Slide supporting material

Lesson 17: Models for Traffic Sources

Giovanni Giambene

Queuing Theory and Telecommunications: Networks and Applications
2nd edition, Springer

In the Past…

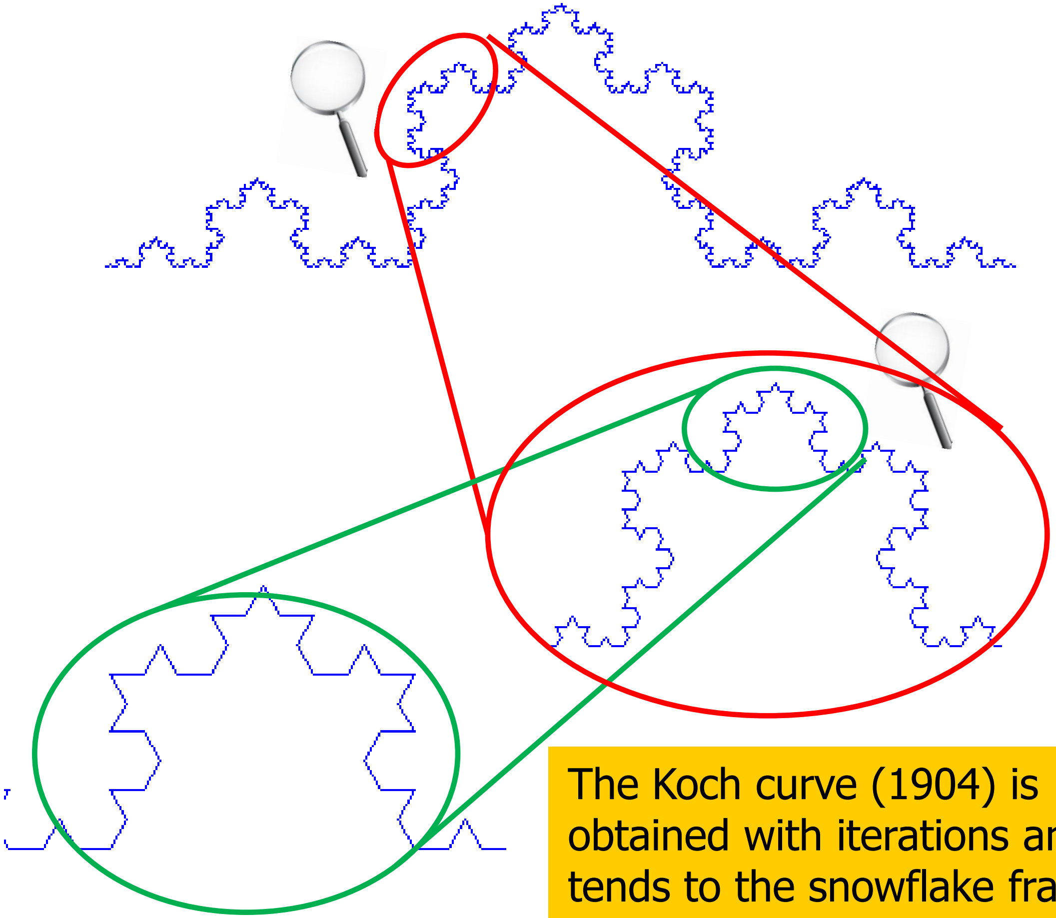
- In the world of telephony, traffic modeling was the basis of initial analytical methods for performance evaluation:
  - **Poisson arrival process**
  - **Exponentially distributed call (service) duration**.
- The strength of these types of traffic models is the **memoryless property of the exponential distribution and the possibility to solve queues by means of Markov chains** (M/M/… queues).
  - Queuing literature was based on these assumptions that allowed a very successful design of telephone networks.

The Initial Discovery

- In 1989, W. E. Leland and D. V. Wilson begun taking high-resolution traffic traces at Bellcore US. New phenomena were highlighted on data traffic:
  - **Highly bursty traffic**;
  - **Burstiness on multiple time scales**: burstiness remains after aggregation on several time scales.
    - If we plot the number of packets arrived per time interval as a function of time, then the **plot looks "the same", regardless of the size of the interval we choose (fractal property).**
  - **Heavy-tailed distributions of file sizes** and corresponding transmission times (~ infinite variance or in any case very high values of the variance).

