eecs589

Advanced Computer Networks

Leland et al., "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking*, 2(1):1–15, Feb. 1994

References

[ENW96] Erramilli, Narayan, and Willinger, "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic," *IEEE/ACM ToN*, 4(2):209–223, Apr. 1996

[W+97] Willinger, Taqqu, Sherman, and Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," *IEEE/ACM ToN*, 5(1):71–86, Feb. 1997

Self-Similarity

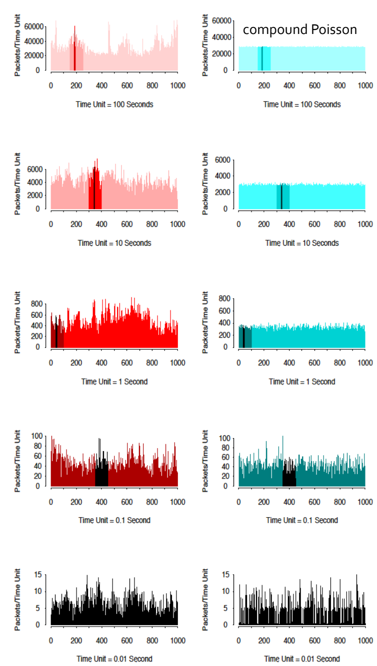
Viewing scale not apparent from object appearance

Object features are statistically similar between object parts and the overall object

For example, we always see a jagged line no matter how close we look at a coastline

Ethernet Traffic …

Absence of a natural length of a "burst": at every time scale from msecs to minutes and hours, bursts consist of bursty subperiods separated by *less bursty* subperiods



Comparable average packet size and arrival rate

Drastically different distribution across time scales

[W+97]

A Traffic Trace

Let X be a covariance stationary stochastic process

What is:

- a stochastic process:
  a time series of a variable that changes randomly
- a stationary process:
  a time series whose statistical properties: mean, variance, autocorrelation, stay constant over time
- covariance:
  by how much two random variables move in tandem

By how much X_t and X_{t+h} move in tandem is not a function of time

Sample Mean

Let X be a covariance stationary stochastic process

$X^{(m)}$, a sample mean, is a new covariance stationary time series obtained by averaging the original series X over non-overlapping blocks of size m

X: 1, 3, 6, 2, 5, 1, 8, 3, 2, . . .

$X^{(2)}$:

$X^{(3)}$:

$X^{(4)}$:

Self-Similarity

A traffic trace is self-similar if (equivalently):

1. the variance of the sample mean remains large even as you sample at larger and larger samples (no smoothing out):

$$\mathrm{var}(X^{(m)}) \sim a_2 m^{-\beta},\ m \to \infty, 0 < \beta < 1, a_2 > 0$$

vs.

$$\mathrm{var}(X^{(m)}) \sim a_4 m^{-1},\ m \to \infty, a_4 > 0$$

Self-Similarity

A traffic trace is self-similar if (equivalently):

2. auto-correlation functions (acf) at various time scales are of the form: $r(k) \sim k^{-\beta},\ k \to \infty, 0 < \beta < 1$

⇒ the acf is non-summable, i.e., traffic exhibits long range dependence (LRD): $\sum_k r(k) \to \infty$

Short-range dependence: $r(k) \sim \rho^k,\ 0 < \rho < 1 \Rightarrow \sum_k r(k) < \infty$

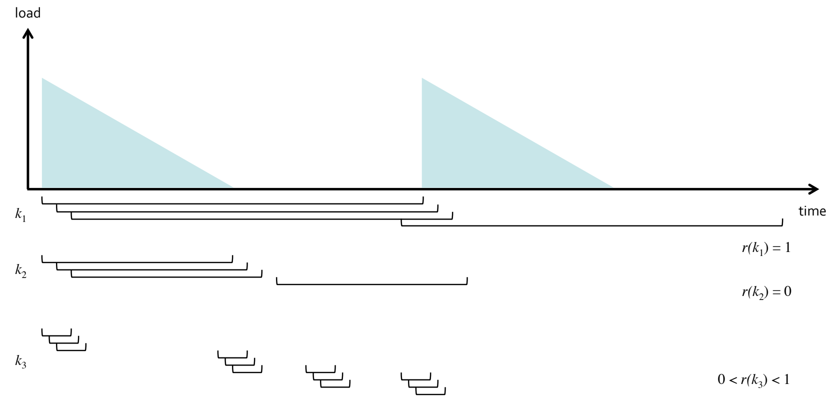
X is (exactly) second-order self-similar if: $r^{(m)}(k) = r(k),\ k \geq 0$

