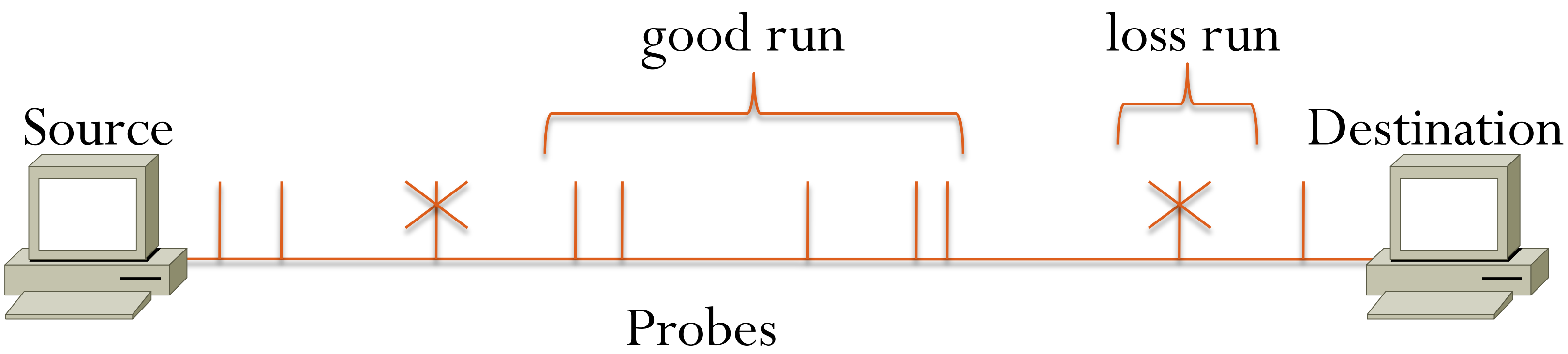
# SAIL: Statistically Accurate Internet Loss Measurements

Hung X. Nguyen and Matthew Roughan
The University of Adelaide, Australia

# Internet Loss Measurement

- Network operators continuously perform loss measurements
  - SLA contracts
  - We need to know that the problem exists before we can fix it
- Active probing: inject probe packets into the network




- Many IETF standards (RFC3357, RFC2330) and commercial products (Cisco IOS IP SLA, Agilent's Firehunter PRO)
  - Poisson Probes – PASTA (Poisson Arrivals See Time Average)
- N samples, typical loss metrics
  - loss rate = # of successes/N (RFC2330)
  - lengths of loss and good runs (RFC3357)

