Analysis of UNC Link Data

1 million packets, from UNC main connection

Time needed: ~ 3 seconds

Study both “incoming” and “outgoing”

View 1 of UNC Link Data

Packet Arrival Times, i.e. “rates”

SiZer version of smooth histogram

Show UncLinkData2p1d1.mpg and UncLinkData2p1d2.mpg

Notes:

- Overall high rate
- SiZer and SiCon show many “statistically significant bursts”
- SiZer patterns similar (for in vs. out), suggesting strong correlation.
- More packets outgoing (see density height and time shift)

View 2 of UNC Link Data

Packet Sizes, in bytes

SiZer version of smooth histogram

Show UncLinkData2p2d1.ps

Nicer on log scale???

Show UncLinkData2p3d1.mpg and UncLinkData2p3d2.mpg

Notes:

- widely separated values
- some common sizes (e.g. 40 bytes)
- min = 28,   max = 1500
- more data outgoing (bigger packets)

View 2 (cont.)

Summary:

Percentages for special sizes:

Size	Incoming	Outgoing
28	0.01%	0.02%
40	37%	25%
1500	12%	33%

View 3 of UNC Link Data

Packet Sizes per unit time

SiZer: averages over different time scales

Show UncLinkData2p5d1.mpg and UncLinkData2p5d2.mpg

Notes:

- Statistically significant changes at many scales
- Bursty behavior
- Outgoing larger than incoming
- No apparent correlation

Views 4 and 5 of UNC Link Data

Views of binned data,

10,000 bins $\quad \Rightarrow$ scale $\quad m \approx 0.02$ sec
& ~ 100 obs. per bin

Study both:

- Counts (i.e. # packets in bin)
- Packet Size Totals (sum'd over bin)

View 4 of UNC Link Data

Autocorrelations

Show UncLinkData2p11d1.ps, UncLinkData2p11d2.ps, UncLinkData2p12d1.ps and UncLinkData2p12d2.ps

Big Surprise:

Have "mixture" of (~10%) long range dep. and (~90%) "white noise"

Possible Reasons???

1. Studying wrong scale?

R&W: $m >$ "packet round trip time" $\Rightarrow$

$\Rightarrow$ Fractional Gaussian Noise

2. Mixture of session/data types?

3. ???

View 4 of UNC Link Data (cont.)

Autocorrelations (cont.)

Notes:

- “exponential intuition” from
  $y \approx a \cdot e^{b \cdot x} \Leftrightarrow \log y \approx b \cdot x + \log a$

- exp. fits with $f \approx 0.98$ suggest “unit root”, i. e. long range dependence

- “polynomial intuition” from
  $y \approx a \cdot x^{b} \Leftrightarrow \log y \approx b \cdot \log x + \log a$

- poly fits, with powers $\in (-.28, -.08)$, suggest Hurst parameters:
  $H \in (0.86, 0.96)$

View 5 of UNC Link Data

Marginal Distributions

Show UncLinkData2p13d1.ps, UncLinkData2p13d2.ps, UncLinkData2p14d1.ps and UncLinkData2p14d2.ps

Notes:

- All are "roughly" both normal and log-normal
- Bin counts more normal?
- Packet Sizes more lognormal?
- Sizes have "few very small values"?

View 6 of UNC Link Data

Autocorrelations across scales

Best case: Packet Sizes, Incoming

Show UncLinkData2p22d1.ps

Coarser Scales $\Rightarrow$

$\Rightarrow$ overall more dependence

$\Rightarrow$ steeper at left

$\Rightarrow$ more variability

View 6 of UNC Link Data (cont.)

Summaries of parameters:

Show UncLinkData2p22d1.ps

R^2 for Long Range Dependence:

- “low” for “small” $m \in (10^{-3}, 10^{-2})$
- “increases” for $m \in (10^{-2}, 10^{-1})$
- “large” for $m \in (10^{-1}, 1)$

Power, f, Hurst Parameter:

- Correlated
- “increasing for small scales m

View 6 of UNC Link Data (cont.)

Other cases (outgoing, Bin Counts)

- somewhat similar
- more “noise” problems???

Worst case: Outgoing Packet Sizes

Show UncLinkData2p22d2.mpg and UncLinkData2p22d2.ps

- Fit lines sometimes slope down
- Then $f > 1$?!?
- and Hurst Param. > 1 ?!?

Should Pursue Further?

1. Study autocorr's at coarses scales m? (needs more data)

2. Modify simulations, to show observed autocorr. Structure?

3. Careful look at sessions (to explain autocorr.)?

4. Other explanations of autocorr.?

5. Investigate "few small sizes"?