Fractals

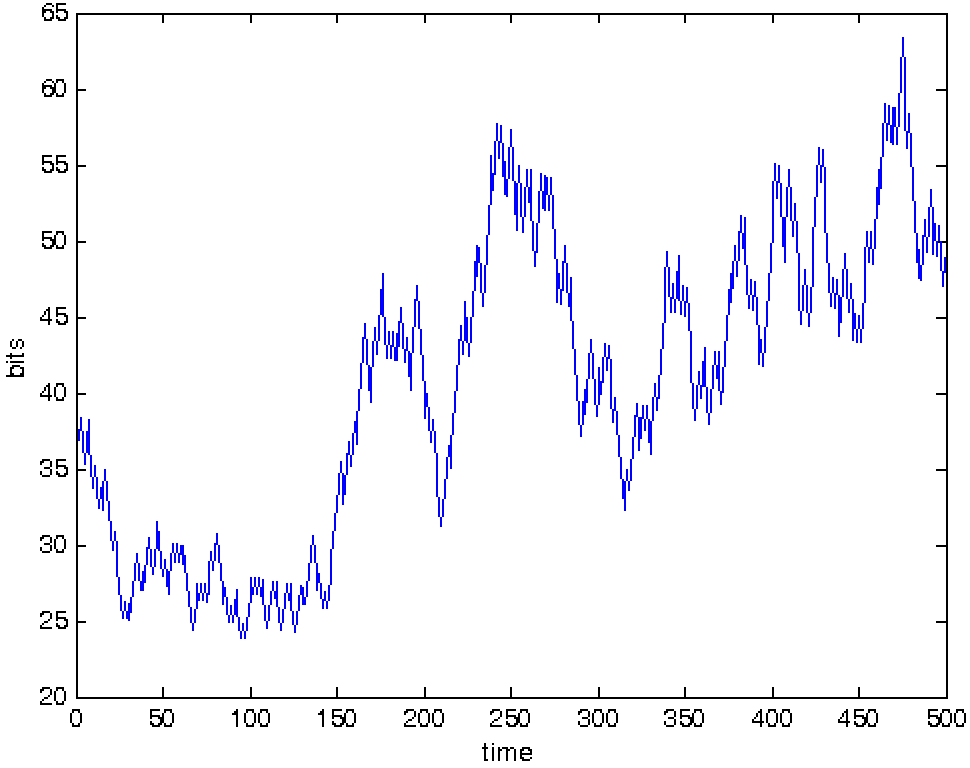
- A fractal is a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole.
- **The traffic in the network can have fractal characteristics in relation to time**: as we aggregate the traffic on larger time scales, we achieve the same traffic profile.



The Koch curve (1904) is obtained with iterations and tends to the snowflake fractal curve.

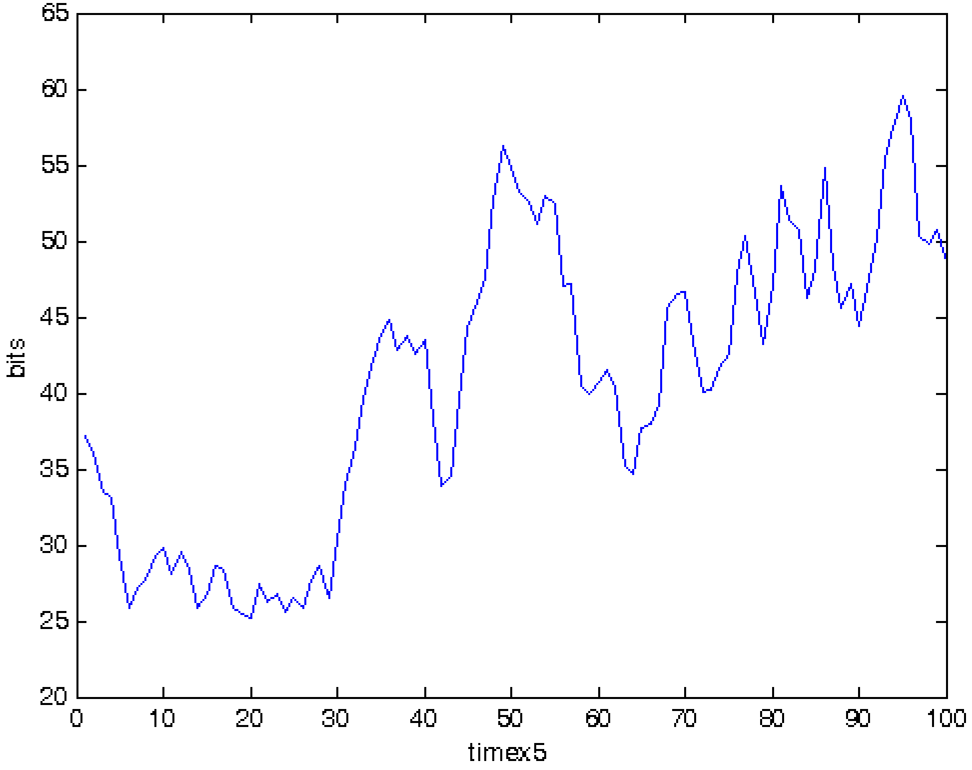
Example of Traffic with Fractal Property

