Modeling and Analysis of Network Applications and Services

Adrian Popescu
Department of Telecommunications and Mathematics
University of Karlskrona/Ronneby

Outline

- Background
- Research Issues
- Estimation of Long-Range Dependence
- Multi-Resolution Analysis and A Wavelet-Based Estimation
- Stationarity
- Implementation
- Measurements
- HTTP Interactions with TCP
- Performance Model
- Web Document Structure
- Network Level Simulation
- Modeling Aspects of Client-Server Interactions

- Simulation Setup
- Effects of TCP/IP on Application Parameters
- End-to-End Performance
- Conclusions

Background

- Research project
  Traffic Control in ATM Networks: Engineering Impacts for Realistic Traffic Processes
- Support
  - Nutek (SWAP II)
  - Blekinge Forskningsstiftelse
- Focus
  - Self-similarity in teletraffic, origins and impacts
  - Controls for QoS with bursty traffic

Research Issues

- Network application traffic characterization and analysis (HTTP; VBR video; FTP; Webcast)
- Evaluation of schemes for modeling and generation of HTTP and VBR bursty traffic
- Multi-fractal analysis and modeling of network applications and services
- Understanding the origins and impacts of self-similarity
- Effect of controls on self-similar traffic
- Interaction between protocols
- Evaluation of different classes of VBR traffic control
- Design of new VBR control mechanisms

Preliminiary results

- "Traffic Characterization Using Wavelet-Based Techniques", D. Boulliant, P. Pruthi and A. Popescu, SPIE International Conference on Performance and Control of Network Systems, Boston, USA, November 1998
- "HTTP Interactions with TCP", P. Pruthi, A.K. Jena and A. Popescu, 11th ITC Specialist Seminar, Yokohama, Japan, October 1998
-
-
-

Estimation of Long-Range Dependence

- Estimating H
  - Aggregation-based estimators (R/S analysis; variance-time analysis)
    Not precise
    Rough estimate of H
    May get seriously biased
  - Asymptotically unbiased estimators derived from Gaussian Maximum Likehood Estimation MLE (Whittle-type estimators)
    Good statistical performance
    Drawbacks: parametric estimators ; computational intensive; risk that the estimator may become biased
  - Wavelet-based estimation

Multi-Resolution Analysis and a Wavelet-Based Estimation

- Multi-Resolution Analysis (MRA)

$$x(t) = \sum_k a_x(J,k)\Phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_x(j,k)\Psi_{j,k}(t)$$

where

$$\left\{\Phi_{J,k}(t) = 2^{-\frac{j}{2}}\Phi_0(2^{-j}t - k)\right\} \quad \text{and} \quad \left\{\Psi_{j,k}(t) = 2^{-\frac{j}{2}}\Psi_0(2^{-j}t - k)\right\}$$

- Discrete wavelet transform: collection of coefficients

$$\{d_x(j,k), j = 1 \ldots J, k \in Z\} \quad \text{and} \quad (a_x(J,k), k \in Z)$$

as defined by the inner products

$$d_x(j,k) = (x,\Psi_{j,k}) \quad \text{and} \quad a_x(j,k) = (x,\Phi_{j,k})$$

Multi-Resolution Analysis and a Wavelet-Based Estimation (cont.)

- Applying the wavelet transform to a wide-sense stationary stochastic process of type LRD
  - Power-law behavior of LRD processes with H

$$\gamma_x(\tau) \sim c_\gamma \tau^{-(2-2H)} \quad , \quad \tau \to \infty$$

  - Power spectrum

$$\Gamma(\upsilon) \sim c_f |\upsilon|^{-(2H-1)} \quad , \quad \upsilon \to 0$$

  - Statistics of the wavelet coefficients

$$E[d_x(j,k)] = 0$$

$$E\left[|d_x(j,k)|^2\right] = \int \Gamma(2^j\upsilon)|\Psi_0(\upsilon)|^2 d\upsilon = c_f 2^{-j(1-2H)} \int |\upsilon|^{(1-2H)} |\Psi_0(\upsilon)|^2 d\upsilon$$

Practically

$$E\left[|d_x(j,k)|^2\right] = \frac{2^j}{n_0}\sum_k |d_x(j,k)|^2$$

Multi-Resolution Analysis and a Wavelet-Based Estimation (cont.)