# Accuracy of Loss Measurements

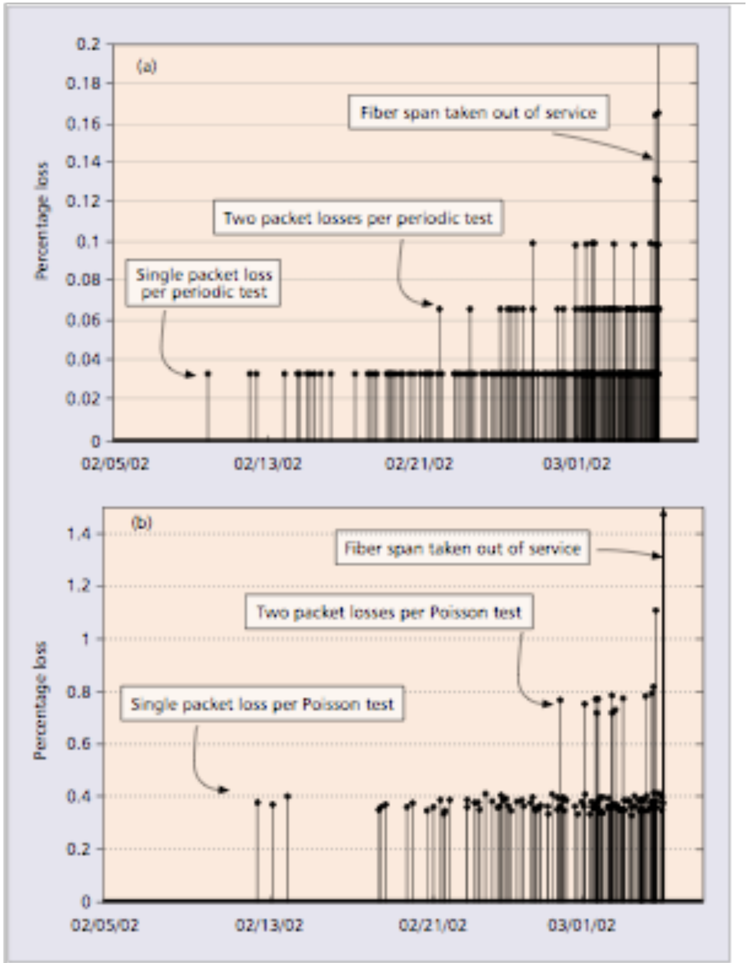


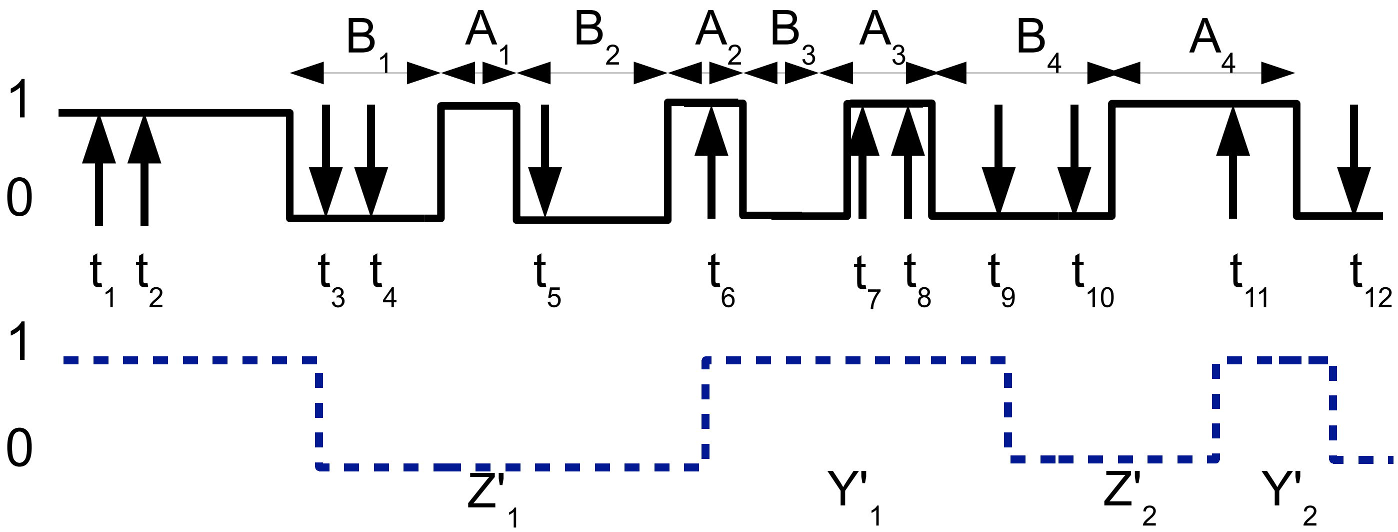

**Figure 2.** *Onset of low-level loss observed by a) periodic and b) Poisson probes.*

| | Loss rate | Loss run length mean (std)(second) |
|---|---|---|
| Web-like traffic | | |
| **True values** | **0.93%** | **0.136 (0.009)** |
| Poisson probes (10Hz) | 0.14% | 0(0) |
| Poisson probes (20Hz) | 0.12% | 0.022 (0.001) |
| TCP traffic | | |
| **True values** | **2.65%** | **0.136 (0.009)** |
| Poisson probes (10Hz) | 0.05% | 0 (0) |
| Poisson probes (20Hz) | 0.02% | 0(0) |

AT&T network, Ciavattone et al. 2003

Testbed at Wisconsin, Sommer et al. 2008

# Errors in loss estimates




- PASTA is an asymptotic result ($N \rightarrow \infty$)
- We need to compute the statistical errors of the estimations (e.g., variance)
  - Loss rate: $p = \frac{1}{N}\sum_{i=1}^{N} I_i$ ,
    - $I_i$ is the indicator function of probe ith
  - Variance: $VAR(p) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E[I_i I_j] - p^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} R(\mathrm{T}_{ij})$,
    - R( $\tau_{ij}$) is the auto-covariance function of probes ith and jth
- Probes miss ON/OFF intervals

# The auto-covariance function $R(\tau_{ij})$

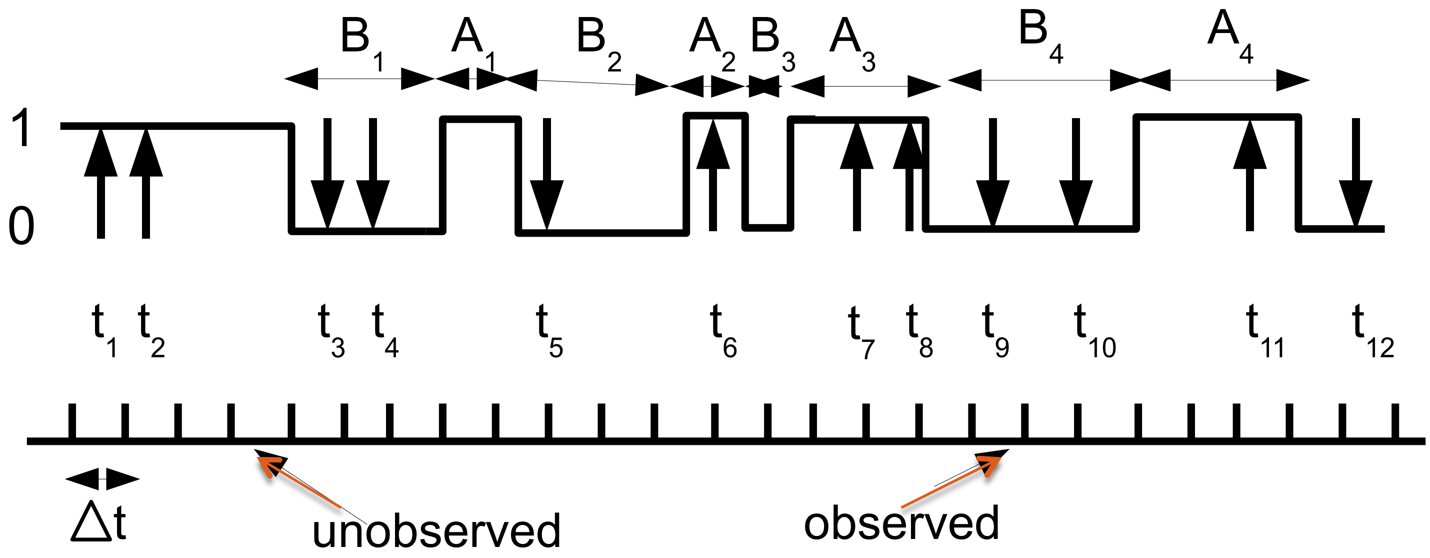
- Empirical computation
  - $R(\tau_{ij})$ can be computed directly from the samples
- Assume independent samples (commonly used)

$$VAR(p) = p(1-p)/N$$

- But losses are correlated, a model for the underlying loss process that captures sample correlation
  - Alternating Renewal ON/OFF model: $\{A_i\}, \{B_i\}$ are independent
  - $\{A_i\}, \{B_i\}$ are Gamma distributed with parameters $(k_0, \Theta_0)$and $(k_1, \Theta_1)$

# Inferring model parameters

- Missing intervals problem
  - Many short ON (or OFF) periods are not observed
  - loss run lengths and good run lengths observed by the probes are much larger than the real values
- Hidden Semi-Markov Model (HSMM) to the rescue



# Forward and Backward Algorithm

- Estimating $(k_0, \Theta_0)$ and $(k_1, \Theta_1)$ is a statistical inference with missing data problem
- Direct Maximum Likelihood Estimation is costly
  - $O(2^U)$, U is the number of un-observed intervals
- Forward and Backward algorithm to speed up
  - Exploiting the renewal properties
  - Expectation-Maximization algorithm
  - $O(2T^2)$, T is the number of intervals
- Knowing $(k_0, \Theta_0)$ and $(k_1, \Theta_1)$ , compute $R(\tau_{ij})$ using inverse Laplace transform
  - Numerical inversion
  - Simulation

