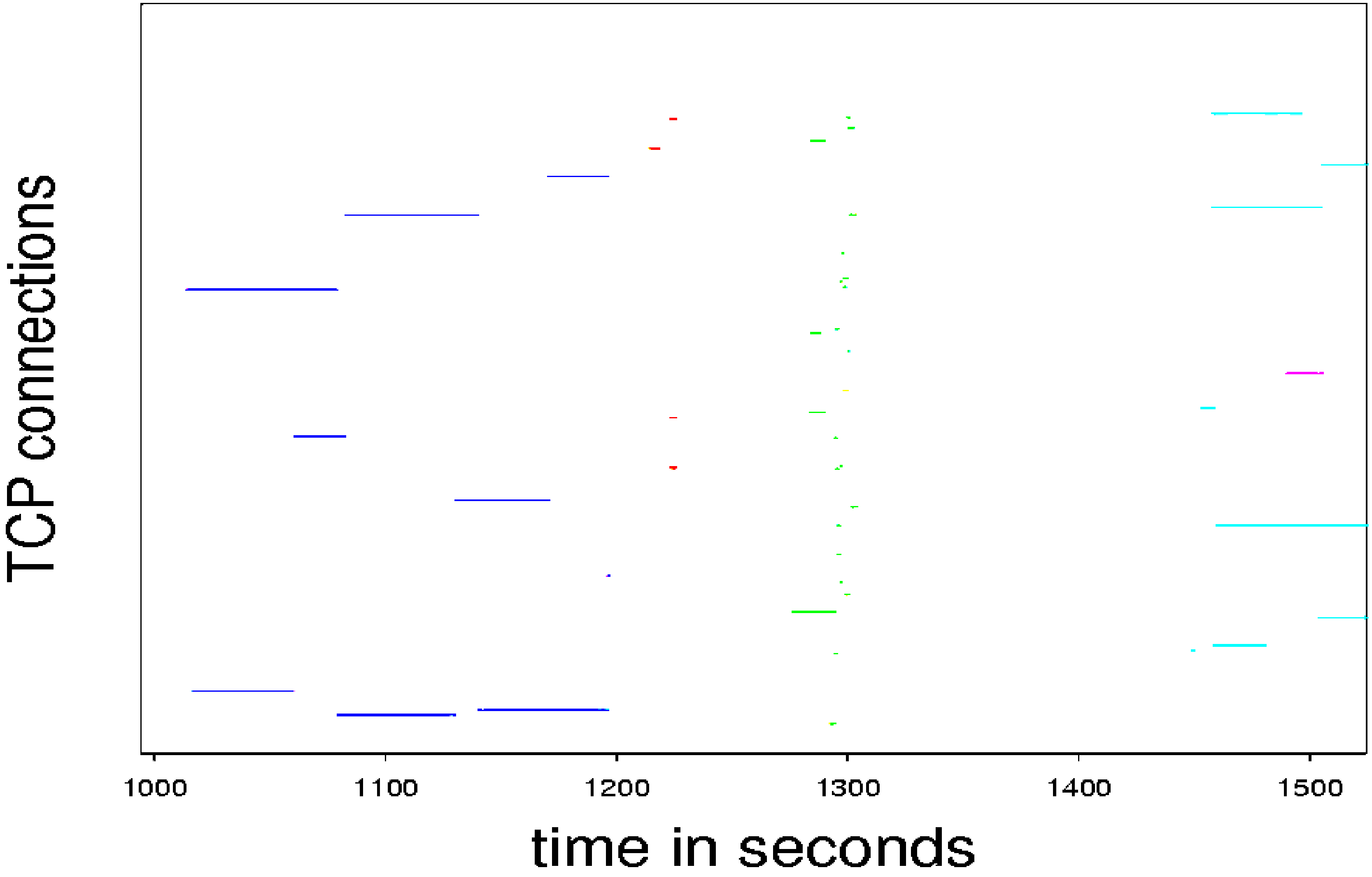
Durations of modem calls



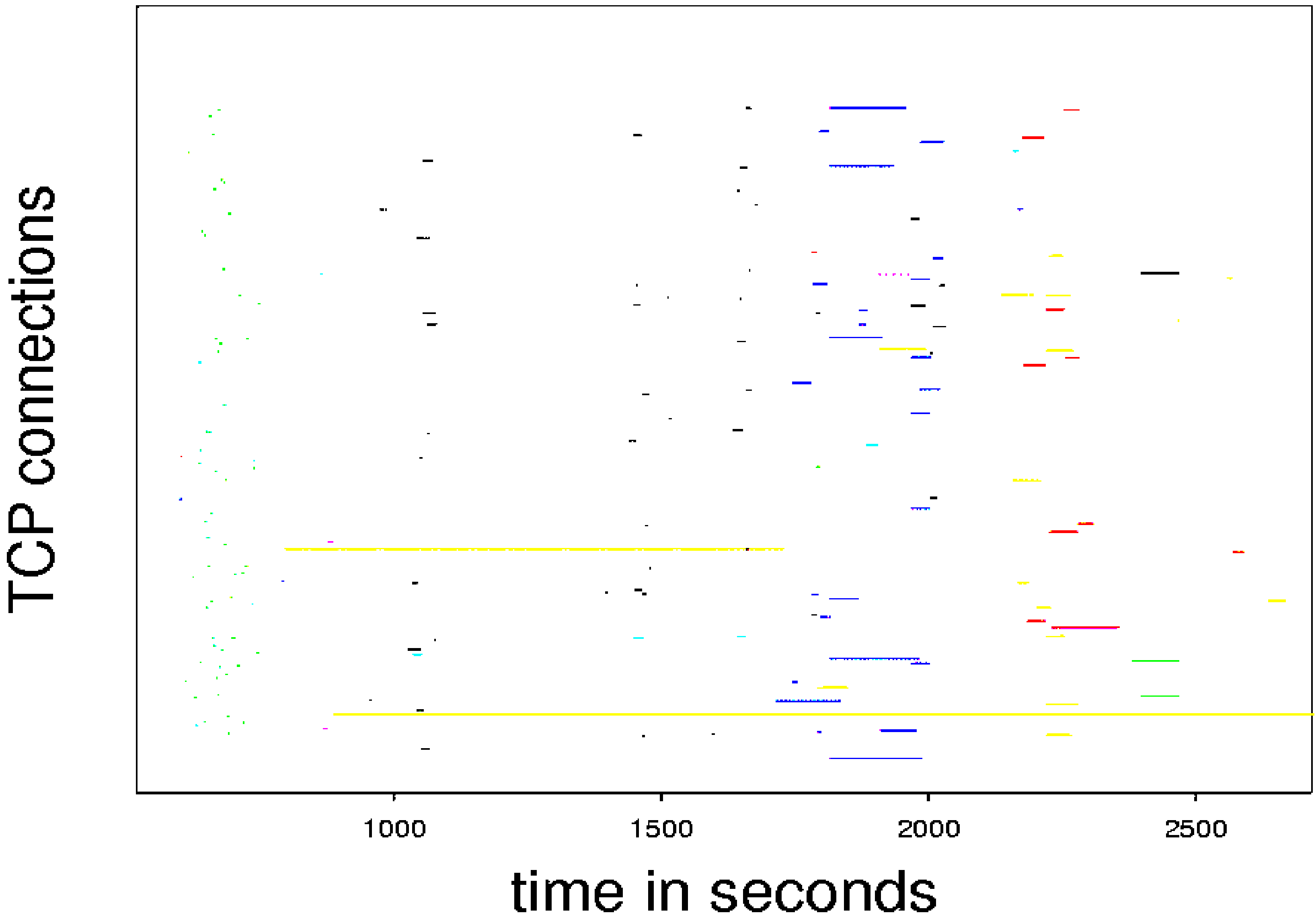