- Unbiased estimation of the parameter H: linear fitting in a log-log plot

$$\log_2\left(\frac{2^j}{n_0}\sum_k |d_x(j,k)|^2\right) = (2\hat{H}-1)j + \hat{c}$$

where

$$\hat{c} \quad \text{estimates} \quad \log_2\left(c_f \int |\upsilon|^{(1-2H)} |\Psi_0(\upsilon)|^2 d\upsilon\right)$$

Number of vanishing moments

$$N > H - 1$$

Stationarity

- Difficulties in distinguishing a stationary process with LRD from a non-stationary one
- Need to test for stationarity assumptions in the process
- Method:
  Partition the sequence
  Perform H estimations on each subsequence
  Compare with the H estimation on the entire sequence
  Should the process be stationary, the mean of the H estimations should not differ in any larger extend from the H estimate taken over the entire data set
- Condition: independence of the partial estimates

Implementation

- Discrete wavelet transform
  Fast pyramidal filter-bank based algorithm followed by a time-averaging unit
- Problem: boundary effects
- Solution: discard the corrupted wavelet coeffiecients from the data set
- Risk: loss of accuracy
- Good practical compromise:
  N = H+1
- Advantages in terms of memory usage
- Validation:
  Traces of fBm and fGn
  Excellent agreements with theoretical results

Measurements

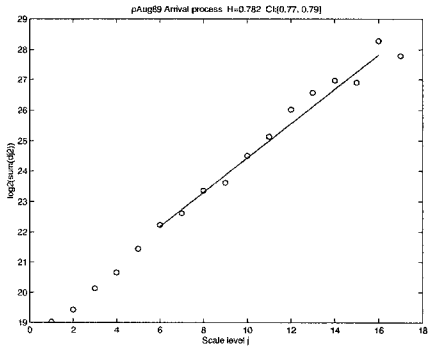
- **Ethernet data**
  Bellcore Ethernet traces (Aug '89)

  Focus: discrete versions of the arrival process and the corresponding point-process
  Sampling interval: 6 msec

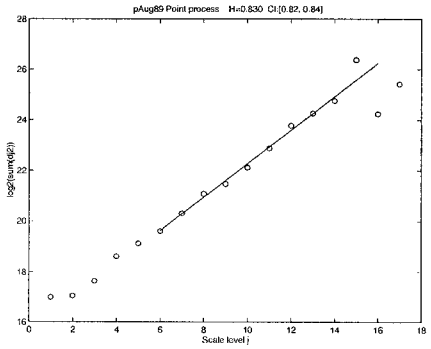
  Focus: stationarity condition

  Wavelet: Daubechies wavelet with N = 2

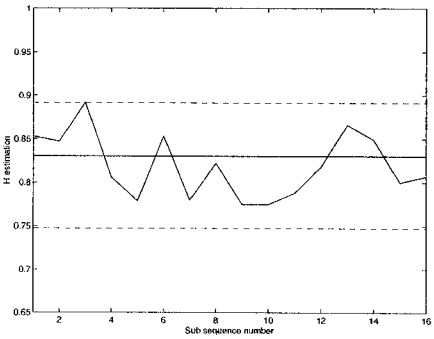
Measurements (cont.)



Measurements (cont.)



Measurements (cont.)



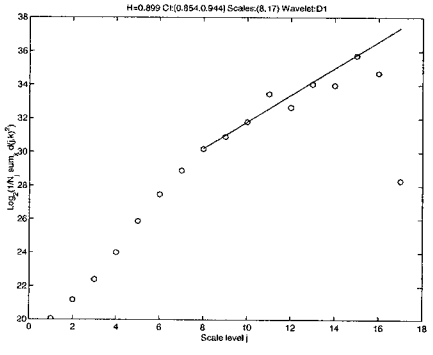
Measurements (cont.)

- **VBR video data**
  Movie "Star Wars" traces
  Encoded using MPEG-1
  Movie length: ~ 2 hours

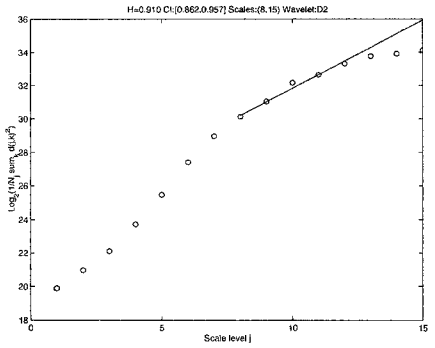
  Wavelets: Haar wavelet, Daubechies 2 wavelet and Daubechies 3 wavelet

  Question: non-stationarity vs estimator sensitivity (when strong short-lag correlations)

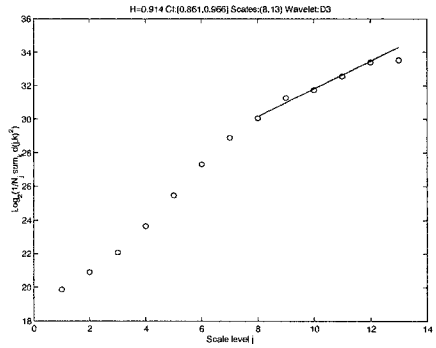
Measurements (cont.)