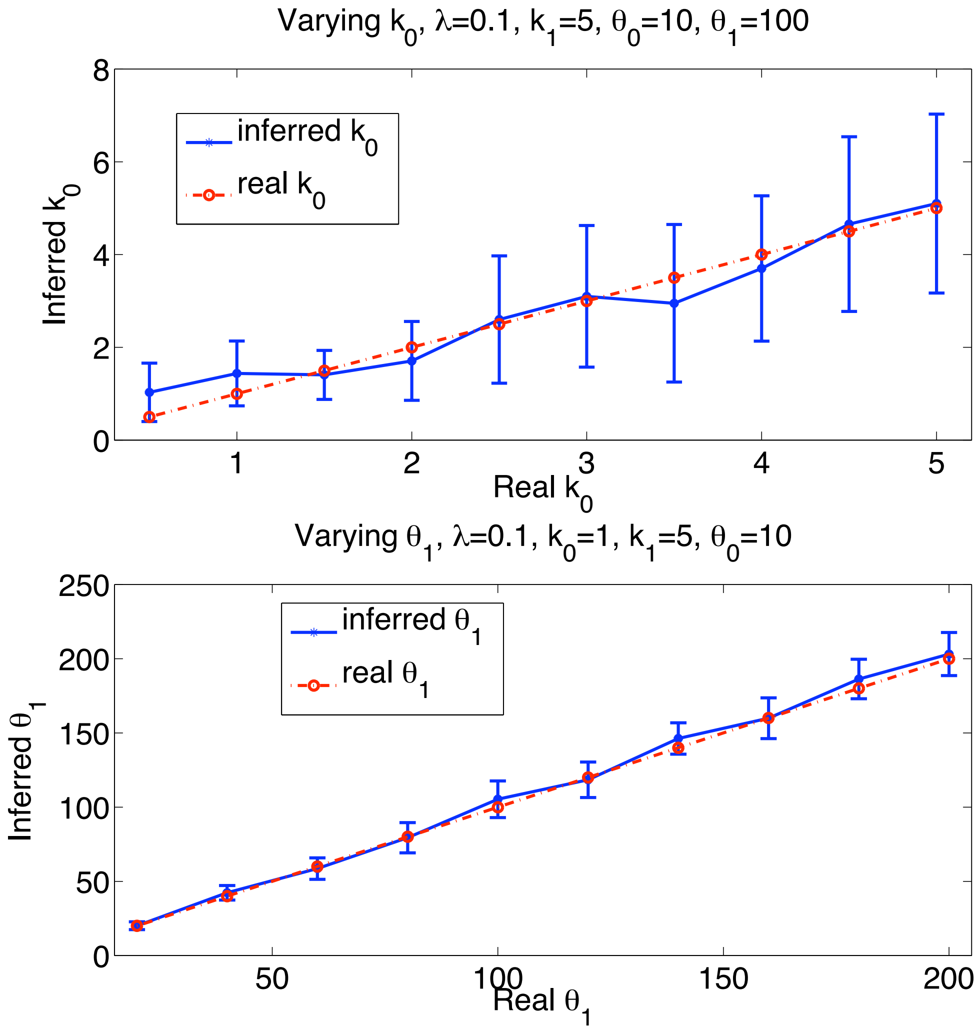
# SAIL

- Input
  - Probe sending times $\{t_1, \ldots, t_N\}$
  - Probe outcomes $\{I_1, \ldots, I_N\}$
  - The length of the discrete time interval $\Delta T$
- Algorithm
  - Apply the forward and backward algorithm to compute $(k_0, \Theta_0)$ and $(k_1, \Theta_1)$
  - Apply the inverse Laplace transform to find $R(\tau)$
  - Compute the loss rate and its variance
- Output
  - The loss rate and its confidence intervals
  - The parameters $(k_0, \Theta_0)$ and $(k_1, \Theta_1)$ of the loss process