What about pkts from modem calls



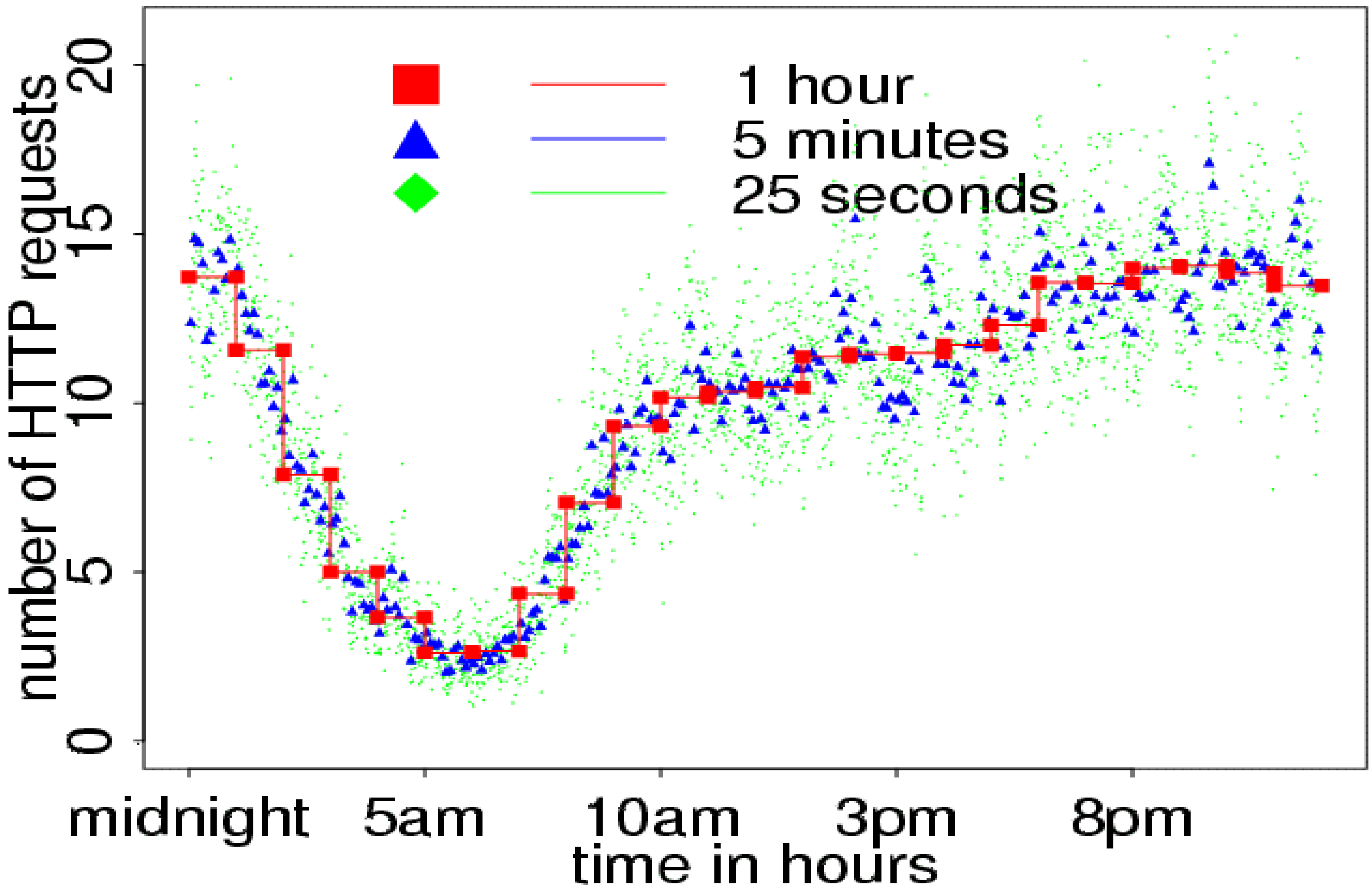
General characteristics of WWW transfers