- **The bits generated as a function of time on 1 s time basis, X.**



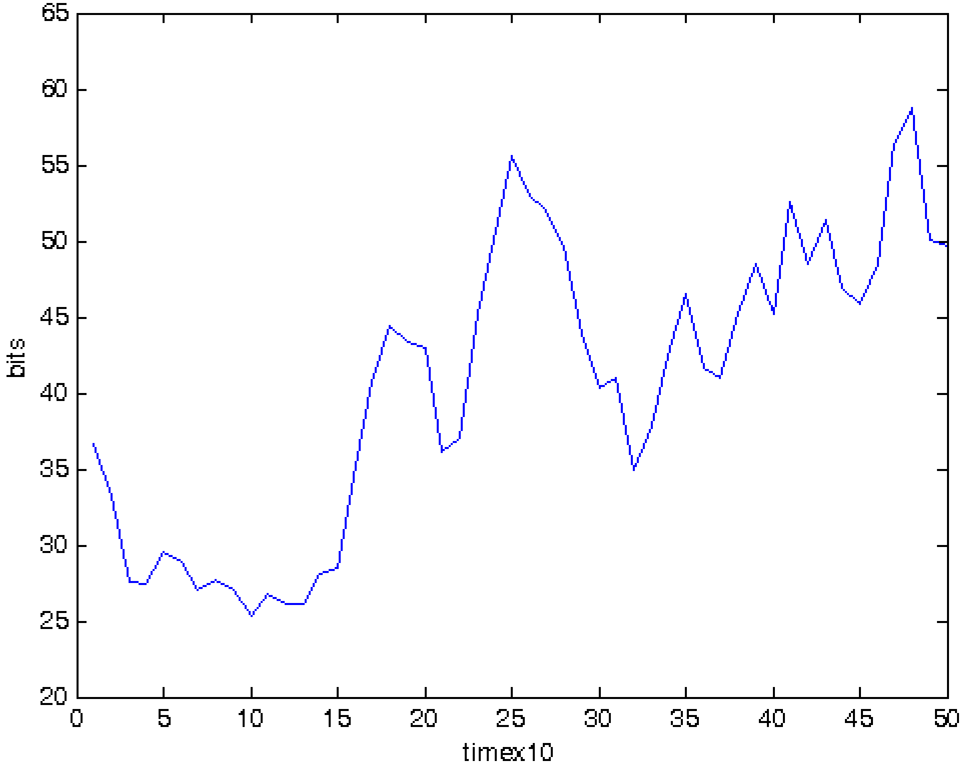
Example of Traffic with Fractal Property

- **The bits generated as a function of time on 5 s time basis, $X^{(5)}$.**



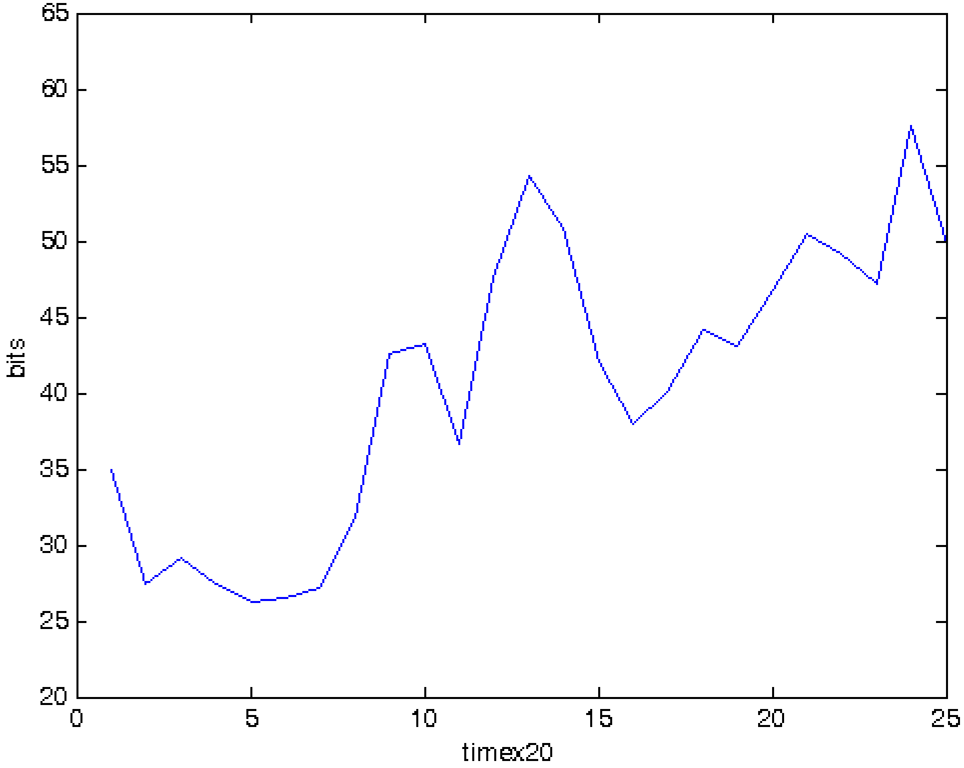
Example of Traffic with Fractal Property

- **The bits generated as a function of time on 10 s time basis, $X^{(10)}$.**



Example of Traffic with Fractal Property

- **The bits generated as a function of time on 20 s time basis, $X^{(20)}$.**



The traffic profile (burstiness) does not change by aggregating the traffic on wider time intervals.

Important Literature on Self-Similar Traffic

W. E. Leland, M. Taqqu, W. Willinger, D. Wilson, “On the Self-Similar Nature of Ethernet Traffic”, *IEEE/ACM Trans. Networking*, 1994.

V. Paxson, S. Floyd, “Wide-Area Traffic: The Failure of Poisson Modeling”, *IEEE/ACM Trans. Networking*, 1995.

M. Crovella, A. Bestavros, “Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes”, *IEEE/ACM Trans. Networking*, 1997.

W. Willinger, M. S. Taqqu, R. Sherman, D. V. Wilson, “Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level”, *IEEE/ACM Trans. Networking*, Vol. 5, No. 1, pp. 71-86,1997.

