Self-similar Network Traffic

The notions and effects of
self-similarity and
long-range dependence

Glen Kramer
kramer@cs.ucdavis.edu

Discrepancy between experiment and analysis

- Markovian models (Poisson)
  - Accurately describe voice traffic
  - Allow tight bounds on performance parameters

Is Poisson model valid for data traffic?

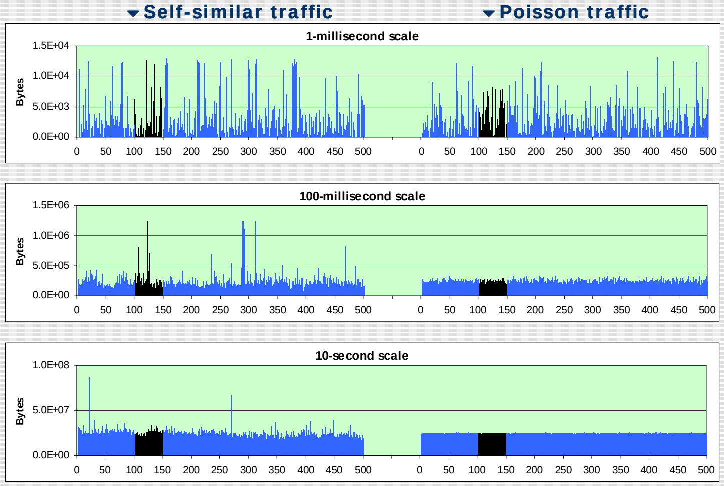
- Observed:
  - Traffic is bursty at many timescales
  - Burstiness increases as number of active sources increases (contrary to Poisson)

Discovery of traffic self-similarity

- Leland et al. – self-similarity in LAN
  W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "*On the Self-Similar Nature of Ethernet Traffic (Extended Version)*," IEEE/ACM Transactions on Networking, 2(1), pp. 1-15, February 1994.
- Paxson and Floyd – self-similarity in pre-Web WAN
  V. Paxson and S. Floyd, *"Wide-area traffic: the failure of Poisson modeling,"* IEEE/ACM Transactions on Networking 3, pp. 226-244, 1994.
- Crovella and Bestavros –self-similarity in modern WAN
  M. Crovella and A. Bestavros, *"Self-similarity in World Wide Web traffic: evidence and possible causes,"* IEEE/ACM Transactions on Networking 5, pp. 835-846, 1996.

What is self-similar traffic?

Self-similar traffic **Poisson traffic**



Definitions

$Y(t)$ – cumulative process (packet/byte arrivals up to time t)

X_t – incremental process of $Y(t)$:

$$X_t = Y(t+1) - Y(t)$$

$X_s^{(m)}$ – aggregated process of X_t:

$$X_s^{(m)} = \frac{1}{m}[X_{sm-m+1} + X_{sm-m+2} + ... + X_{sm}]$$

$\gamma(k)$ – autocovariance function

$$\gamma(k) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

5/21/2001 UC Davis 5

Definition of self-similarity 1

- Part resembles the whole
- Self-similarity is too restrictive: original and scaled aggregated processes should be undistinguishable.
- Usually consider second-order self-similarity: autocovariance function $\gamma(k)$ of the original and scaled process is the same.
- Process is exactly (second-order) self-similar if

$$\gamma^{(m)}(k) = \gamma(k)$$

- Process is asymptotically (second-order) self-similar if

$$\lim_{m\to\infty} \gamma^{(m)}(k) = \gamma(k)$$

5/21/2001 UC Davis 6

Definition of self-similarity 2

- Traffic is self-similar with parameter *H* (0 < *H* < 1) if for all *k* > 0 and *t* ≥ 0,

$$Y(t) \overset{\Delta}{=} k^{-H} Y(kt) \qquad (H \text{ – Hurst parameter})$$

- We are interested in X_t behavior (discrete time process)

Considering X_t to be a stationary increment process, $X^{(m)}$ can be viewed as a sample mean. Then,

$$X^{(m)} = \frac{1}{m}\sum_{t=1}^{m} X(t) = \frac{1}{m}\left[Y(m) - Y(0)\right] = \frac{m^H}{m}\left[Y(1) - Y(0)\right] = m^{H-1} X$$