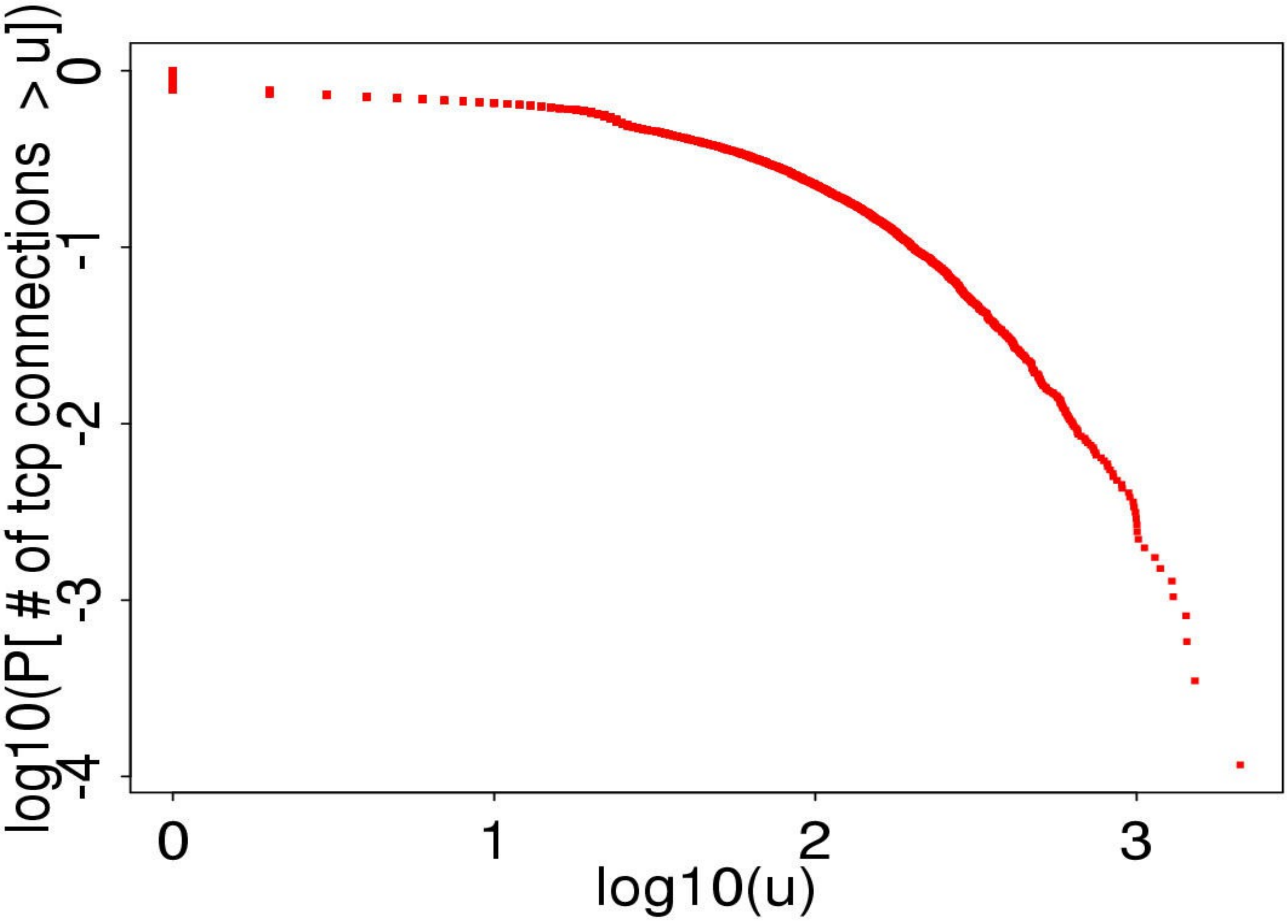
General characteristics of WWW transfers