Definitions

- Let us consider a stochastic time series $X = (X_t; t = 1, 2, 3, \ldots)$ representing the **amount of data generated in consecutive time intervals of equal size**.
  - If X_t is a second-order stationary process, then **the mean $E(X_t) = \mu$ and the variance $\mathrm{Var}(X_t) = \sigma^2$ are time-independent**, and the autocorrelation depends only on the lag k between the instants t+k and t as:

$$r(k) = \frac{E\left[\left(X_{t+k} - \mu\right)\left(X_t - \mu\right)\right]}{\sigma^2}$$

Note: $r(0) = 1$

  - We define the **m-aggregated series** $X^{(m)} = (X_n^{(m)}; n = 1, 2, 3, \ldots)$ by averaging the original series X over non-overlapping blocks of size m:

$$X_k^{(m)} = \frac{X_{(k-1)m} + \ldots + X_{km-1}}{m}$$

Definitions (cont'd)

- **Positive correlation in a process**: big observation usually followed by another big observation, or small observation followed by small observation.
  - **Traffic traces showed strong positive correlations on a broad range of timescales**.
- Negative correlation in a process: big observation usually followed by small, or small by big.
- **There is no correlation in a process if observations are unrelated.**
  - This is the case of a Poisson process that is known to be a memoryless process.

Historical Details

- Harold E. Hurst was a hydrologist who began working on the Nile river dam project in about 1907.
- He spent 40 years studying almost 800 years of records of Nile river.
    - Hurst observed that the records of flow or levels at the Roda gauge, near Cairo, did not vary randomly, but showed series of low-flow and high-flow over years.
    - His problem was: How much discharge could be set, such that the Nile reservoir never overflowed or emptied ?
- The **Hurst parameter H is used to characterize the fractal property of a process, in our case the traffic flow**.

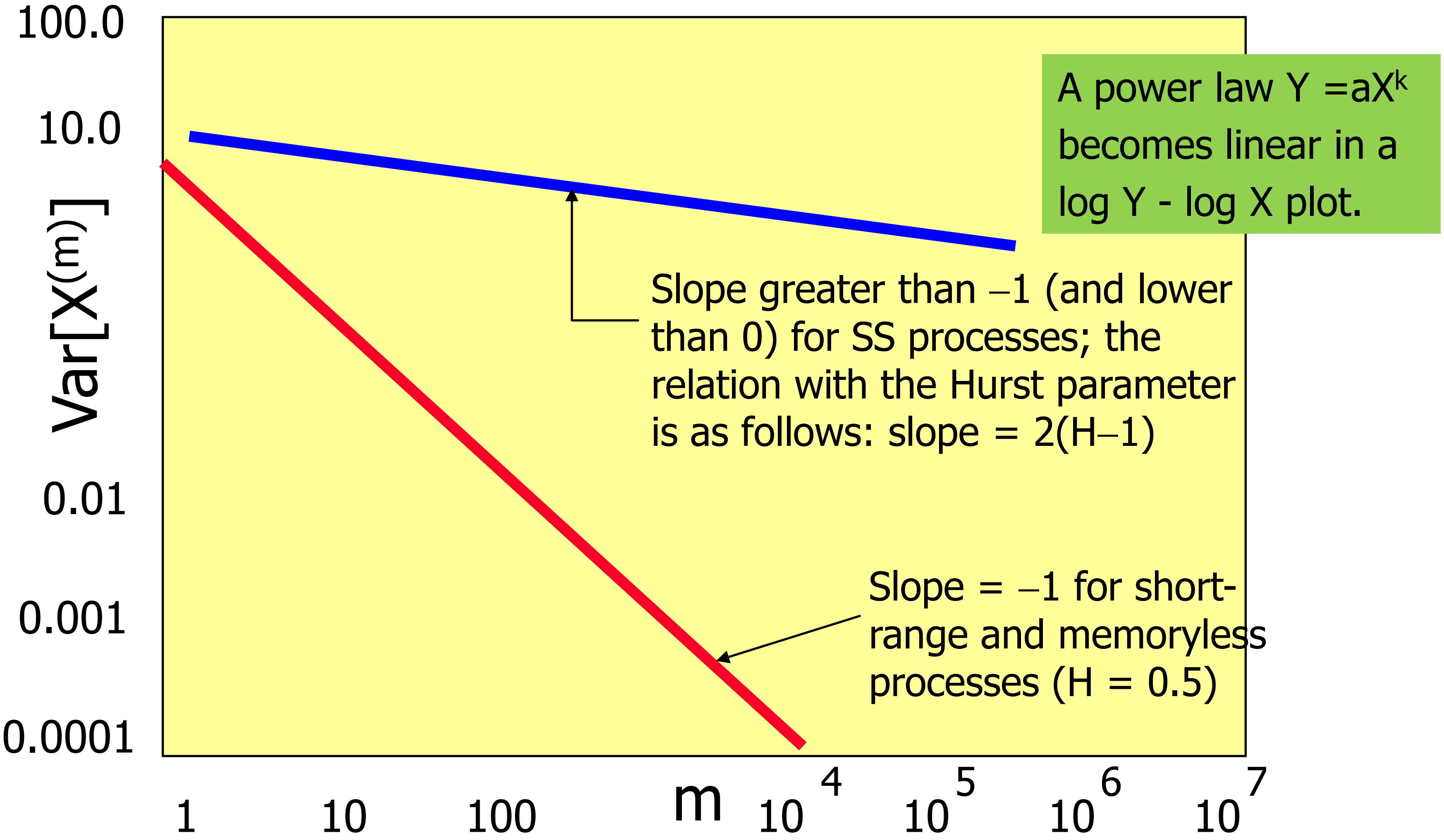
Long-Range Dependent Traffic: Definition and Characteristics