and (asymptotically) second-order self-similar if:

$$r^{(m)}(k) \to r(k),\ m \to \infty$$

White noise (not self-similar): $r^{(m)}(k) \to 0,\ m \to \infty$

Auto-correlation

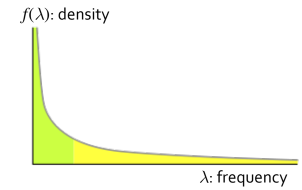


Self-Similarity

A traffic trace is self-similar if (equivalently):

3. taking the time series into the frequency domain (Fourier transform), the low frequency components obeys a power-law near the origin (a low frequency is proportionally denser than its next higher frequency):

$$f(\lambda) \sim a_3 \lambda^{-\gamma},\ \lambda \to 0, 0 < \gamma < 1, a_3 > 0$$
$$\gamma = 1 - \beta$$



Cf. Zipf distribution

Self-Similarity

A traffic trace is self-similar if (equivalently):

4. the expected *rescaled adjusted range statistic*:

$$\mathbb{E}[R(n)/S(n)] \sim a_5 n^H,\ n \to \infty, a_5 > 0$$

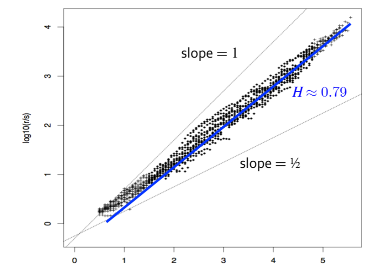
has Hurst parameter $½ < H < 1$

The Hurst parameter expresses the speed of decay of the acf

- $H \leq ½$: short-range dependent processes, e.g., Poisson, batch-Poisson, Markov-modulated Poisson
- $H > 1$: non-stationary process

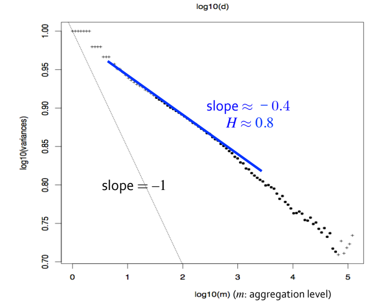
Detecting LRD in Ethernet Trace

H can estimated directly from R/S statistic



Or from variance time plot

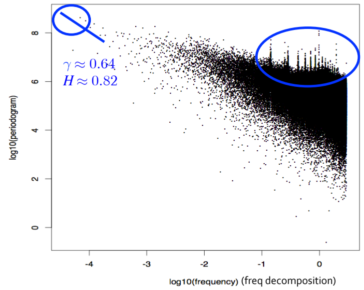
- slope $-\beta, 0 < \beta < 1$
- $H = 1 - \beta/2$



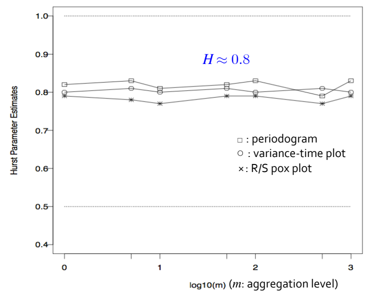
Detecting LRD in Ethernet Trace

Or from periodogram, slope of 10% of the lowest frequencies, near 0

$H = (1+\gamma)/2$

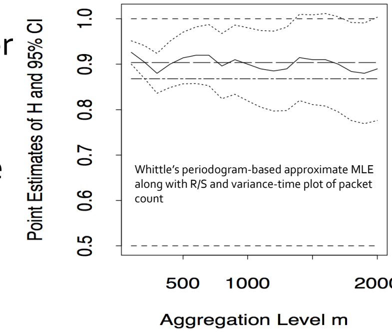


Hurst parameter stays constant across traffic aggregation levels



Detecting LRD in Ethernet Trace

H can also be estimated with maximum-likelihood estimator (MLE) based on the periodogram (Whittle estimator) with the advantage of computing 95% confidence interval



Detecting LRD in Ethernet Trace

H increases as traffic load increases!