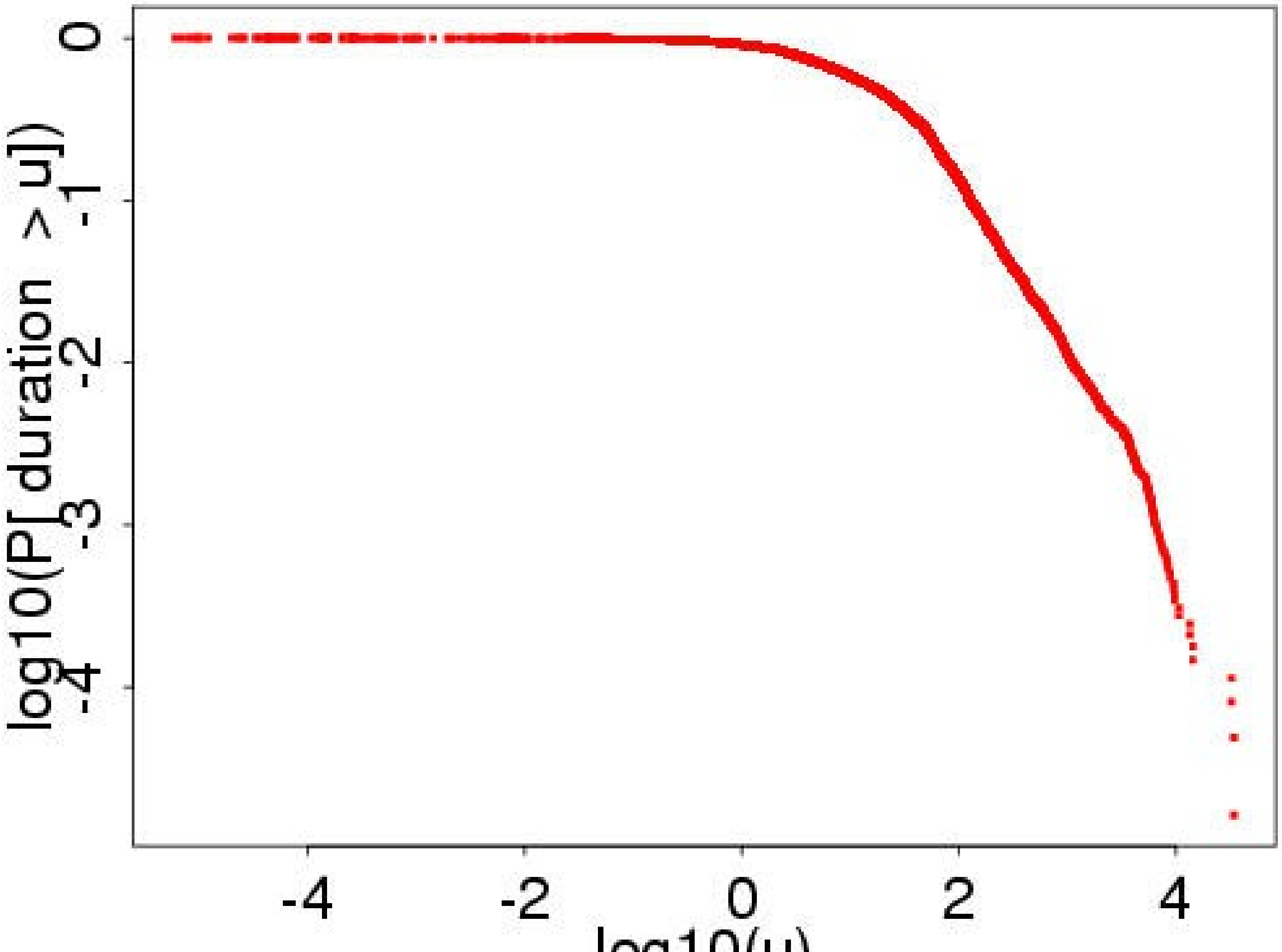
General characteristics of WWW transfers