- **Self-Similarity (SS)**
  - X is self-similar if X and $m^{1-H} X^{(m)}$ have the same variance and autocorrelation. H is the **Hurst parameter**. This is the **fractal property** of the traffic. **The original traffic trace and its m-aggregations have the same bursty profile.**
- **Long-Range Dependency (LRD)**
  - $\Sigma r(k) = \infty$, i.e., **autocorrelation is not summable** [if $\Sigma r(k) < \infty$, the process has a short-range dependence].
  - Its autocorrelation $r(k) \sim k^{-\beta}$ as $k \to \infty$ (with $0 < \beta < 1$), which means the process follows a power law, rather than exponential decaying.
  - **An SS process is LRD if $0.5 < H < 1$** with **$\beta = 2(1-H)$**: self-similar processes are the simplest way to obtain LRD processes.
  - The degree of SS and LRD (autocorrelation) increases as $H \to 1$. Whereas, a **H value of 0.5 indicates the absence of long-range dependence (short range dependent processes as well as pure random processes have H = 0.5)**.
- **Heavy-tailed distributions are involved in generating LRD traffic.**

Graphical Tests for Self-Similarity

- There are different ways to graphically test the SS characteristics of a traffic trace. The method we consider here is the **variance-time plot**:
  - Rely on slowly-decaying variance of self-similar series.
  - **The variance of $X^{(m)}$ is plotted versus m on a log-log plot**
    - A slope $-\beta$ greater than –1 is indicative of SS (and LRD): $-\beta = 2(H - 1) > -1$