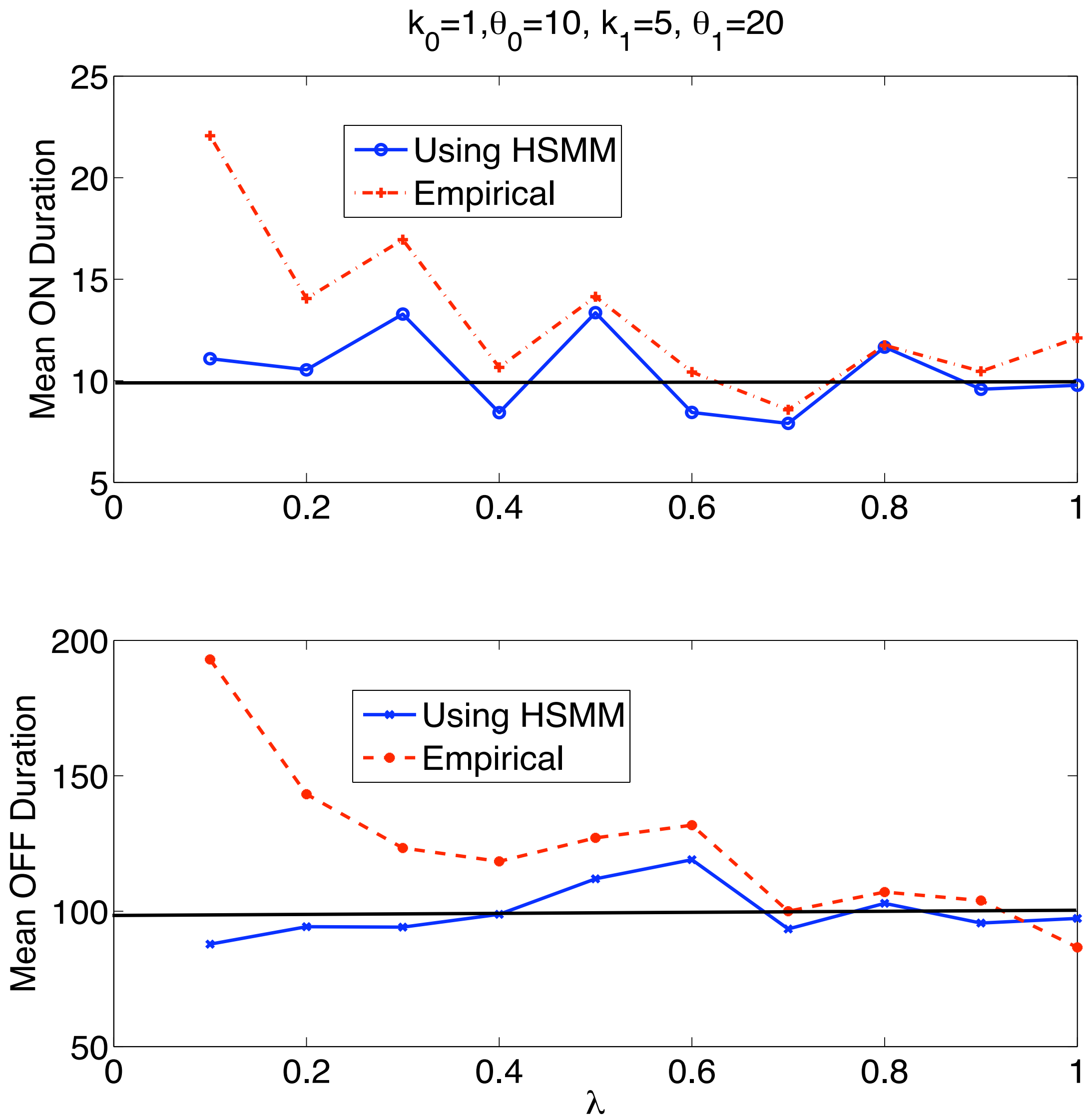
# Simulation

- Alternating ON/OFF renewal process with Gamma intervals, 4 parameters $\{A_i\}:(k_0, \Theta_0)$ and $\{B_i\} : (k_1, \Theta_1)$
- Poisson probes with rate $\lambda$

SAIL works when the model assumptions are correct

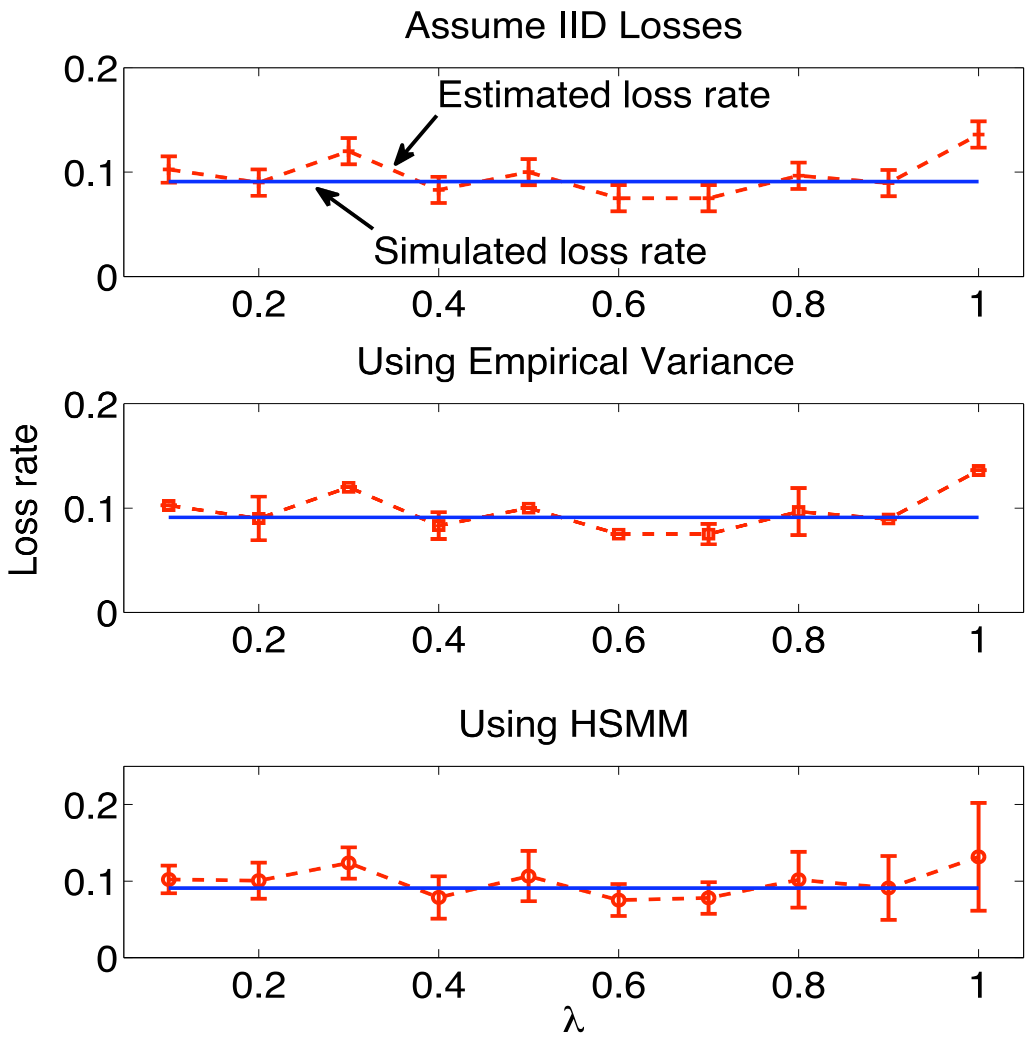


# Simulation- ON/OFF duration




SAIL can correct the missing intervals problem and is needed.

