Modeling and analyses of traces

- what do real workloads look like?
- what is performance of real systems
  \[\Rightarrow\] measurements and analysis

- trace collection
- trace analysis
  - what are statistical properties?
- development of models
  - descriptive models
  - explanatory models
Self similarity of network traffic
Time series Analysis

- time series \( \{X_t : t = 1, 2, \ldots \} \) covariance stationary process
  - mean \( \mu \), variance \( \sigma^2 \)

- covariance stationary process if
  - \( \text{Cov}(X_t, X_{t+k}) = E[X_t X_{t+k}] - \mu^2 \) doesn’t depend on \( t \) for all \( k \)

- autocorrelation function \( r(k) \)
  - \( r(k) = \frac{\text{Cov}(X_t, X_{t+k})}{\sigma^2} \)
  - \( r(0) \geq |r(k)|, k = 1, \ldots \)
  - \( r(k) = r(-k), k = 1, \ldots \)
Examples

- Poisson process \( \{N_t\} \)
  - \( X_t = N_{t+1} - N_t \)
  - \( r(0) = 1; \ r(k) = 0, \ k = 1,... \)

- Renewal process with hyperexponential, Erlang distributions
  - \( r(k) \sim \exp(-\alpha k), \text{ as } k \to \infty \)

- Birth death process
  - \( r(k) \sim \exp(-\alpha k), \text{ as } k \to \infty \)
Long range dependence

- $X = \{X_t: t = 1,2,...\}$ is covariance stationary process
  - mean $\mu$, variance $\sigma^2$

- $X$ is long range dependent if

$$
\sum_{k=0}^{\infty} r(k) = \infty
$$

satisfied if $r(k) \sim k^{-\beta}$ as $k \to \infty, 0 < \beta < 1$
Self similarity

- \( X^{(m)} = \{X_k^{(m)}\} \) where elements are average over non-overlapping blocks of \( m \)
- Suppose that \( r(k) \sim k^{-\beta}, 0 < \beta < 1 \)
- \( X^{(m)} \) is second-order self-similar with Hurst parameter \( H = 1 - \beta / 2 \) if for all \( m, \) (.5 < H < 1)
  - \( \text{Var}(X^{(m)}) = \sigma^2 m^{-\beta} = \sigma^2 m^{-2(1-H)} \)
  - \( r^{(m)}(k) = r(k), k = 0, ... \)
- \( X^{(m)} \) is asymptotically second-order self-similar if
  - \( r^{(m)}(k) \sim r(k) \) as \( k \to \infty \)
Ethernet and WAN traffic appear self-similar (Leland etal 94)

visual inspection

- $x$ - time in varying units
- $y$ - packets/unit time
Power spectral density

- if given $X$, power spectral density $f(w)$ is discrete time Fourier transform of $r(k)$

\[ f(w) = \sum_{k=-\infty}^{\infty} r(k)e^{-iwk} \]

because $r(k)$ is real valued,

\[ f(w) = r(0) + 2\sum_{k=1}^{\infty} r(k)\cos(wk) \]

- if $X$ is LRD, then $f(w) \sim w^{-(1-\beta)}$ as $w \to 0$
  - $f \to \infty$, $w \to 0$
Statistical tests for LRD

- $n$ samples $X_1, \ldots, X_n$
- Variance-time method
  - Second-order self-similar process $X^{(m)}$
  - $\text{Var}(X^{(m)}) = \sigma^2 m^{-\beta}$
  - Plot $\log(\text{Var}(X^{(m)}))$ against $\log m$
- Periodogram
  - Estimate PSD
    \[ I(w_j) = \hat{r}(0) + 2 \sum_{k=1}^{n-1} \hat{r}(k) \cos(w_j k), \quad w_j = 2\pi j / n, \quad j = 1, \ldots, n/2 \]
  - Check for $I(w) \sim w^{-(1-\beta)}$
- Wavelet - PSD in time domain
Models