Variance-Time Plot: Slope Yields the Hurst Parameter

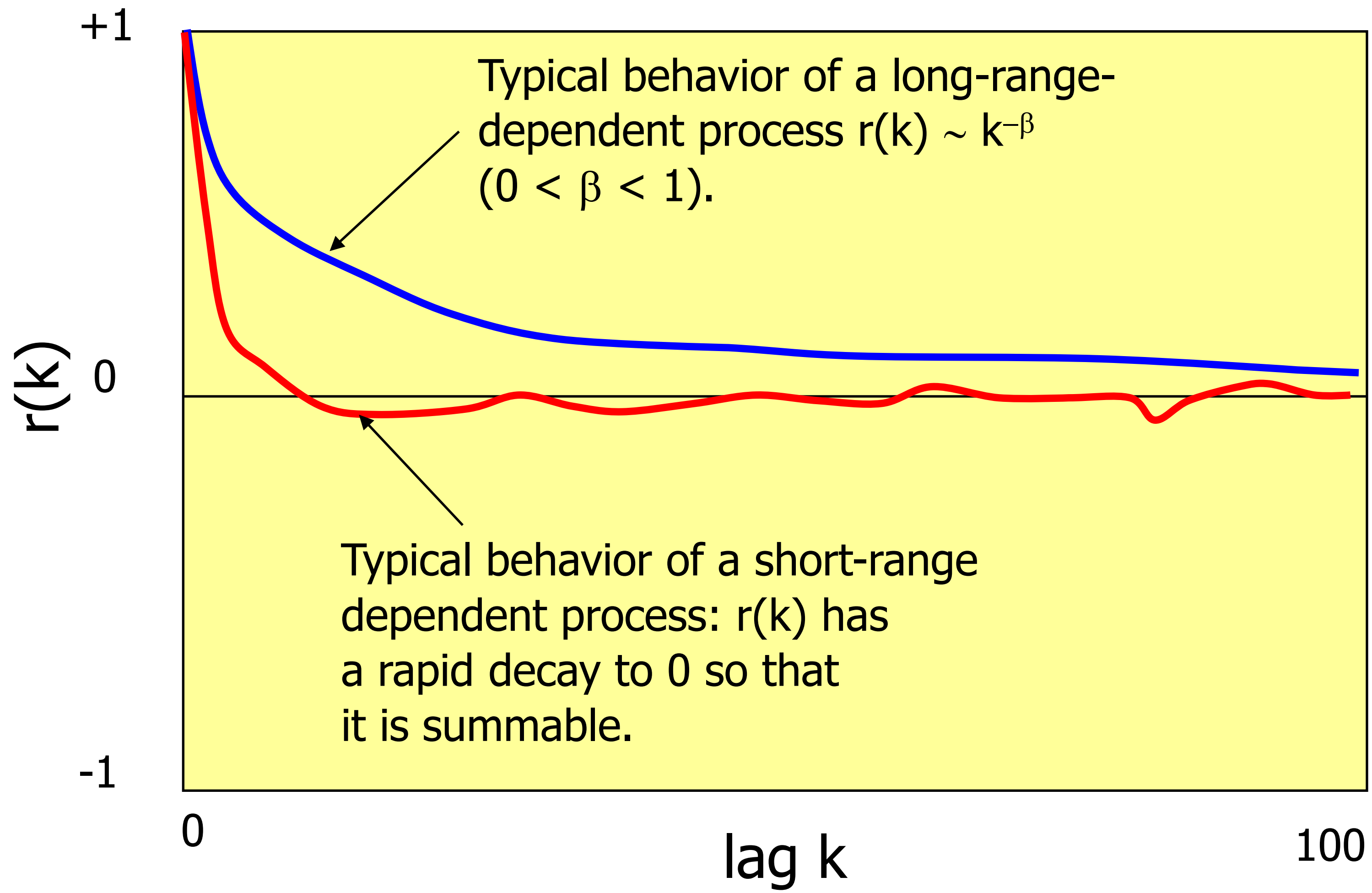


A power law $Y = aX^k$ becomes linear in a log Y - log X plot.

Slope greater than –1 (and lower than 0) for SS processes; the relation with the Hurst parameter is as follows: slope = 2(H–1)

Slope = –1 for short-range and memoryless processes (H = 0.5)

Autocorrelation Function and LRD Property



Typical behavior of a long-range-dependent process $r(k) \sim k^{-\beta}$ $(0 < \beta < 1)$.

Typical behavior of a short-range dependent process: r(k) has a rapid decay to 0 so that it is summable.

Key Concepts for LRD Traffic

- **Heavy tails in the file size** entail high variance in the transmission times.

- **High variances of transmission times** entail self-similarity and long-range dependent behavior at the session level and traffic burstiness on multiple scales of aggregation.

- **Traffic burstiness** causes higher queuing delays in the nodes of the networks (see the next example on the impact of H on the queue behavior).

Poisson vs. Self-Similar

- A Poisson process
  - appears bursty when observed on a fine time scale;
  - flattens when aggregated on a coarse time scale.
- A self-similar (fractal) process
  - **When aggregated over wide time scales (i.e., considering $X^{(m)}$ for increasing m values), this traffic maintains its bursty profile** (shaping is not effective in smoothing this traffic; a too long time is needed to reach the regime condition). **This is different from a Poisson process where the aggregation of traffic leads to an almost-constant profile.**
- For emerging future networks, the Poisson model fails to capture the traffic characteristics.

Characteristics of Network Traffic

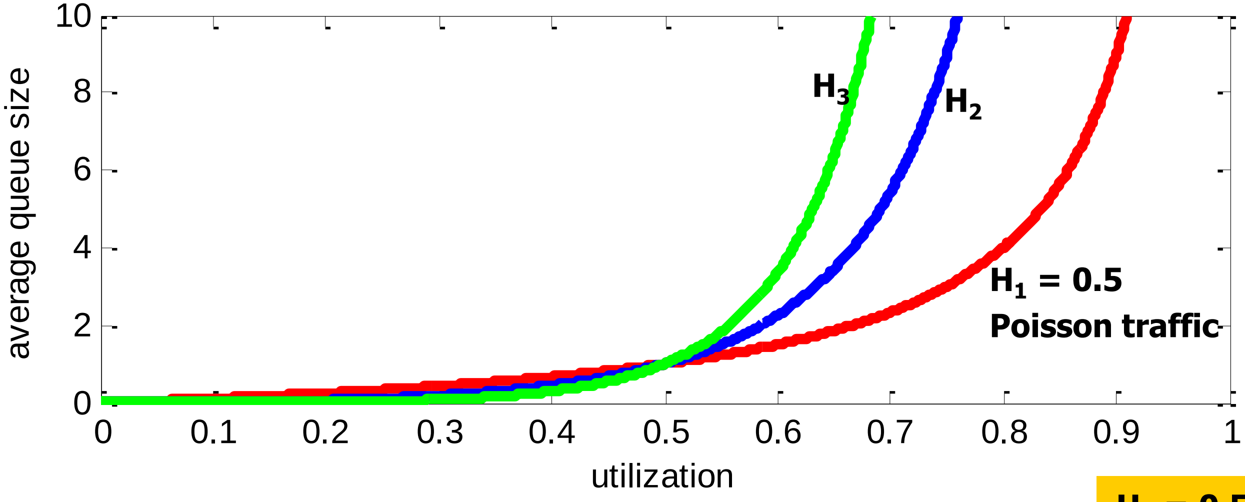
- LAN, ATM, TCP/IP, 3G core network, and video VBR traffics are both SS and LRD.
- The analysis of Ethernet traffic in terms of packets/time unit has shown that **H is between 0.8 and 0.95**.
- Isolated voice sources are, on the contrary, well described by fluid-flow ON-OFF Markov models (short-range dependent traffic).
- **The greater H, the longer the traffic correlation degree (traffic peaks last for longer intervals) and the worse the queuing performance.**

SS Traffic and Queue Analysis

- Three probability distributions play an important role in modeling SS traffic characteristics, such as **Pareto, Lognormal, and Weibull**. These distributions are heavy-tailed.
- If we like to analyze Internet congestion using queuing theory, then we have to deal with the Pareto, Lognormal, and Weibull distributions.
- **Many of the available results from queuing theory require the existence of the Laplace transform of the underlying interarrival or service time distributions**.
- **Pareto, Lognormal, and Weibull distributions do not have closed-form expressions for their Laplace transforms. This entails some problems in applying the results of the classical queuing theory.**

The Impact of H on Queue Performance

- The queue length distribution
  - Has an exponential decrease with traditional Poisson traffic (H = 0.5);
  - Decreases much more slowly with SS traffic: queue length distributions have SS and LRD characteristics.
- The network could have a bad performance if we do not take SS&LRD traffic characteristics into due account in the design phase.