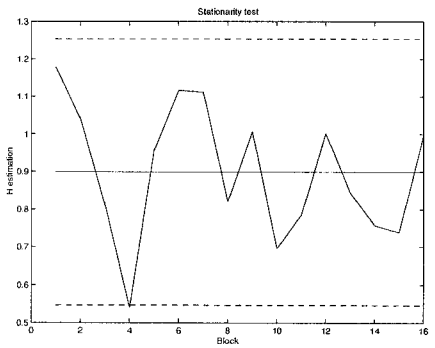
Measurements (cont.)



Measurements (cont.)



Measurements (cont.)



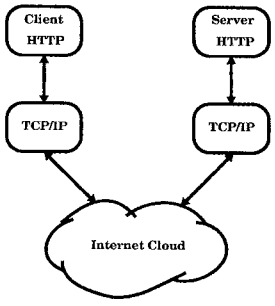
HTTP Interactions with TCP

- **Question: understanding the self-similar phenomenon in network packet traffic**
- **Underlying physics that could give rise to self-similarity:**
  - **User (human) behavior (sociology, Lotka's law; medicine, cardiac spectrum; biology, Willis' law; economics, Pareto's law; etc)**
  - **Data generation, organization and retrieval (application characteristics: heavy-tailed as well as light-tailed packet arrival and file size distributions; ON-OFF sources with infinite variance syndrome, Noah effect)**
  - **Computational models (client/server)**
  - **Traffic aggregation (Joseph effect)**
  - **Network control with bounded resources**
  - **Feedback-based (flow) control mechanisms**
  - **Protocol interactions (cascading effects)**
  - **Network evolution**

HTTP Interactions with TCP (cont.)

- **Management of application traffic has requirements that are different from those for the management of transport-level traffic**
- **Application-level traffic management: decisions on the admission control and resource management hierarchy**
- **Transport- and network-level management: requires the delivery of implied (e.g., IP) or contracted (e.g., ATM) QoS**
- **Conclusion: need to understand the traffic parameters at the application layer AND to model and quantify the effects of diverse control mechanisms**
- **Problem: lack of results showing quantitative contributions of protocol stacks and their effects on source parameters**
- **Main question: how specific LRD parameters (application-level) get modified by TCP/IP?**
- **Application of interest: WWW**

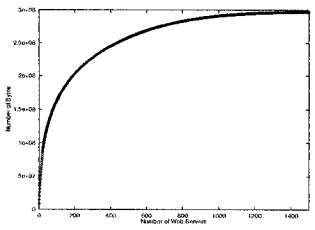
Performance Model



Web Document Structure

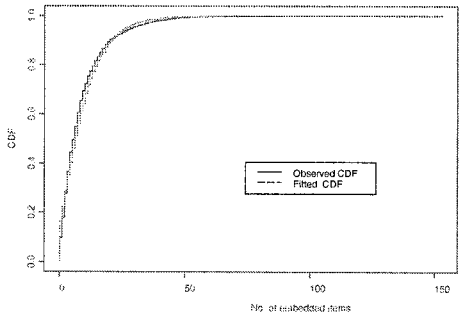
- Analysis of data (web traffic) collected during a busy hour from a production Frame Relay network (part of Pacific Bell's Internet access network)
- About 1800 web servers were accessed by hundreds clients during this interval via one Frame Relay trunk
- A number of about 25000 web pages were analyzed
- Contributions:
  - Distribution of the total number of bytes accessed from various servers on the Internet during a busy hour
  - Distribution of the number of embedded items within web documents

Web Document Structure (cont.)



Total number of bytes received during a busy hour from the first X most popular servers

Web Document Structure (cont.)