Definition of self-similarity 3

- Process is exactly (second-order) self-similar if

$$X^{(m)} = m^{H-1} X$$

- Process is asymptotically (second-order) self-similar if

$$\lim_{m\to\infty} X^{(m)} = m^{H-1} X$$

Definition of Long Range Dependence

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

- autocorrelation function

For $0 < H < 1,\ H \neq \frac{1}{2}$ $\quad r(k) \sim H(2H-1)k^{2H-2}, \quad k \to \infty$

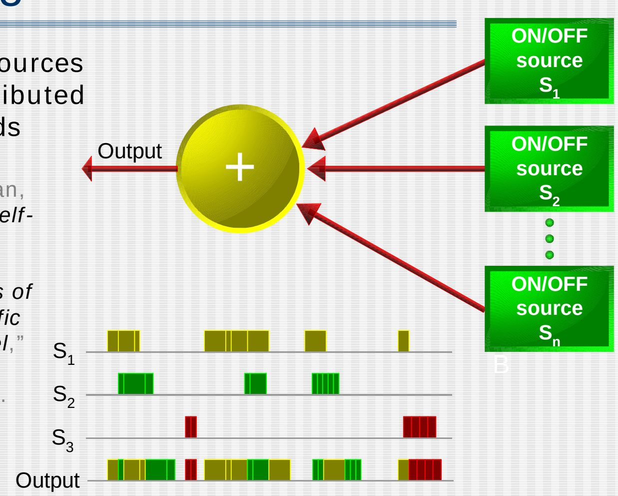
If $\frac{1}{2} < H < 1$, then $\sum_{k=-\infty}^{\infty} r(k) = \infty$ - property of LRD

Long Range Dependence ≠ Self-similarity, but for ½<H<1, process is both LRD and self-similar

Traffic generation model

Aggregate *n* sources of Pareto-distributed ON/OFF periods

W. Willinger, M. Taqqu, R. Sherman, and D. Wilson. "*Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level*," In Proc. ACM SIGCOMM '95, pp. 100-113, 1995.



Pareto distribution

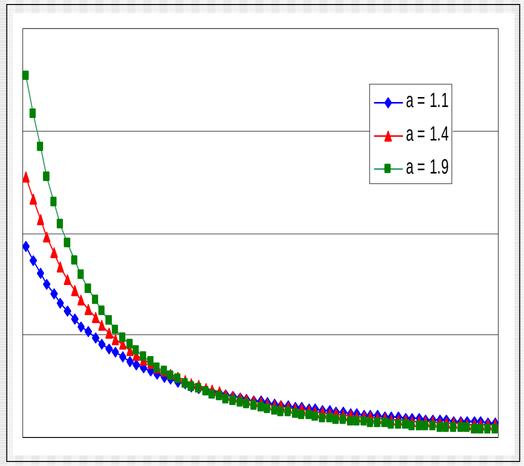
$$P(x) = \frac{\alpha b^{\alpha}}{x^{\alpha+1}} \quad , \quad x \geq b$$

α - shape parameter

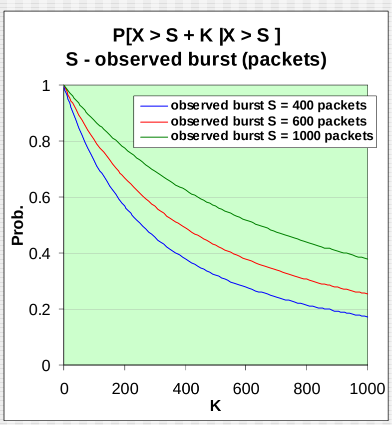
$(1 < \alpha < 2)$

b - location parameter

$$E(x) = \frac{\alpha b}{\alpha - 1}$$



Traffic is not memoryless



$$P[X > S + K \mid X > S] = \frac{P[X > S + K]}{P[X > S]}$$

$$= \frac{\int_{S+K}^{\infty} \frac{\alpha b^{\alpha}}{x^{\alpha+1}} dx}{\int_{S}^{\infty} \frac{\alpha b^{\alpha}}{x^{\alpha+1}} dx} = \frac{\left(\frac{b}{S+K}\right)^{\alpha}}{\left(\frac{b}{S}\right)^{\alpha}} = \left(\frac{S}{S+K}\right)^{\alpha}$$

Thus, probability that current burst will continue for K packets depends on how many packets this burst delivered so far (S), i.e., it depends on burst history.

Larger bursts are more likely to continue !!!