$H_1 = 0.5 < H_2 < H_3 < 1$

Short-Range-Dependent Traffic Models

Web Traffic Characteristics

- As for Web traffic, we have to consider that it uses HTTP (Hyper Text Transfer Protocol) over TCP (Transmission Control Protocol).
- It is complicated to use models or real traffic traces for TCP-based traffic, since they should account for the **feedback nature of TCP with the related round trip time**.
- To overcome these difficulties, the **Markovian model** in the next slide has been proposed (ETSI, 3GPP) to describe the arrival process of IP packets due to Web downloading traffic.
- For a more recent model for Web traffic (HTTP traffic), we can consider the **PackMime-HTTP model** that has been developed in the ns-2 simulation environment. This bursty traffic generator is characterized by a single parameter that represents the connection arrival rate per HTTP source cloud. Web site: http://www.cs.odu.edu/~mweigle/research/netsim/packmime-nsdoc.pdf
- Recent **statistics on the Web traffic (2010)** are avaiable at the following Web site: https://developers.google.com/speed/articles/web-metrics

Web Traffic Characteristics

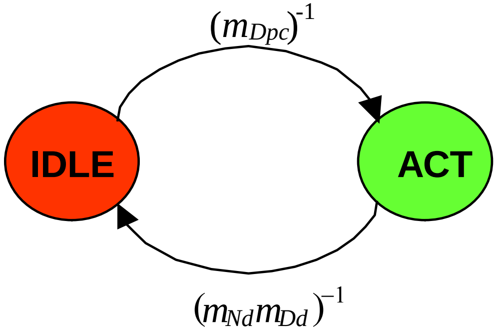
- As for Web traffic, we have to consider that it uses HTTP (Hyper Text Transfer Protocol) over TCP (Transmission Control Protocol).
- It is complicated to use models or real traffic traces for TCP-based traffic, since they should account for the **feedback nature of TCP with the related round trip time**
- To overcome these been proposed (ETS due to Web downlo
- For a more recent r **PackMime-HTTP** environment. This b parameter that repr cloud. Web site: http://www.cs.odu.edu/~mweig .ch/netsim/packmime-nsdoc.pdf

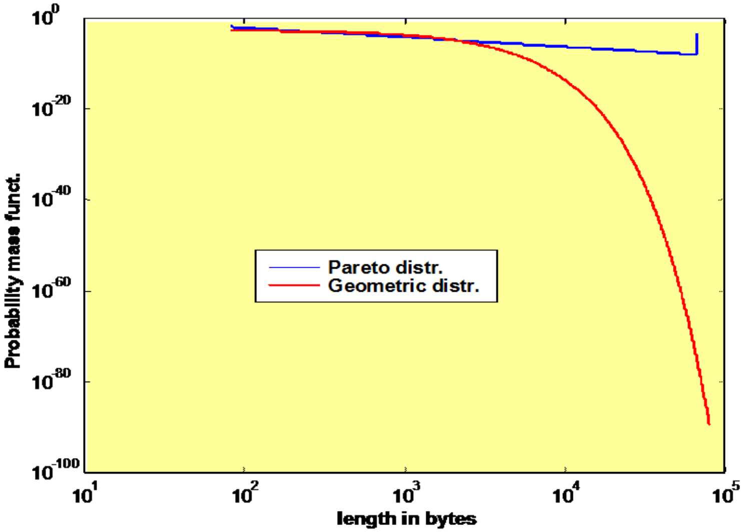
According to the data of year 2010, the average size of a Web page is 320 kB, the average number of objects (gets) per Web page is 44.5, the average number of hosts to connect to retrieve objects when accessing a Web page is 7.

- Recent **statistics on the Web traffic (2010)** are avaiable at the following Web site: https://developers.google.com/speed/articles/web-metrics

Web Downloading Traffic Model

- **Web traffic is generated according to an ON-OFF Markov model (UMTS 30.03):**
  - In the **activity phase (ACT)** traffic is generated as:
    - A geometrically-distributed number of datagrams (mean $m_{Nd} = 25$) is produced.
    - Exponentially distributed interarrival times, mean $m_{Dd} = 1/(2q)$ s; parameter q allows modulating the burstiness of the traffic source.
    - Each datagram has a truncated Pareto distribution in bytes (mean length L_w pkts).
  - In the **inactivity phase (IDLE)** no traffic is produced: the time spent in this state is exponentially distributed with mean m_{Dpc} (= 4 s).