- fractional Gaussian noise (FGN)
  - Gaussian process with mean $\mu$, variance $\sigma^2$ and autocorrelation function
  - $r(k) = \frac{1}{2}(|k+1|^{2H}-|k|^{2H}+|k-1|^{2H})$, $k=1,\ldots$

- discrete time $M/G/\infty$ input model
  - service time $X$; residual service time $Y$
    - $P(Y=k) = P(X>k)/E[X]$
  - $\{N_t\}$ has autocorrelation $r(k) = \lambda E[X] P(Y>k)/\sigma_N^2$
  - $\{N_t\}$ is LRD when $E[X] < \infty$, $\sigma_X^2 = \infty$
    - example, $X$ given by Pareto distribution, $P(X>k) \propto k^{-\alpha}$, $1<\alpha<2$
Where does LRD come from?

- Web object sizes come from heavy-tailed distribution
  - approximately Pareto (Bestavros, Crovella 97)

- p2p apps make it more pronounced
The use of aggregation: pitfalls to avoid

- simulation of single link
  - $N = 40$ flows;
  - buffer $B = 20$
  - bandwidth $1$ Mbps
  - RTT $\sim 200$ ms

- single TCP session generates LRD traffic
  - Hurst param $\sim 0.8$
  - loss prob. $\sim 0.16$

(Veres, Boda, Infocom 2000)
Statistical tests

Absolute value method

\[ X_k^{(m)} = \sum_{1 \leq i \leq m} X_{ik} \]

\[
\mu(m) = \frac{1}{N/m} \sum_{k=1}^{N/m} X_k^{(m)} - \frac{1}{N} \sum_{i=1}^{N} X_i
\]

\[ \mu(m) \sim m^{-\beta/2} \]
Analysis of four hours trace, 100ms samples

Absolute values method

\[ \log_{10}(m) \]

Periodogram method

\[ \log_{10}(\text{frequency}) \]

\[ \log_{10}(\mu(m)) \]  

H=1/2

self-similarity; Hurst parameter \( \approx 0.8 \)
Statistical Pitfalls

- use all data
  - 15 levels (base 2)
  - 4.5 levels (base 10)

- perform aggregation
  - absolute moments method
  - periodogram method
  - compare trace to fractional gaussian noise (FGN)

- collect longer traces (if possible)
Absolute moments method
TCP trace

Aggr = 100

FGN

log10(m)

log10(μ(m))

log10(μ(m))

Aggr = 1000

log10(m)

log10(m)
Periodogram method

TCP trace  Original  FGN

\[ \text{Aggr} = 10 \]
Successively Higher Aggregates

TCP trace  \[ \text{Aggr} = 100 \]

FGN  \[ \text{Aggr} = 1000 \]
End-to-end loss traces
Outline

- trace analysis
  - stationarity
  - autocorrelation
  - burst length distributions

- modeling
  - k-th order Markov models
  - hidden Markov models
Trace Collection

- one mobile host and one fixed end host
- approximately 215 minutes, 610K frames
- frame w/o error - record '0';
- frame with error - record '1';
- details: Konrad et al, MSWiM 2001
- URL:
  http://www.cs.berkeley.edu/~almudena/traces/index.html
Whole trace, Window Size = 2,000

Smoothed loss rate

Sequence number
Whole trace, Window Size = 5,000

Smoothed loss rate

Sequence number

0 200000 400000 600000

0 0.05 0.1 0.15 0.2 0.25 0.3
Whole trace, Window Size = 10,000

Smoothed loss rate

Sequence number
GSM Frame Error Trace Analysis
- Stationary Test

- frame-error trace exhibits three distinguishable trace parts

- stationarity test:
  - divide each trace part into segments; segment size may be varied
  - hypothesis tests detects underlying trend for aggregations of the time series

- part 3 does not pass
- part 1 passes with segment sizes >2200
- part 2 passes with segment sizes > 800
First- and Second-order Statistics

- frame Error Rate (FER)
- error and error-free burst lengths
  - coefficient of variation (Cov)
  - complementary cumulative distribution function (ccdf)
- auto-correlation function
Estimators