Number of internally embedded items: actual vs as fitted by a negative binomial distribution

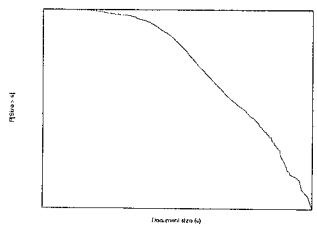
Network Level Simulation

- Simulation environment: OPNET simulation tool
- Clients and servers grouped into two 10 Mbps Ethernet networks
- Interconnection of Ethernets: 192 kbps link
- Model with multiple client nodes (up to 60 users) and a single server
- Experiments focused onto the bottleneck link

Modeling Aspects of Client-Server Interactions

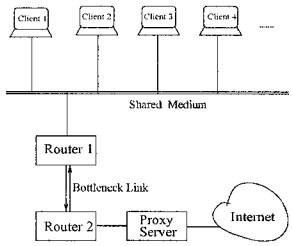
- Client model: ON/OFF source
- Server model: FIFO
- Probability of success in accessing a web server (web page request): 90%
- Probability of objects of type HTML (server traffic): 50%
- Browser caching not modeled
- However, probability of no change for a specific document (since last transferred): 4%

Modeling Aspects of Client-Server Interactions (cont.)



Typical CCDF (log-log plot) of web document sizes

Simulation Setup

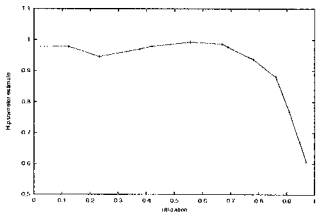


Simulation setup for analyzing the effect of TCP on HTTP performance

Effects of TCP/IP on Application Parameters

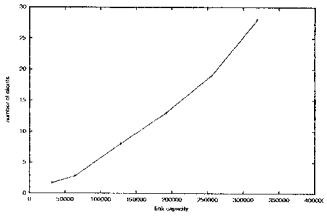
- Aggregate traffic on the Bottleneck Link: aggregation of several TCP sessions
- Duration of TCP sessions: governed by the distribution of objects as well as the effects of controls and congestion
- Application traffic: Pareto model
- Network traffic: fBm model
- Long estimations (H parameter) under conditions of high load
- Estimates (simulation):
  - H as a function of link utilization
  - Number of users as a function of link capacity
  - fBm peakedness as a function of link utilization

Effects of TCP/IP on Application Parameters (cont.)



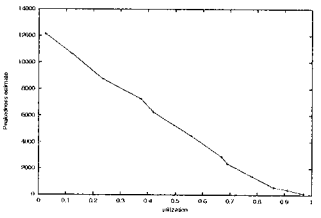
Hurst parameter as a function of utilization on the bottleneck link from server to client

Effects of TCP/IP on Application Parameters (cont.)



Number of web users supported by a channel with capacity C

Effects of TCP/IP on Application Parameters (cont.)

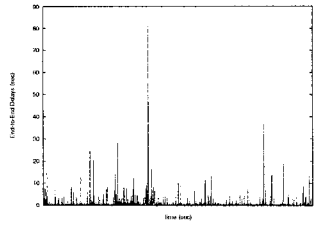


fBm peakedness as a function of link utilization

End-to-End Performance

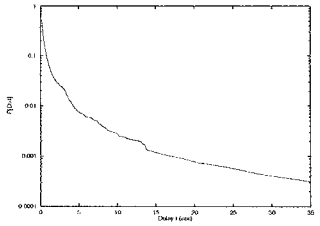
- Main goal: end-to-end application level performance (delay)
- Conditions for simulation experiments:
  - Bottleneck link: 192 kbps
  - Router buffers: 57.6 kb
- Extreme difficulties when computing analytically end-to-end delay (mixing of negative binomial distribution with convolutions of Weibull distributions)

End-to-End Performance (cont.)



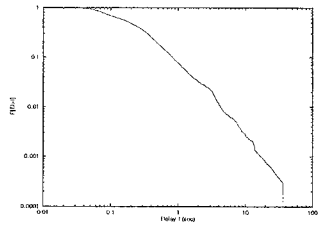
Observed end-to-end web object retrieval delays (sample trace)

End-to-End Performance (cont.)



CCDF of the observed end-to-end application delays (semi-log plot)

End-to-End Performance (cont.)



CCDF of the observed end-to-end application delays (log-log plot)

Conclusions

- Contributions
  - Wavelet-based tool for the analysis of LRD traffic
  - Distribution of the total number of bytes accessed from various servers during a busy hour
  - Distribution of the number of embedded items within web documents
  - Simulation of a client-server environment for WWW
  - Interactions between HTTP and TCP/IP protocols in terms of self-similarity in network traffic
- Further topics of interest
  - Wavelet-based analysis of VBR traffic
  - Multi-fractal properties of traffic
  - Engineering of trunks and network resources for end-to-end QoS
  - Interactions among protocols (cascading effects)