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 = 5,000

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 \( p_X(x \mid \theta) = P(X=x) \) where \( \theta \) is a parameter

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

\[
L(\theta \mid x) = \prod_{i=1}^{n} p_X(x_i \mid \theta)
\]

- maximum likelihood estimate \( \hat{\theta}(x) \) maximizer of \( L(\theta \mid x) \)
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

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

- $\hat{\theta}(x)$ is asymptotically unbiased if
  
  $$E[\hat{\theta}(x)] = \theta$$
  
  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 \( \tau(\hat{\theta}(x)) \) is the MLE for any function \( \tau(\theta) \)
Returning to trace

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

**CCDF** \(- P(X \geq k), k=1, \ldots\)

---

**Graph**

- **X-axis:** Burst length
- **Y-axis:** CCDF
- **Legend:**
  - Error-free Burst
  - Error Burst
An Aside: how not to look for power laws

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

Topology from BGP tables of 18 routers
Faloutsos³ (Sigcomm'99)

- frequency vs. degree

topology from BGP tables of 18 routers
Power Laws

• Faloutsos\(^3\) (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

Autocorrelation Function of Part2
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 \( q_{y|x} = P(X_t = y | Y_t = x) \)

- **k-th order Markov chain**
  - \( S = \{0,1\}^k \)
  - if \( x = b_1...b_k \), then \( q_{y|x} = 1 \) if \( y = b_1 \), and 0 otherwise.
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
    \[ \hat{p}_{y,z} = n_{y,z} / (n-1); \quad q_{x,y} = m_{x,y} / m \]

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
  $\Rightarrow$ on/off model

- Bursty loss and loss-free burst lengths
  $\Rightarrow$ mixtures of Geometric phases

Diagram:
- Geometric mixture for ‘1’ bursts
- Geometric mixture for ‘0’ bursts

Off state $\rightarrow$ 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></th>
<th>Frame Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM trace (part2)</td>
<td>0.0134</td>
</tr>
<tr>
<td>4\textsuperscript{th}-order MTA</td>
<td>0.0209</td>
</tr>
<tr>
<td>4\textsuperscript{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
Evaluation – Burst Length Distributions

- extended on/off model captures tail behaviors of burst lengths significantly better
Evaluation - Autocorrelation Function

- Extended on/off model captures auto-correlation significantly better.
Summary
topics covered

- Poisson processes, Markov processes
- queueing theory
- stochastic differential equations (engineering viewpoint)
- scheduling theory
- queueing networks
- approximations
- TCP
- time series analyses
... not covered (well)

- measurement techniques
  - measurement-based modeling
- simulation techniques
  - output analysis (confidence intervals, ...)
- HW stressing
  - modeling studies with validation
  - analyses of measurement traces
- tie between stochastic ordering and performance modeling