# Simulation- Loss rate



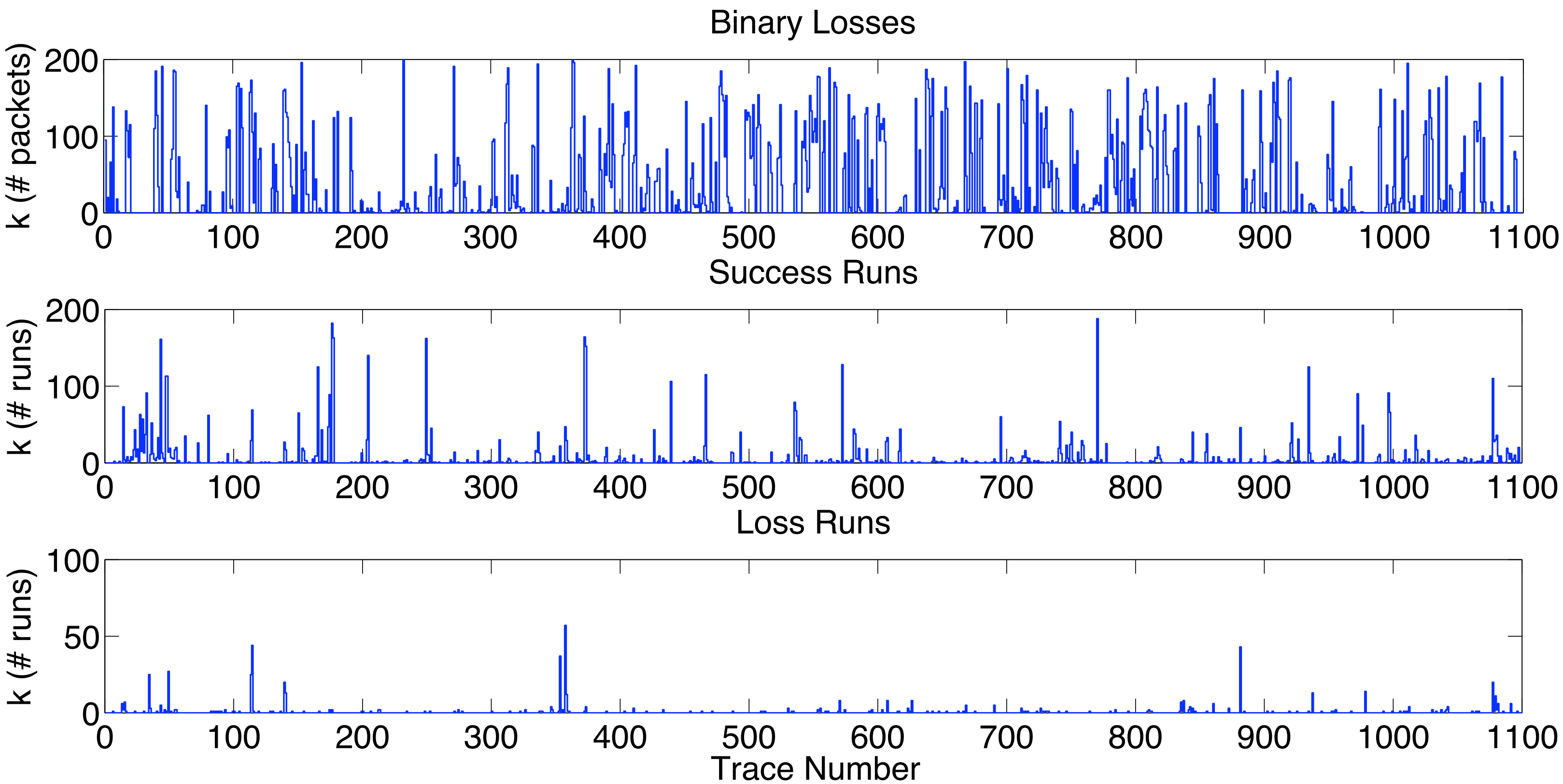

SAIL is more accurate than other methods in computing the statistical errors

# Measurements - Datasets

- UA-EPFL: 1 host at the University of Adelaide and 1 at EPFL, Switzerland
- PlanetLab: randomly selected source and destination pairs
- Poisson probes with small packet size (40 bytes)
- 1 hour traces, in each trace the probing rate is a constant
- Stationarity tests using heuristics (no big/sudden jump and no gradual trend in the moving average loss rate)

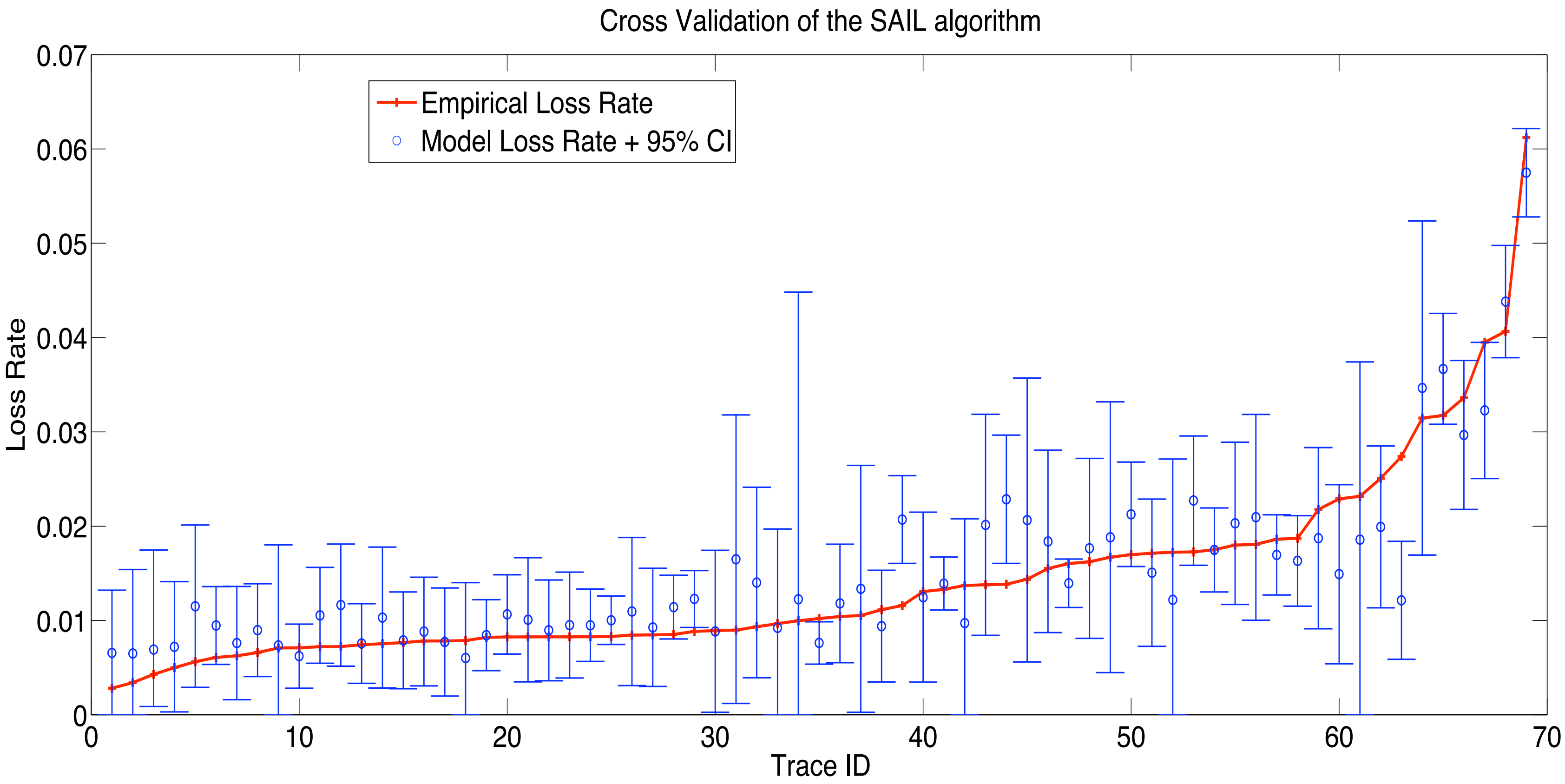
| | UA-EPFL | PlanetLab |
|---|---|---|
| Hours | 100 | 5246 |
| # stationary traces | 10 | 1090 |

