Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.829 Fall 2001 **Traffic Self-Similarity** December 3 2001

Overview. Self-similar nature of traffic.

1 Stochastic process background

A stochastic process X can be thought of as a (potentially infinite) sequence of values drawn from some random distribution. I.e., X is a stochastic process defined as $X = \{X_1, X_2, ..., X_n, ...\}$ (E.g., think of X_i as the bandwidth in packet/second at time i on an Ethernet.)

X is *covariance stationary* or *wide-sense stationary (WSS)* iff it can be completely characterized by three things:

- Its mean μ (the mean of the distribution it's drawn from). The mean doesn't change with time if the process is WSS.
- Its variance, σ^2 (which is defined as $E[X^2] - (E[X])^2$). The variance doesn't change with time for a WSS process.
- Its autocorrelation function $r(i, j)$ is a function only of $|j - i|$.

The autocorrelation function r$(i, j) = E[(X(i) - mu) * (X(j) - mu)]$. That is, $r(i, j)$ measures how correlated two time-series samples are across time. In the limiting case when $i = j$, you get the variance of the process. For $i \neq j$, this tells you how much impact what you saw at time i (X_i) has at time j.

The beauty of WSS is that it is "time-invariant" as far as the autocorrelation is concerned: it doesn't matter what the actual time values i and j are, it *only* matters how much further into the future (or past!) j is compared to i.

Many natural processes are WSS. This paper *assumes* that Ethernet bandwidth consumption is WSS and then proceeds from there.

It is important to distinguish a WSS process from a *stationary stochastic process.* The latter is a process where the X_i's are drawn from precisely the same distribution for all time instants i. That is, $F(x_1, x_2, \ldots, x_k) = F(x_{i+1}, x_{i+2}, \ldots, x_{i+k}) \forall i, k$. Here, F is the joint cumulative distribution function of the time series.

All stationary processes are WSS. The converse is *not* true. Only the mean and variance need to be time-invariant for WSS, in addition to $r(i, j)$ depending only on the time difference $|j - i|$ between the observations.

2 Intuition and main contributions of paper

A self-similar process "looks the same" across multiple time scales spanning many orders of magnitude. Network traffic seems to show this across four or five orders of magnitude (e.g., beyond the several hour time-scale, there is clear evidence of daily periodicity caused by human work patterns). At time scales below this, the analysis shows burstiness across aggregates, in contrast to aggregates of Poisson arrivals that quickly smooth out over time.

Intuitively, in a self-similar process, there's a "lot of memory." The autocorrelation, which decays quickly as a function of the time difference k for a "normal" process, decays rather slowly for a self-similar process. $r(k)$ goes as $k^{-\beta} 0 < \beta < 1$. (This is not even as fast as $1/k$.) And in addition, if you block together sequences of X_i's and treat them as one (with the average value used) to get a new stochastic process (called the "blocked process"), the autocorrelation *remains the same* asymptotically! I.e., aggregating into bigger time chunks does nothing to make the system "forget."

This paper makes three important contributions:

- Analysis of large-scale Ethernet traces. Impressive, meticuluous data collection and systematic analysis. Data available at the Internet Traffic Archive (ITA).
- The demonstration of self-similarity and fractal behavior showing long-range dependence of traffic patterns.
- Demonstration of the failure of Poisson modeling of aggregate traffic behavior.

3 Formalization

Suppose that X has an autocorrelation function of the form $r(k) \approx k^{-\beta}, 0 < \beta < 1$. Now, consider the *blocked process* of X. Parametrized by a number m, it is obtained by taking non-overlapping blocks of m consecutive samples from X and replacing each block by one average number, the average of the m blocked samples. Call this new process $X_k^{(m)}$.

X is *self-similar* if two conditions hold:

- The variance of the blocked process is a slowly decaying function of m, the size of the aggregate block. More specifically, $\mathrm{Var}(X_k^{(m)}) = \sigma^2 m^{-\beta}$, and
- The autocorrelation of the blocked process is the same as the original process: $\lim_{m\to\infty} r^{(m)}(k) = r(k)$.

One more definition before we get to the implications and experimental analysis: the *Hurst parameter H* is defined as $1 - \beta/2$.

4 Implications

What are some mathematical implications of self-similarity of a stochastic process? There are two worth noting:

- Slowly decaying variance. The variance of the sample mean decays very slowly, slower than the reciprocal of the sample size. What this says is that the process has a *heavy tail*, since the tail probabilities don't fall off sharply (large variance).

- Autocorrelations decay slowly. The autocorrelation decays hyperbolically rather than exponentially fast. Because $r(k)$ goes as $k^{-\beta}, \sum_k r(k)$ diverges to ∞. This is what is meant by *long-range dependence* or LRD.

Typical Poisson-based models, in aggregate, tend to second-order white noise, where $r(k)$ quickly decays to 0. ("White" or "Gaussian" noise is a WSS process with $r(k) = 0$ for $k \neq 0$.) Aggregates of long-range dependent traffic are very different.

5 Experimental tests for self-similarity

- Method 1: Variance-time plots.
  1. For each block size m calculate the mean $X_k^{(m)}$ and find its variance.
  2. Plot this vs. m on a log-log scale.
  3. For large values of m, fit the best straight line and calculate its slope. If the process is self-similar, the slope will be $-\beta = -2(1 - H)$ (where H is the Hurst parameter introduced previously).

- Method 2: R/S statistic.
  1. Calculate S^2, the sample variance of the stochastic process X.
  2. Calculate $R_n/S_n = \max[0, W_1, \ldots, W_n] - \min[0, W_1, \ldots, W_n]$, where $W_i = X_1 + \ldots + X_i - i\mu$ (μ is the sample mean).
  3. $E[R_n/S_n]$ turns out to be n^H for a self-similar process, where H is the same Hurst parameter introduced before. $0.5 < H < 1$ for a self-similar process.
  4. Plot $\log R_n/S_n$ vs. $\log n$ and take the slope.

6 Experimental results

The key result of the paper is that **Ethernet traffic is self-similar.** The "proof" is by meticulous empirical validation using three independent tests of four different huge data sets. Very thorough analysis.

They observe LRD ("fractal") behavior over many (but not all) time scales. The more loaded the Ethernet, the larger the Hurst parameter H and the "higher" the long-range dependence.

Why does this happen? It turns out that if you have a large number of individual sources that are each "ON/OFF" sources with *heavy-tailed* active (ON) and inactive (OFF) periods, the resulting aggregate is long-range dependent.

7 Why should we care?

This is the hardest part to discuss. Here are some reasons why we should care:

- Deeper understanding. It is *always* worth the effort to understand phenomena deeply and further our state of understanding of a complex system. Even if it isn't immediately obvious today, it may well be that this deeper understanding leads to big practical impact in the future. Mathematics and physics are replete with examples of such findings.
- This analysis clearly shows the failure of Poisson modeling, the method hitherto used and recognized as appropriate.
- This analysis shows the failure of conventional notions and metrics of "burstiness," which do not comprehend infinite variance and LRD.
- It might influence the design of buffer sizes and management schemes in switches and routers, although (in my opinion) this seems rather unlikely to make much practical impact in this area.