of TCP connections per session



Flow durations

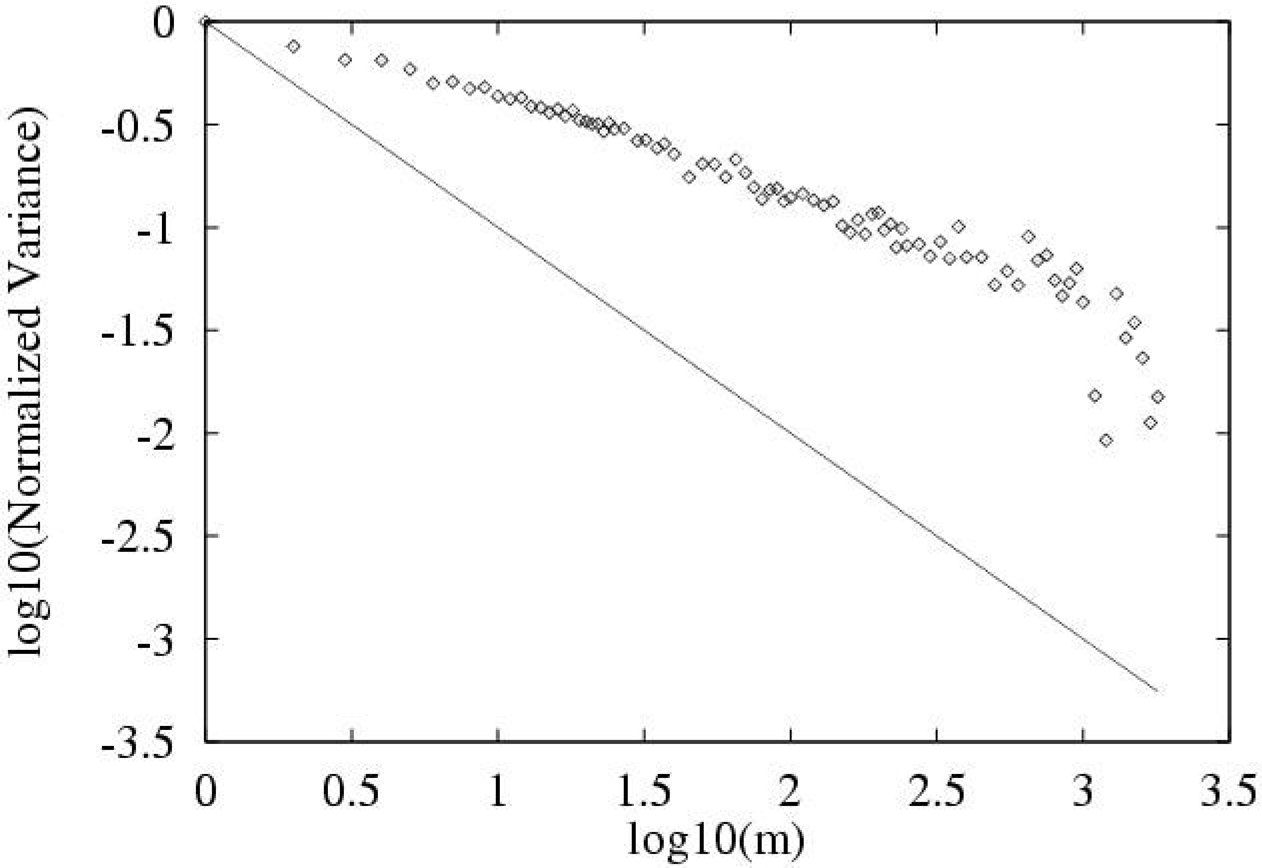


Web client trace analysis 1995

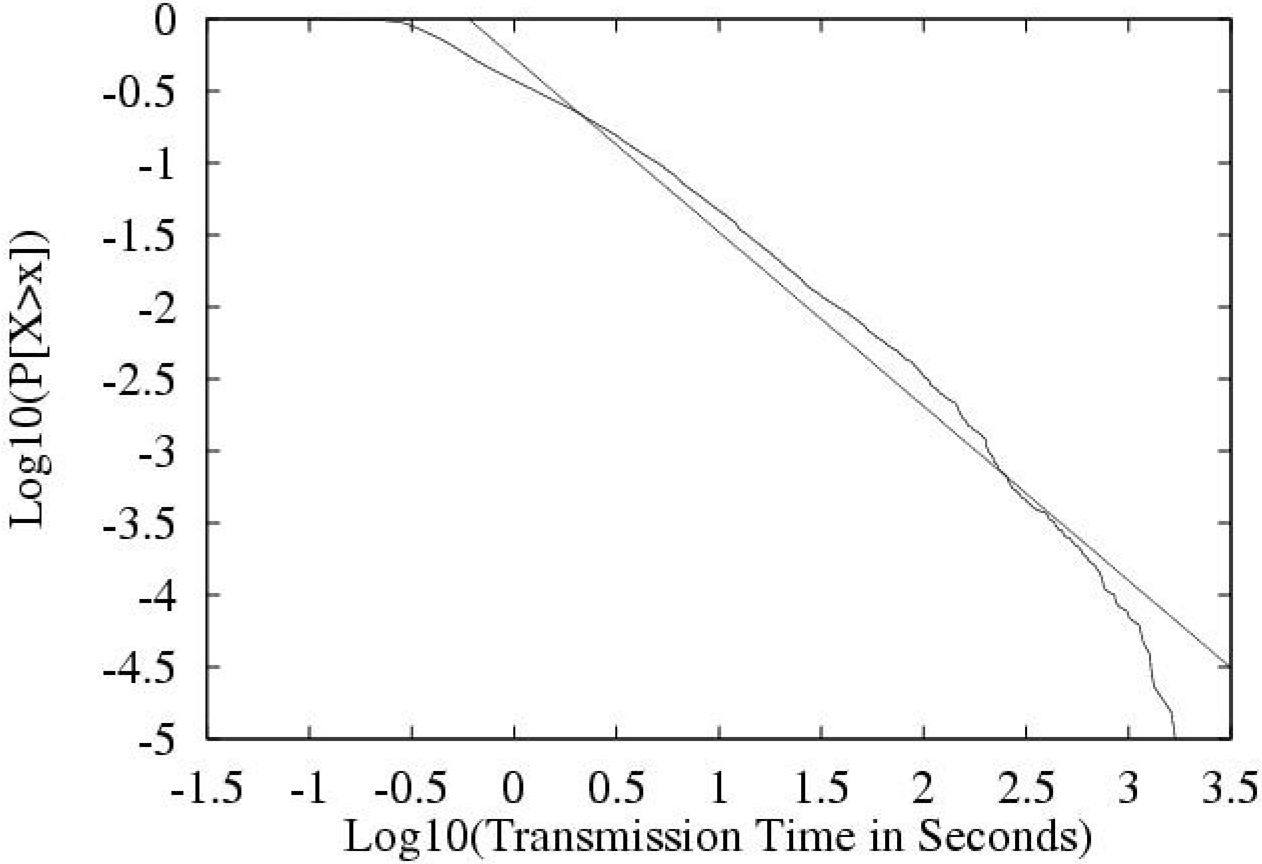
- Modified Web browser (Mosaic)
- Population: students at BU
- Duration: 21 Nov 94 to 8 May 95