# Renewal Properties



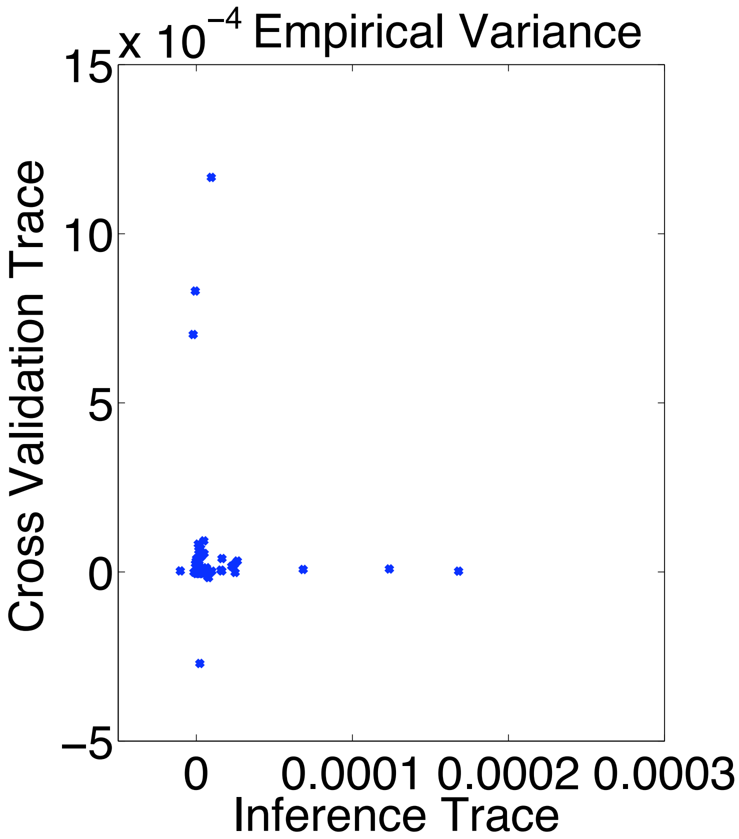

Autocorrelation function test to verify renewal properties

# Cross validation

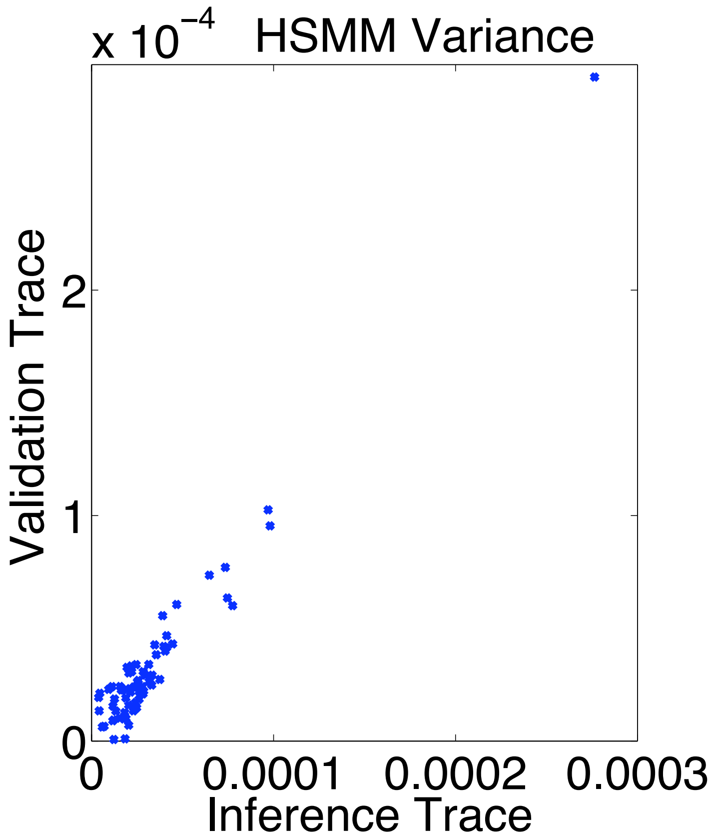



Traces are divided randomly into two sub-segments of equal length. Each sub-segments can be viewed as Poisson samples with rate $\lambda/2$.