- maximum likelihood estimate
- moment-based estimates
- Bayesian estimation
maximum likelihood estimation

Example: $X_1, \ldots, X_n$ - i.i.d. random variables with probability $p_X(x | \theta) = P(X=x)$ where $\theta$ is a parameter

- likelihood function $L(\theta | x)$ where $x=(x_1, \ldots, x_n)$ is set of observations

$$L(\theta | x) = \prod_{i=1}^{n} p_X(x_i | \theta)$$

- maximum likelihood estimate $\hat{\theta}(x)$ maximizer of $L(\theta | x)$
**MLE**

- typically easier to work with log-likelihood function, $C(\theta | x) = \log L(\theta | x)$

  $$C(\theta | x) = \sum_{i=1}^{n} \log p_x(x_i | \theta)$$

- example, Bernoulli process with parameter $p$, $p_x(1|p) = p$, $p_x(0|p) = 1 - p$

  $$C(p | x) = \sum_{i=1}^{n} (x_i \log p + (1 - x_i) \log(1 - p))$$
  $$= n_1 \log p + n_0 \log(1 - p)$$

where $n_1$ is number of 1s, $n_0$ - number of 0s, $n_1 + n_0 = n$
Bernoulli process example
continued

\[ \frac{dC}{dp} = \frac{n_1}{p} - \frac{n_0}{(1-p)} \]

- setting \( \frac{dC}{dp} = 0 \) and solving yields
  \[ p = \frac{n_1}{n} \]
Properties of estimators

- An estimator $\hat{\theta}(x)$ is unbiased if
  \[ E[\hat{\theta}(x)] = \theta \]

- $\hat{\theta}(x)$ is asymptotically unbiased if
  \[ E[\hat{\theta}(x)] = \theta \text{ as } n \to \infty \]
Properties of MLE

- asymptotically unbiased, i.e.,
  \[ \hat{\theta}(x) \to \theta \text{ as } n \to \infty \]

- asymptotically optimal, i.e., \( \hat{\theta}(x) \) has minimum variance as \( n \to \infty \)

- invariance principle, i.e., if \( \hat{\theta}(x) \) is the MLE for \( \theta \) then \( f(\hat{\theta}(x)) \) is MLE for any function \( f(\theta(x)) \)
Returning to trace

<table>
<thead>
<tr>
<th></th>
<th>FER</th>
<th>Cov of Burst Lengths</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>error burst</td>
<td>error-free burst</td>
</tr>
<tr>
<td>Part1</td>
<td>0.0503</td>
<td>2.064</td>
<td>5.116</td>
</tr>
<tr>
<td>Part2</td>
<td>0.0134</td>
<td>1.087</td>
<td>3.452</td>
</tr>
<tr>
<td>Part3</td>
<td>0.0799</td>
<td>2.850</td>
<td>5.875</td>
</tr>
</tbody>
</table>
Burst length distributions

CCDF – $P(X \geq k)$, $k=1,…$
An Aside: how not to look for power laws

Faloutsos$^3$ (Sigcomm'99)
- frequency vs. degree

topology from BGP tables of 18 routers
Faloutsos$^3$ (Sigcomm'99)
- frequency vs. degree

Power Laws

topology from BGP tables of 18 routers
Power Laws

Faloutsos\textsuperscript{3} (Sigcomm'99)

- frequency vs. degree
- empirical ccdf
  \[ P(d>x) \sim x^{-\alpha} \]
Power Laws

- Faloutsos\(^3\) (Sigcomm'99)
  - Frequency vs. degree
  - Empirical ccdf
    \[ P(d>x) \sim x^{-\alpha} \]
  - \( \alpha \approx 1.15 \)
Autocorrelation function
Markov Models

- discrete-time Markov model \( \{ Y_t \} \), state space \( S \) with output \( \{ X_t \}, X_t \in \{0,1\} \)
  - \( \{ Y_t \} \) has transition probability matrix \( P \)
  - \( \{ X_t \} \) is described by \( p_{x,y} = P(X_t = y | Y_t = x) \)