Truncated-tail Pareto distribution

$X_{PARETO} = \frac{b}{U^{1/\alpha}}$, U - uniform r.v. $(0 < U \le 1)$

But computers generate discrete values

Denote S - smallest non-zero U, then $Q = \frac{b}{S^{1/\alpha}}$ - largest value (tail truncation)

Find truncated p.d.f. $f_T(x)$: $\int_b^Q f_T(x)\,dx = 1 \;\Rightarrow\; f_T(x) = \frac{f(x)}{\int_b^Q f(x)\,dx} = \frac{\alpha b^\alpha}{x^{\alpha+1}} \times \frac{1}{1-\left(\frac{b}{Q}\right)^\alpha}$

Then $E(X) = \int_b^Q x\, f_T(x)\,dx = \frac{\alpha b^\alpha}{1-\left(\frac{b}{Q}\right)^\alpha} \times \left.\frac{x^{1-\alpha}}{1-\alpha}\right|_b^Q = \frac{\alpha b}{\alpha-1} \times \frac{1-\left(\frac{b}{Q}\right)^{\alpha-1}}{1-\left(\frac{b}{Q}\right)^\alpha} = \frac{\alpha b}{\alpha-1} \times \frac{1-S^{\frac{\alpha-1}{\alpha}}}{1-S}$

Traffic parameters

ON-periods are Pareto-distributed with $\alpha_{ON} = 1.4$ (s.t., $H = (3\text{-}\alpha)/2 = 0.8$)
and $b_{ON} = 1$ packet size.

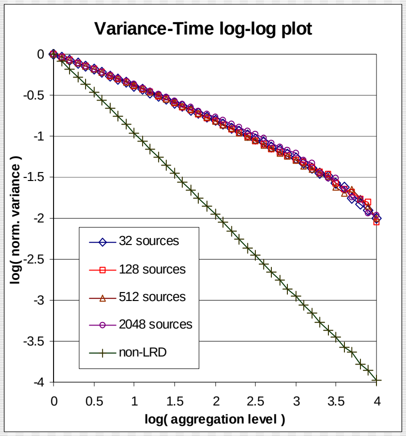
OFF-periods are Pareto-distributed with $\alpha_{OFF} = 1.2$
and b_{OFF} calculated based on desired load:

Total load $L = \sum_{i=1}^{N} L_i \;\Rightarrow$ (for identical sources) $L_i = L/N = \frac{E[ON]}{E[ON]+E[OFF]}$

$E[OFF] = E[ON] \times \left(\frac{1}{L_i} - 1\right) \;\Rightarrow\; \frac{\alpha_{OFF} b_{OFF}}{(\alpha_{OFF}-1)} \times \frac{1-S^{\frac{\alpha_{OFF}-1}{\alpha_{OFF}}}}{1-S} = \frac{\alpha_{ON} b_{ON}}{(\alpha_{ON}-1)} \times \frac{1-S^{\frac{\alpha_{ON}-1}{\alpha_{ON}}}}{1-S} \times \left(\frac{1}{L_i} - 1\right)$

From here find $b_{OFF} = b_{ON} \times \frac{\alpha_{ON}}{\alpha_{OFF}} \times \frac{\alpha_{OFF}-1}{\alpha_{ON}-1} \times \frac{1-S^{\frac{\alpha_{ON}-1}{\alpha_{ON}}}}{1-S^{\frac{\alpha_{OFF}-1}{\alpha_{OFF}}}} \times \left(\frac{1}{L_i} - 1\right)$

Self-similarity verification



$$X^{(m)} = m^{H-1} X$$

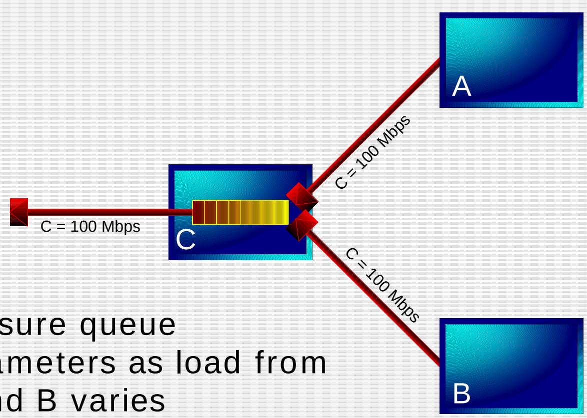
$$\Rightarrow \quad \mathrm{Var}(X^{(m)}) = m^{2(H-1)} \, \mathrm{Var}(X)$$

$$\Rightarrow \quad \log\left(\frac{\mathrm{Var}(X^{(m)})}{\mathrm{Var}(X)}\right) = (2H-2) \times \log(m)$$