# Empirical Variances



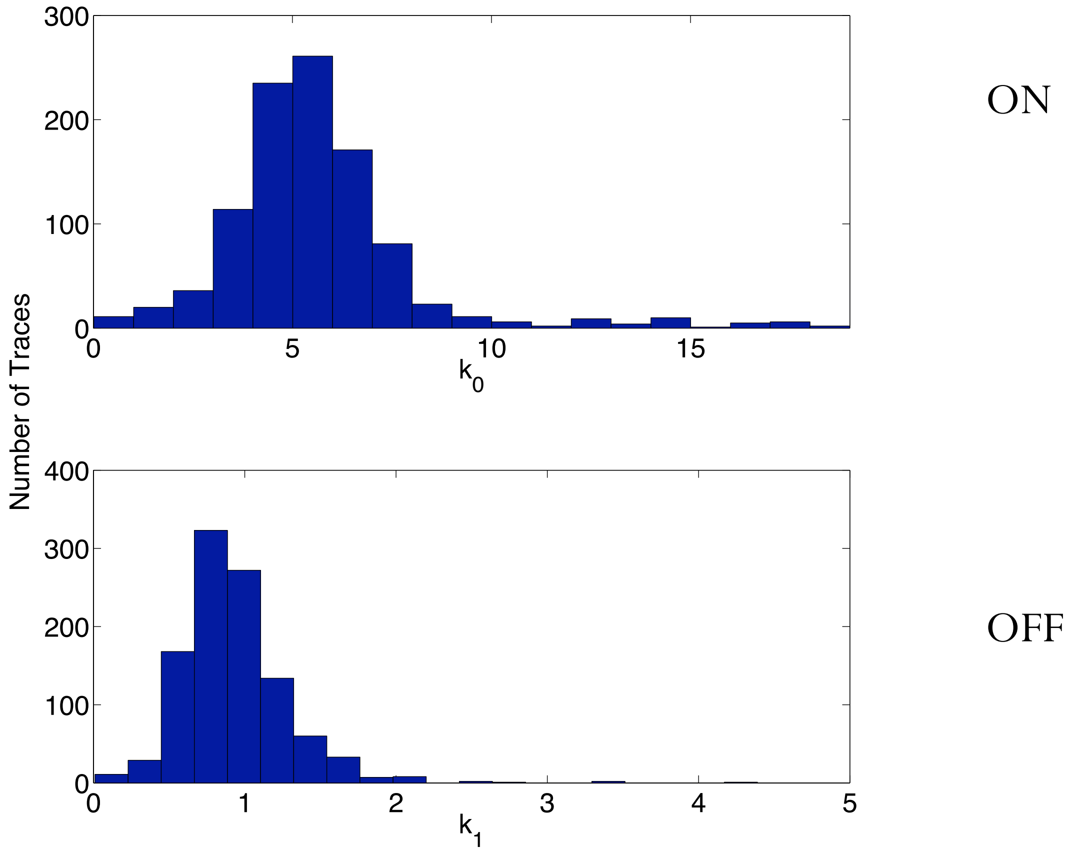

Empirical




SAIL

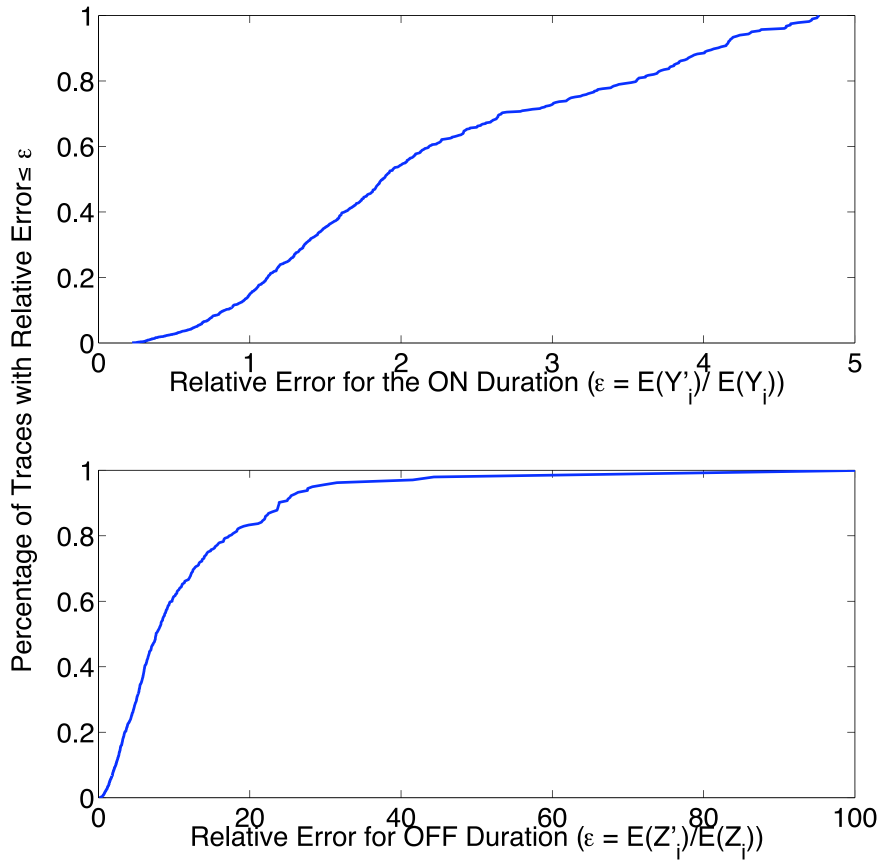
It is important to use a correct method to compute the variance (e.g., SAIL)

