Network traffic: Scaling

Ways of representing a time series

Timeseries: information in time domain
Ways of representing a time series

Timeseries: information in time domain
FFT: information in frequency (scale) domain

Timeseries: information in time domain
FFT: information in frequency (scale) domain
Wavelets: information in time and scale domains
Wavelet Coefficients: Local averages and differences

Intuition:
- Finest scale:
  - Compute averages of adjacent data points
  - Compute differences between average and actual data
- Next scale:
  - Repeat based on averages from previous step

Use wavelet coefficients to study scale or frequency dependent properties

Wavelet example
Wavelets

**FFT:** decomposition in frequency domain

**Wavelets:** localize a signal in both **time** and **scale**
Discrete wavelet transform

Definition:
- From 1D to 2D: \( X \leftrightarrow \{d_{j,k} : j \in \mathbb{Z}, k \in \mathbb{Z}\} \)
- Wavelet coefficients at scale \( j \) and time \( 2^k \)
  \[ d_{j,k} = \int X(s) \psi_{j,k}(s) ds, \quad j \in \mathbb{Z}, k \in \mathbb{Z} \]
- Wavelets: \( \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k) \)
- Wavelet decomposition: \( X(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t) \)

Global scaling analysis

Methodology: Exploit properties of wavelet coefficients
- Self-similarity: coefficients scale independent of \( k \)
  \[ d_{j,k} \approx 2^{j(1+2H)} \text{ for all } j \]

Algorithm:
- Compute Discrete Wavelet Transform
- Compute energy of wavelet coefficients at each scale
  \[ \log_2 E_j = \log_2 \left( \frac{1}{N_j} \sum_k |d_{j,k}|^2 \right) \approx -j(1+2H) \]
- Plot \( \log_2 E \) versus scale \( j \)
- Identify scaling regions, break points, etc.
- Hurst parameter estimation

Ref: AV IEEE Transactions on Information Theory 1998
Motivation

Scaling

- How does traffic behave at different aggregation levels

Large time scales: User dynamics => self-similarity

- Users act mostly independent of each other
- Users are unpredictable: Variability in
  - Variability in doc size, # of docs, time between docs

Small time scales: Network dynamics

- Network protocols effects: TCP flow control
- Queue at network elements: delay
- Influences user experience

How do they interact???
Global scaling analysis (large scales)

\[ \text{Energy}_j = \frac{1}{N_j} \sum_k |d_{j,k}|^2 \]

- Trivial global scaling == horizontal slope (large scales)
- Non-trivial global scaling == slope > 0.5 (large scales)

Self-similar traffic
Self-similar traffic

Adding periodicity

- Packets arrive periodically, 1 pkt/2³ msec
- Coefficients cancel out at scale 4
Effect of Periodicity

Actual traffic: Different time periods
Actual traffic: different subnets

A simple topology

Used to measure before bottleneck

Used to vary delay

Used to limit capacity

Clients

- Used to vary delay
  - vary delay
  - access speed
Impact of RTT on global scaling

- **Workload**
  - Web (Pareto dist.)

- **Network**
  - Single RTT delay
  - Examples
    - scale 15 (24 ms)
    - scale 10 (1.3 s)

- **Conclusion**
  - Dip at smallest time scale bigger than RTT
A more complex topology

Impact of different RTTs on global scaling

- Network variability (delay) => wider dip
- Self-similar scaling breaks down for small scales
A more complex topology

Impact of different bottlenecks on global scaling

- Network variability (delay) => wider dip
- Network variability (congestion) => wider dip
- Simulation matches traces without explicit modeling
Impact of different bottlenecks on global scaling

- Network variability (delay) => wider dip
- Network variability (congestion) => wider dip
- Simulation matches traces without explicit modeling
Small-time scaling - multifractal

Wavelet domain:
- **Self-Similarity**: coefficients scale independent of $k$
- **Multifractal**: scaling of coefficients depends on $k$

  local scaling is time dependent

Time domain:

Traffic rate process at time $t_0$ is:
- # of packets in $[t_0, t_0 + \delta t]$

  **Self-Similarity**: traffic rate is like $(\delta t)^H$
  **Multifractal**: traffic rate is like $(\delta t)^{\alpha(t_0)}$

Conclusion

Scaling

- **Large time scales**: self-similar scaling
  - User related variability
- **Small time scales**: multifractal scaling
  - Network variability
    - Topology
    - TCP-like flow control
    - TCP protocol behavior (e.g., Ack compression)
Summary

- Identified how IP traffic dynamics are influenced by
  - User variability, network variability, protocol variant
- Scaling phenomena
  - Self-similar scaling, breakpoints, multifractal scaling
- Physical understanding guides simulation setup
  - Moving towards right “ball park”
- Beware of homogeneous setups
  - Infinite source traffic models