Sessions	4,700
Users	591
URLs Requested	575,775
Files Transferred	130,140
Unique Files Requested	46,830
Bytes Requested	2,713 MB
Bytes Transferred	1,849 MB
Unique Bytes Requested	1,088 MB

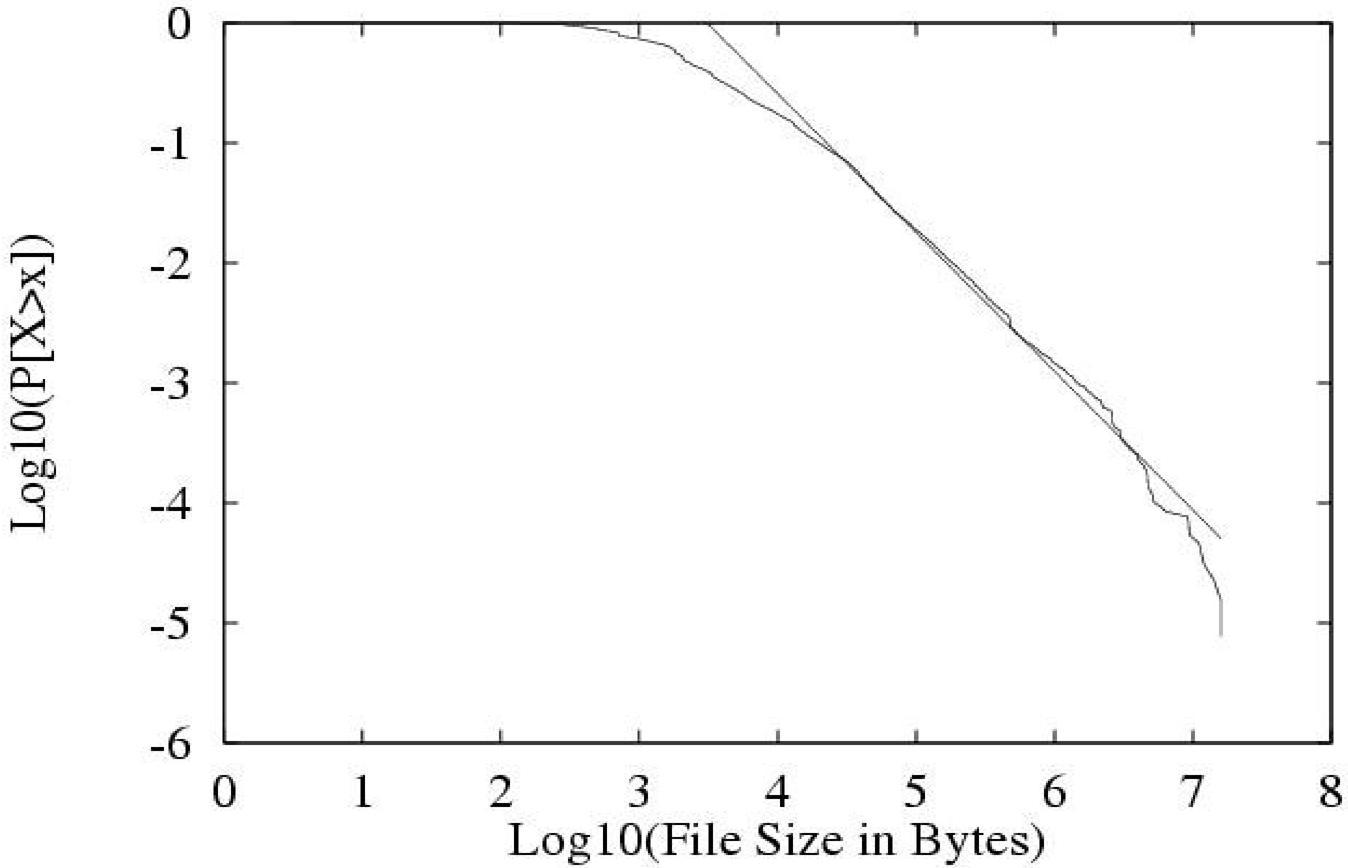
What about WEB traffic



Durations of WEB transfers???

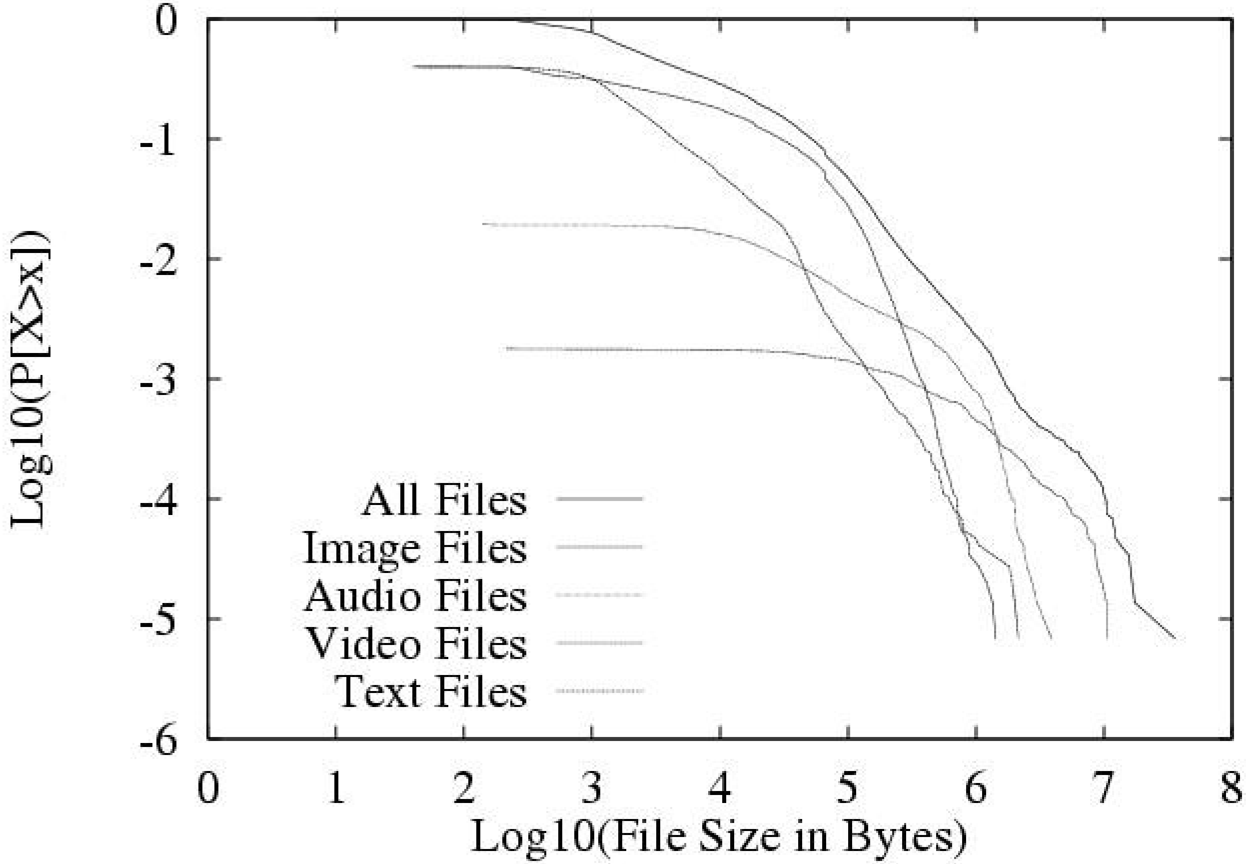


File size of WEB transfers???