# Shape Parameters of the Loss Processes



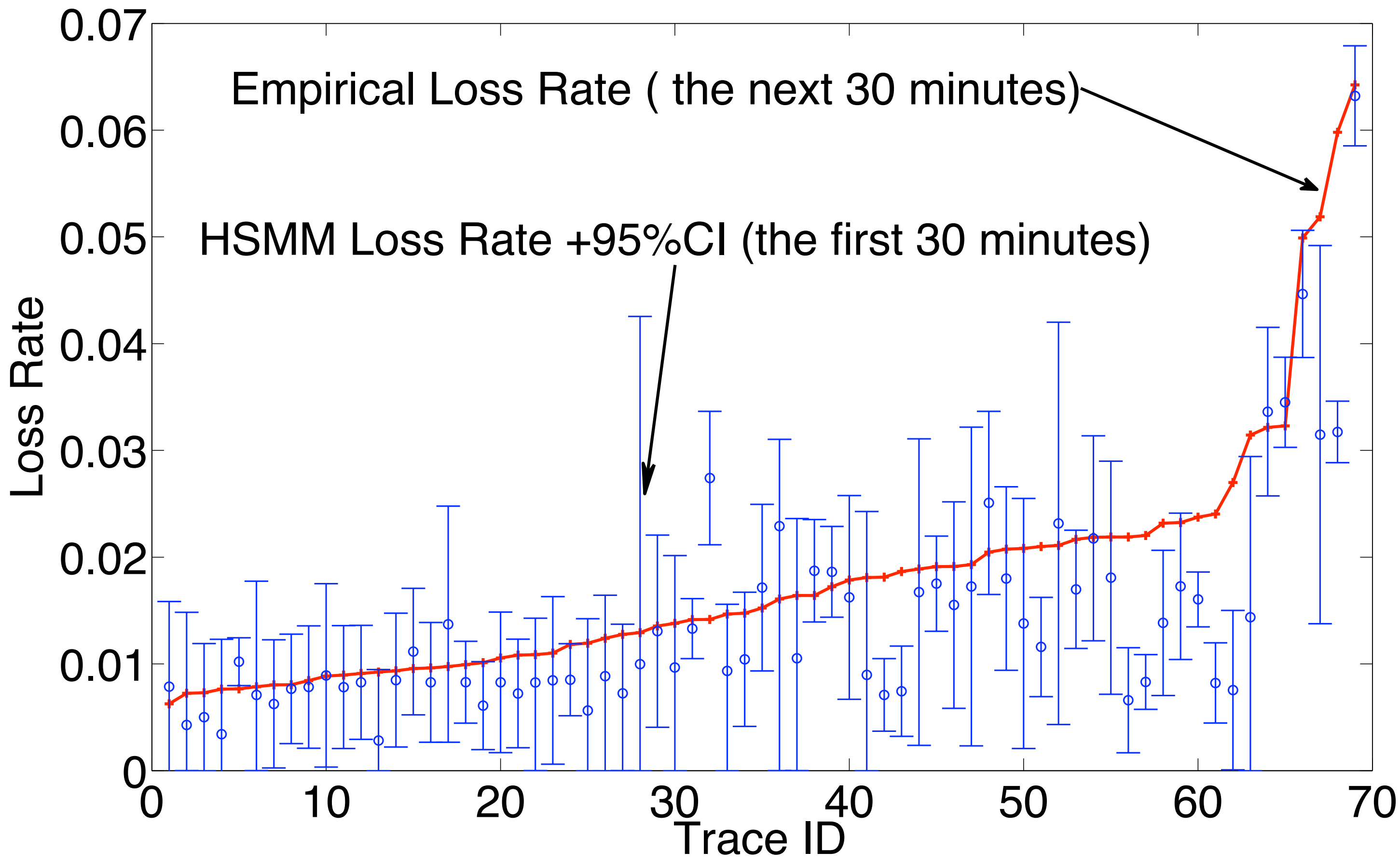

The OFF periods appear to be exponentially distributed

# Errors in Estimating ON/OFF durations



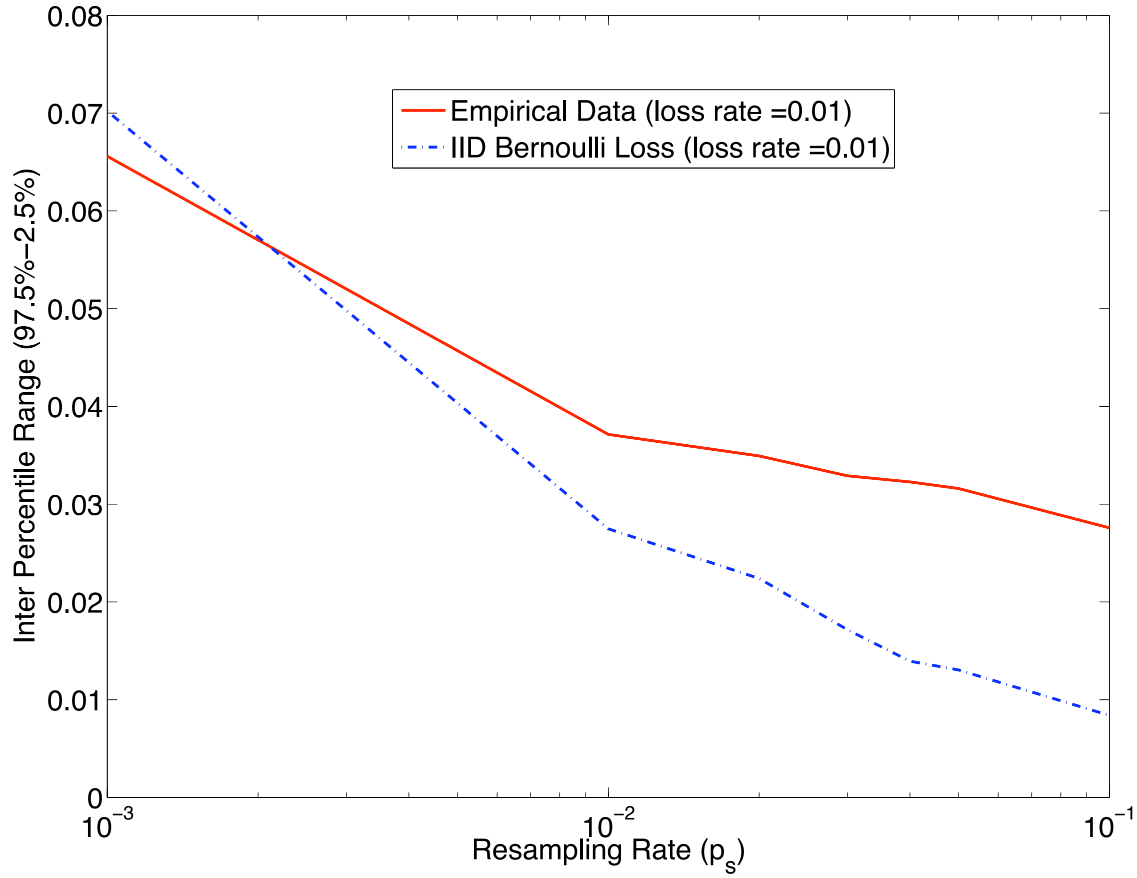

Errors can be quite large because of the missing (short) ON/OFF intervals problem

# Prediction




SAIL can be used to estimate future loss rate

# How Many Probes




Increasing sampling rate only yields small improvements in the variance

# Summary

- SAIL: accurately computes errors in loss estimates
- Better than any existing alternative
- Future work:
  - Faster inference algorithm
  - Non-parametric models for the loss process
  - On-line
  - Make SAIL available to network operators/users
- Code is publicly available, please try
- Thanks!