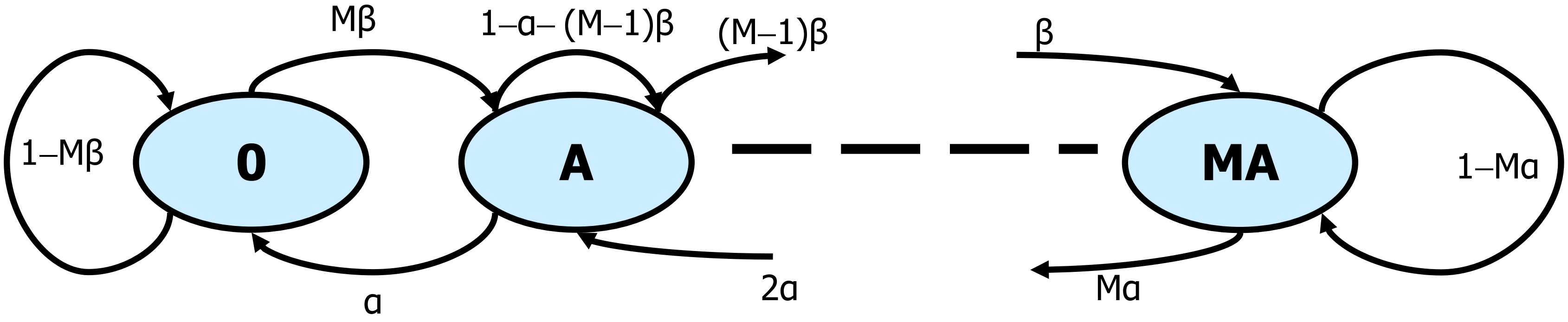


- **This is a 2-state MMPP traffic source of IP datagrams.**

A. Andreadis, G. Benelli, G. Giambene, B. Marzucchi, "A Performance Evaluation Approach for GSM-based Information Services", *IEEE Transactions on Vehicular Technology*, Vol. 52 , No. 2, pp. 313-325, March 2003.

Fluid-Flow Traffic Model for Video Sources

- Video sources (real-time, conversational or streaming class) generate traffic according to a fluid-flow model with bit-rate modulated by a discrete-time Markov chain. A time slot is used as the basis of the traffic generation.
- A video source can be obtained as the superposition of M ON-OFF minisources (a minisource in the ON state produces a constant bit-rate equal to A).



The time interval spent in ON (OFF) state by a minisource is geometrically-distributed with mean p (q).

$$\alpha = \frac{1}{p} \qquad \beta = \frac{1}{q}$$

Fluid-Flow Traffic Model for Video Sources (cont'd)

- Each mini-source in ON produces a constant bit-rate A [bit/s]:

mean bit-rate of a video source

$$A = \left(\frac{\mu}{M}\right) + \left(\frac{\sigma^2}{\mu}\right)$$

standard deviation of the bit-rate of a video source

  - We have assumed: $\mu/\sigma = 16$.
  - One IP-video packet is generated every 10 - 20 ms (video frame).
  - M = 10 or even greater.
  - All these parameters can be derived through a fitting process with a real video trace (according to different possible standards and formats).

O. Casals, C. Blondia, "Performance Analysis of Statistical Multiplexing of VBR Sources", in *Proc. of INFOCOM'92*, pp. 828-838, 1992.

Long-Range-Dependent Traffic Model

M/Pareto Model

- A **typical SS and LRD traffic source is given by the M/Pareto** model.
  - M/Pareto traffic is generated as Poisson arrivals (mean rate λ) of overlapping bursts.
  - The arrival of packets during a burst is constant for its duration with rate r packets/s.
  - The duration of each burst is a random variable, according to a Pareto distribution (parameter γ) with finite mean and infinite variance in order to have **heavy tails**.
- **This traffic model corresponds to an M/G/∞ system (Poisson arrivals of bursts/General burst duration/infinite bursts can be simultaneously present).**

M/Pareto Model (cont'd)

- Let λ denote the mean arrival rate of bursts according to a Poisson process:

$$\text{Prob}\{A(t)=n\}=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

- The Pareto distribution has complementary distribution as:

$$\text{Prob}\{X>x\}=\begin{cases}\left(\frac{x}{\delta}\right)^{-\gamma}, & x\geq\delta\\ 1, & \text{otherwise}\end{cases}$$

if $0 < \gamma < 2$ (and $\delta > 0$), the Pareto distribution is heavy-tailed and we need to have $1 < \gamma < 2$ in order for this traffic source to generate self-similar traffic with Hurst parameter in (0.5, 1).

- The mean of X is $\delta\gamma/(\gamma-1)$ s and its variance is infinite.
- The mean number of packets within one burst is $r\delta\gamma/(\gamma-1)$.
- The mean traffic produced by an M/Pareto source is $r\delta\gamma\lambda/(\gamma-1)$ in packets/s.
- The M/Pareto model generates SS traffic with Hurst parameter $H = (3-\gamma)/2$.

Final Comments on Traffic Models

- FTP session arrivals are well modeled by Poisson processes.
- A number of WAN traffic characteristics are well modeled by heavy-tailed distributions.
- The packet arrival process for typical Internet applications as well as the aggregate Internet traffic is self-similar.

T. B. Fowler. "A Short Tutorial on Fractals and Internet Traffic", *Telecommunication Review*, Vol. 10, 1999.

Thank you!

giovanni.giambene@gmail.com