Thus, expect slope *2H-2.*

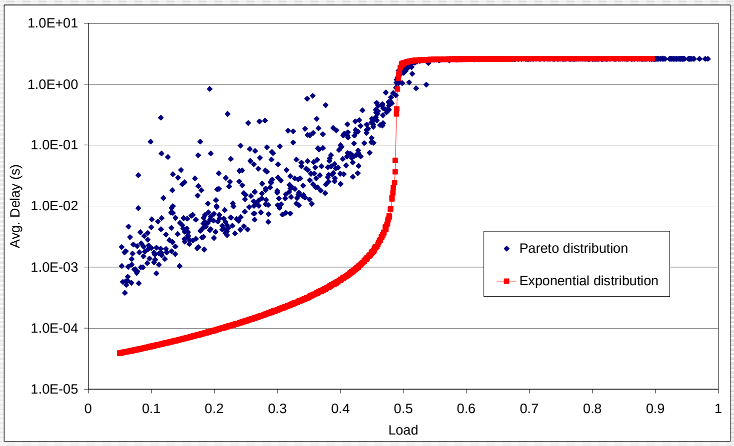
H = 0.8 => slope *= -0.4*

Simple experiment



Measure queue parameters as load from A and B varies

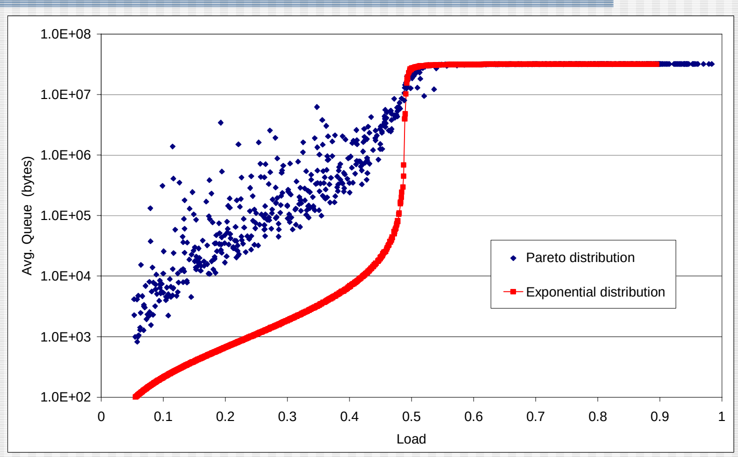
System performance (latency)



Each point is an average of 10 mln.(!) packets

5/21/2001 UC Davis 17

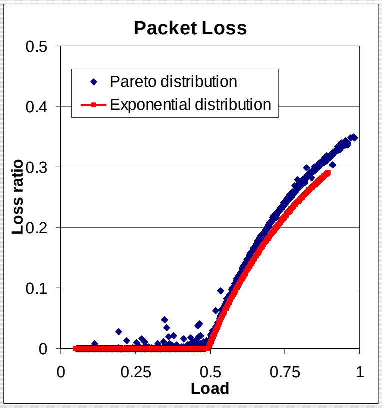
System performance (queue size)

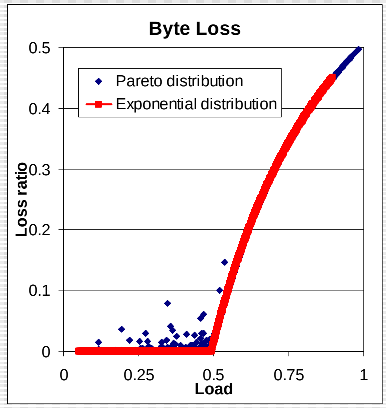


Each point is an average of 10 mln.(!) packets

5/21/2001 UC Davis 18

System performance (packet loss)





Question of the day: **why at maximum load (2x100Mbps) only 35% of packets lost?**

Conclusion

n^{th} moment of Pareto (any heavy-tail) distribution has form $\int^{\infty} x^n \frac{c}{x^\beta} dx$. Therefore, no moments above $n = \lfloor \beta - 1 \rfloor$ exist. For Pareto distribution the mean is finite, but the variance is infinite

=> cannot build closed form analytical expression

=> must use simulation

=> it is important to use self-similar traffic to get realistic results