- k-th order Markov chain
  - \( S = \{0,1\}^k \)
  - if \( x = b_1...b_k \), then
    \[
    p_{x,y} = \begin{cases} 
    a_{xy} & y = 1b_1b_2\cdots b_{k-1} \\
    1 - a_{xy} & y = 0b_1b_2\cdots b_{k-1} \\
    0 & \text{otherwise}
    \end{cases}
    \]
**k-th order Markov model**

Q: how to determine $P$ from set of observation $x$?

likelihood function

$$L(P \mid x) = \prod_{i=k}^{n-1} p_{x_{i-k+1} \ldots x_i, x_{i-k+2} \ldots x_{i+1}}$$

log likelihood function

$$C(P \mid x) = \sum_{x,z \in S} n_{x,z} \log p_{x,z}$$

where $n_{x,z}$ - number of transitions from $x$ to $z$

$$\hat{p}_{x,z} = \frac{n_{x,z}}{n - k + 1}$$
Hidden Markov Model (HMM)

- **general case**
  - complete likelihood function given $x$ (observable) and state sequence $y$ (not observable)
    
    $$L(P, q \mid x, y) = \prod_{i} p_{y_{i}, y_{i+1}} q_{x_{i}, y_{i}}$$

  - complete log likelihood function
    
    $$C(P, q \mid x, y) = \sum_{i} \left( \log p_{y_{i}, y_{i+1}} + \log q_{x_{i}, y_{i}} \right)$$

  - **MLE**
    
    $$\hat{p}_{y, z} = \frac{n_{y, z}}{(n-1)}; \quad q_{x, y} = \frac{m_{x, y}}{m}$$

    where $n_{y, z}$ - no. transitions from $y$ to $z$,
    
    $m_{x, y}$ - no. times output = $x$ while in state $y$
    
    $m_{y}$ - no. times state = $y$
**EM algorithm**

- **initialize** $P^{(0)}$, $q^{(0)}$
- **at** $i$-th iteration
  
  **Expectation step**
  
  estimate $n, m$ by conditional expectation given observation data $x$ under $P^{(i-1)}$ and $q^{(i-1)}$

  **Maximization step**

  - compute new estimates $P^{(i)}$ and $q^{(i)}$

- **iterate** E&M steps till convergence
EM algorithm

- EM is algorithm to solve a fixed point problem
- Properties
  - likelihood increases at each iteration
  - if single solution, result is MLE
- Baum-Welch algorithm is instance for HMMs
Model Comparison

- no model captures GSM frame error rate well
- no model captures tail behavior of error-free burst length, or autocorrelation function of empirical trace
Extended On/Off Model

- Cross correlation between loss and loss-free burst lengths negligible
  ⇒ on/off model

- Bursty loss and loss-free burst lengths
  ⇒ mixtures of Geometric phases

Geometric mixture for ‘1’ bursts

Geometric mixture for ‘0’ bursts

Off state

On state
Extended On/Off Model

Model parameters:
- initial model - fit mixture of geometric phases to loss and loss-free burst length distributions
- EM algorithm - to fine tune model parameters
### Evaluation - FER

<table>
<thead>
<tr>
<th>Model</th>
<th>Frame Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM trace (part2)</td>
<td>0.0134</td>
</tr>
<tr>
<td>4(^{th})-order MTA</td>
<td>0.0209</td>
</tr>
<tr>
<td>4(^{th})-order Markov</td>
<td>0.0139</td>
</tr>
<tr>
<td>5-states HMM</td>
<td>0.014 ± 0.006</td>
</tr>
<tr>
<td>Extended On/Off</td>
<td>0.0134</td>
</tr>
</tbody>
</table>

- **extended on/off model better predicts FER of GSM channel**
extended on/off model captures tail behaviors of burst lengths significantly better
Evaluation - Autocorrelation Function

- extended on/off model captures auto-correlation significantly better