On the Limits of Self-Similarity in Network Traffic Models

Mart Molle

Amir Dabirian

UC Riverside

Self-Similar Traffic Models

- 1993, Leland *et al*. introduced the concept of *self-similarity* or *long-range dependence* in Ethernet traffic models
- Many others found the effect in other networks
- 1996, Erramilli *et al*. presented some experiments to show its effect on queueing models of network devices

Outline of this Talk:

- Review Erramilli's experimental methodology
- Show that results obtained in this way are *sensitive to network load* in the trace file
- Explain shuffling by *fixed time intervals*, not fixed sample counts
- Demonstrate that there is a *limited range* to the dependence

Erramilli's Approach

- Start with trace of *packet inter-departure times* $t_1, t_2, t_3, \ldots$
- *Ignore packet service times (length)* and replace by a *fixed* value v.
- Vary v to obtain a delay-throughput curve
- Compare curve to queueing formula with i.i.d. arrivals with similar moments

Block Shuffling by Sample Count

- Divide trace into blocks of *N* consecutive samples
- *Internal shuffle* randomly reorders inter-departure times within each block, but preserves the order of the blocks
- *External shuffle* preserves the ordering within each block, but randomly reorders the blocks

Effects of Shuffling by Counts

- External shuffles with N=25 points has a large effect on the delay-throughput curve
- The effect is still visible, even with N=500
- Internal shuffle with N=25 points has almost no effect on the curve

Bias from Shuffling by Counts

- Network traffic is bursty
  - transmissions are clustered into *busy periods*
  - pick a sample from the inter-departure time sequence and it is likely to be in a busy period
- Therefore, splitting the trace at a given sample is likely to split the busy periods
  - external shuffle breaks up busy periods
  - internal shuffle reorders packets in the b.p.

Shuffling by Time Interval

- Divide trace into intervals of length *L*
- Interval 1 contains n_1 samples, where

$$\sum_{i=1}^{n_1} t_i \le L < \sum_{i=1}^{n_1+1} t_i$$

- Intervals also contain *partial samples* at the beginning and the end

Bias of Time Interval Shuffling

- Block boundaries are more likely to occur between bursts than between packets inside a burst
- *If time between bursts were exponential*, then the two partial samples at the block boundaries would joint to become exponential
- External shuffling would test dependence *between* busy periods

Effects of Time Interval Shuffling

- 30 minute traces were externally shuffled, varying *L* from 0.1 seconds to 24 minutes
- For Erramilli's curves, shuffling block sizes as small as 1 second have no visible effect
- When y-axis scale is increased larger block sizes must be used, but 1 minute is adequate

Trace Decomposition

- Label the intervals into different groups, based on number of samples per interval
- Divide the trace into *sub-traces*, containing only those intervals from a single group
- Run separate experiments on each sub-trace
- Combine the results of each experiment as a weighted average, using proportion of samples contained in each sub-trace

Effect of "Trains" on Delay

- In a "train" model, t_i, t_{i+1}, . . . , $t_{i+k\text{-}1}$ are *small*, and possibly *periodic*; t_{i+k} is *large*
- t_i depends on typical packet size within a transaction
- If Erramilli's service time is less than t_i, then average waiting time is almost zero

Service Time Exceeds Inter-Departure time in Trace File

- Assume inter-departures within "train" are t
- Assume all service times are v
- Average waiting time is

$$\frac{0+(t-v)+2(t-v)+\cdots+(k-1)(t-v)}{k}=\frac{(k-1)(t-v)}{2}$$

- Result is *linear growth* beyond the "knee"

Effect of Recombining Sub-traces

- Traffic on a given subtrace has
  - no long-range dependence, and
  - is less bursty than the original
- Recombining the output from separate experiments on each sub-trace is
  - almost perfect in low load (startup transient)
  - pessimistic in high load (busy periods get huge)

Conclusions

- Erramilli's experiment is flawed, and does not demonstrate long-range dependence
- Shuffling by time interval is more relevant
- Beyond a critical interval size on the order of 1 second, shuffling has very little effect
- Behavior could be explained by exponential gaps between highly variable busy periods