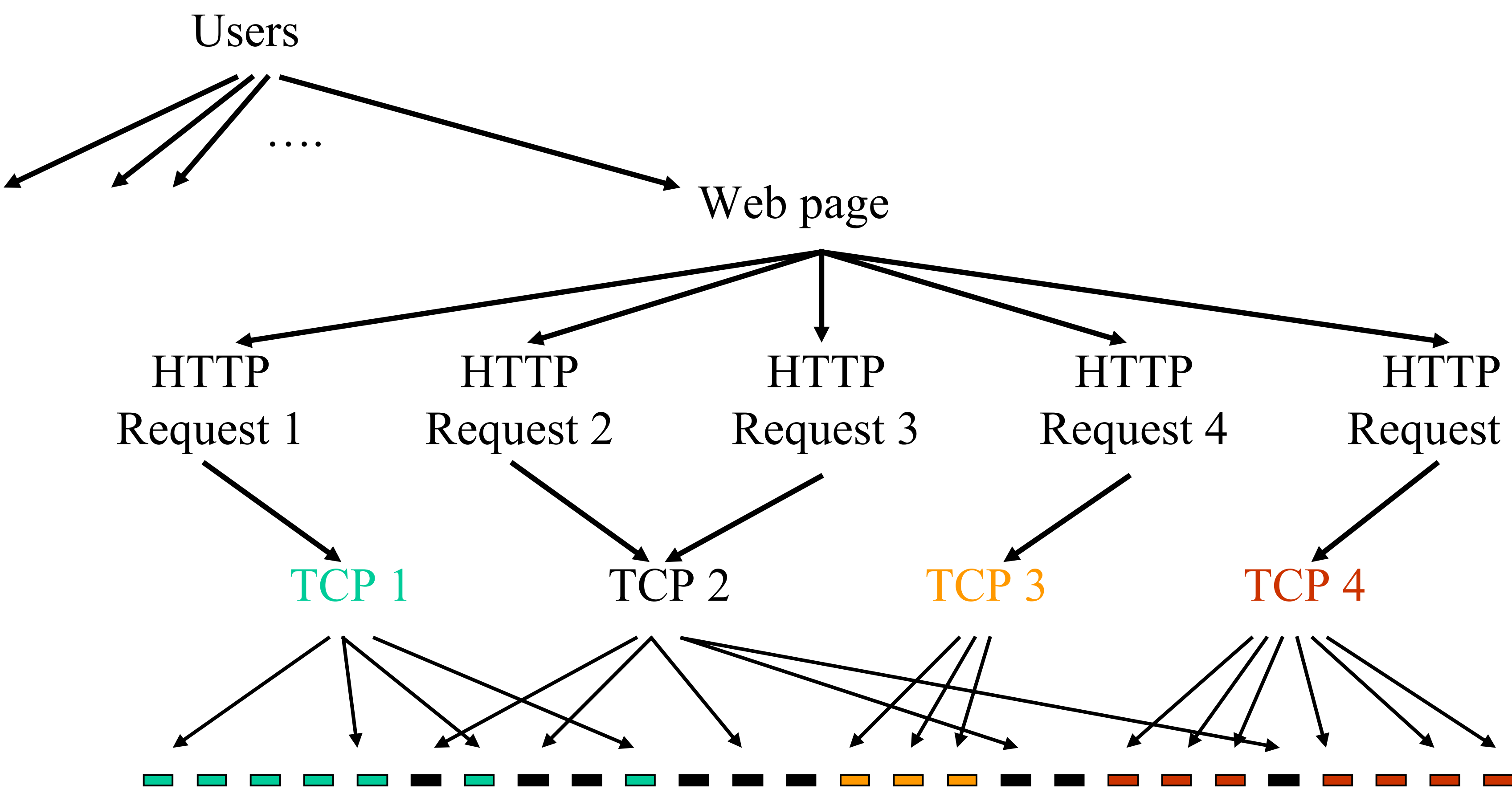


What about the available files?

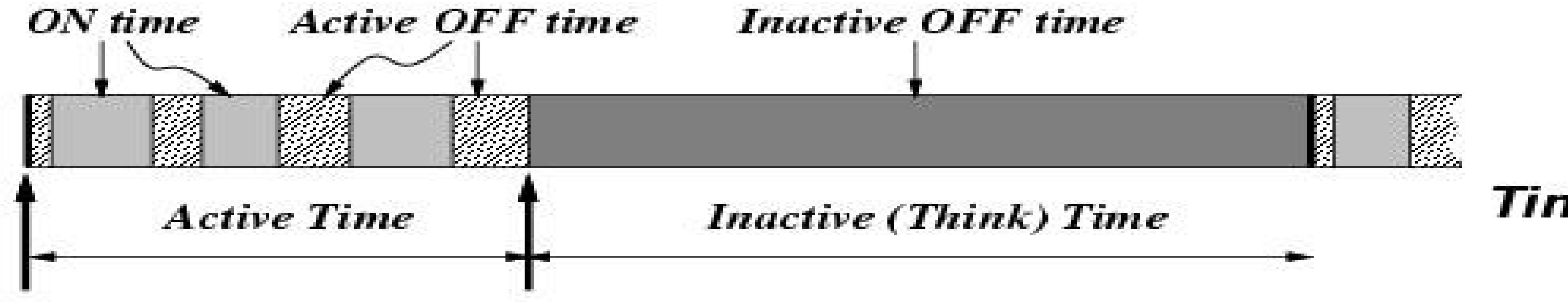
Long tail due to superposition of heterogeneous filesizes/filetypes



What about off times? Process Factor



What about the WEB: Human Factor



Interarrival times of URL requests: Human Factor