Remain true over time (89-92)

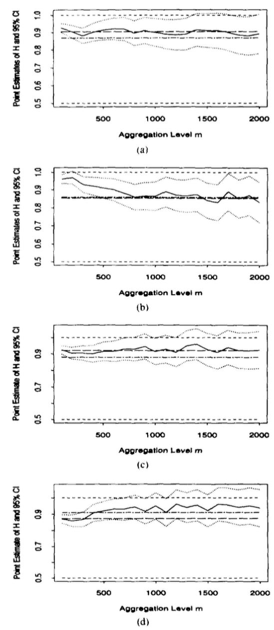


Fig. 6. Periodogram-based MLE/aggregation method for the sequences AUG89.MP, OCT89.MP, JAN90.MP, and FEB92.MP.

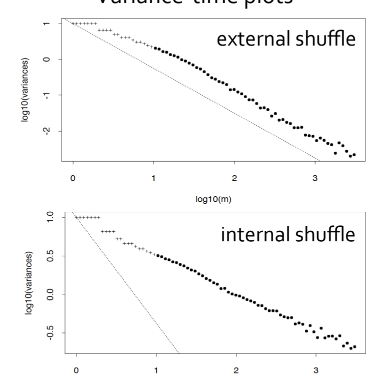
Implication

Long-range dependent traffic effects queueing delay: it makes buffer sizing ineffectual

How do we know LRD causes ineffectual buffering [ENW96]?

- external shuffle experiment: divide traffic into m blocks and shuffle the blocks around preserving the sequence inside each block: destroys LRD, preserves SRD
- internal shuffle experiment: same blocks, shuffle traffic inside each block, keeping the block order: destroys SRD, preserves LRD

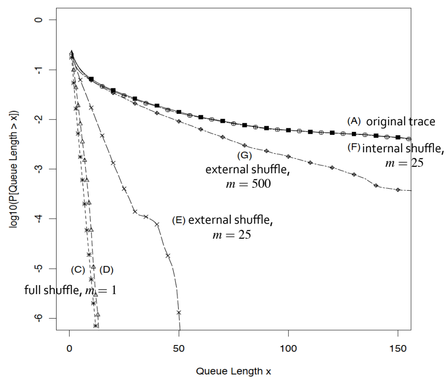
Variance-time plots



[ENW96]

Implication

Resulting queue occupancy statistics:



What can we do about it?

- frequency domain view: traffic can be decomposed into high (spikes), mid (ripples), and low (swells) frequencies
- network must have enough capacity to handle peak rate of low frequency
- buffer space should be used only to handle high-frequency traffic

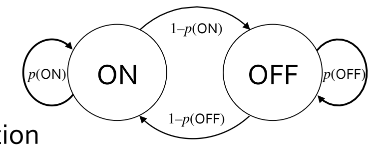
[ENW96]

Causes of LRD

Aggregation of ON/OFF traffic with heavy-tailed OFF time distribution [W+97]

- human "think" time
- effect of TCP congestion avoidance (`cwnd`)
- multimedia sources can also be modeled as ON/OFF

Why does long-tailed ON/OFF distributions cause LRD?

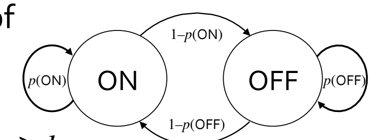


- long OFF time means autocorrelation of bursts at large k, hence $\sum_k r(k) \to \infty$
- long ON time increases the probability of seeing other traffic

Heavy-tailed Distributions

$P[X > x] \sim x^{-\alpha}, x \to \infty,$ $0 < \alpha < 2$

Examples: Pareto, Weibul, Zipf



Pareto distribution:

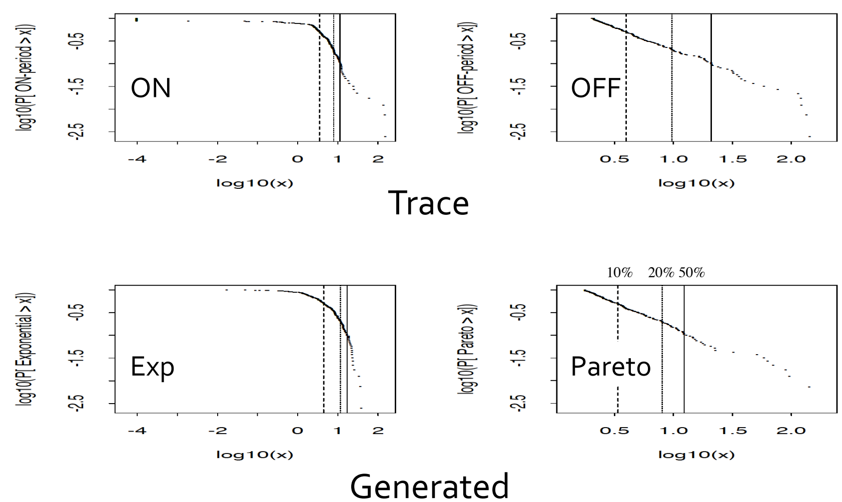
$p(x) = \alpha k^{\alpha} x^{-\alpha-1}$, α and $k > 0$, $x \geq k$

$F(X) = P[X \leq x] = 1 - (k/x)^{\alpha}$

$\alpha < 2$: trace has infinite variance

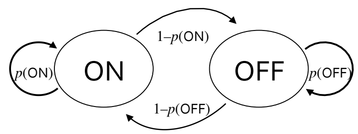
$\alpha < 1$: trace has infinite mean

Heavy-tailed Distributions



[W+97]

Modeling LRD



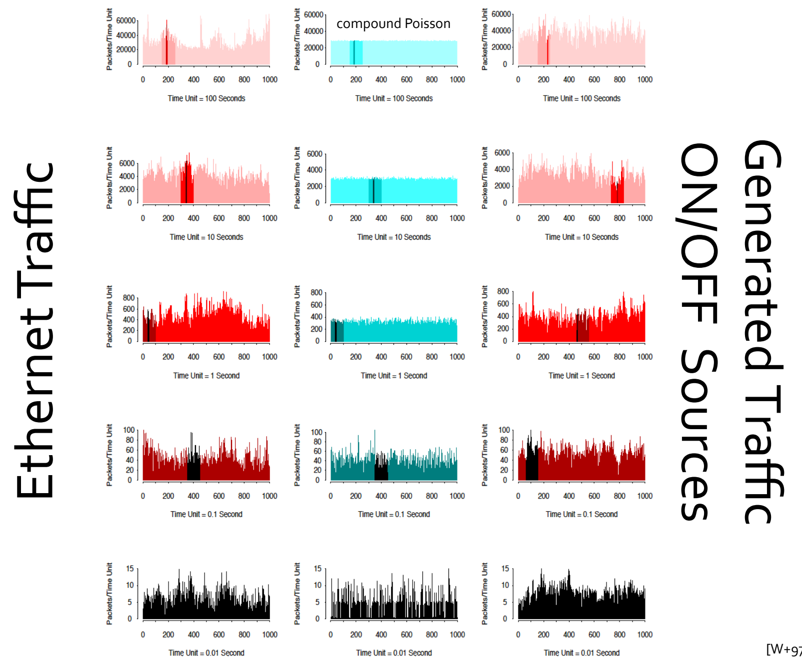
Aggregation of ON/OFF traffic with heavy-tailed OFF time distribution [W+97]

Advantage: parsimonious, only one parameter, α

$$P[X > x] \sim x^{-\alpha}, x \to \infty, 0 < \alpha < 2$$

Alternative models:

- by fitting multiple short-range dependent processes: parameter explosion, no physically meaningful interpretations
- fractional Gaussian noise: does not model short-range dependencies
- fractional ARIMA:
  - can model both short- and long-range dependencies,
  - but still does not provide physical explanation of self-similarity
  - plus, known parameter estimation techniques too expensive



[W+97]

Discussions

Gold standard of measurement study and analysis

Prior to this paper, traffic modeling assumes Poisson distribution

After this paper, traffic modeling uses power law distributions (Pareto, Weibul, Zipf)

A flurry of follow-on papers found power-law distribution everywhere in the network . . .