From measurement to Interpretation

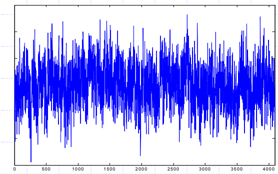
Kavé Salamatian
LIP6-UPMC

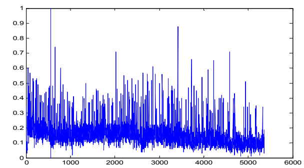
Presentation map

- About interpretation
  - Plato Cave
- Interpretation framework
  - Inverse inference problems
  - Solution methods
- Illustrative Examples

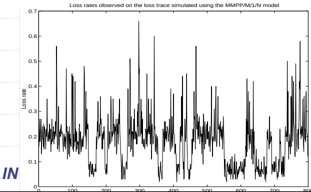
About interpretation

- Measurements
  - But what did they mean ?
- Interpreting?
  - Relating effects to cause
  - Being able to predict the behaviours
    - At different timescales
  - Being able to react
  - Interpretation need *a priori*





Loss rates observed on the loss trace simulated using the MMPP/M/1/N model



Modelling approach in networking

- Descriptive approach
  - Much more used in measurement papers
  - Network is a black box with unknown structure
    - describe observations only through descriptive statistics
      - mean, variance, Hurst or multi-fractal parameters, etc…
  - Top-down approach
    - Begin with observations and derive descriptive parameters
  - Drawbacks
    - It does not explain why?
    - It does not answer what if ?
    - It is difficult to interpret them
      - Interpretation need *a priori*
    - It does not use all the available information
      - We may have *a priori* information on the process generating the observation
- Constructive approach
  - Classical approach
  - Derive IP performance through an explicative model of the process involved into the network
    - Network is constituted of queues and routers, …
      - Uses simulation by ns or analytic queuing theory, network calculus, etc
    - Down top approach
      - Begin from input scenario and network structure and derive performance measures
  - Drawbacks
    - Generalization is difficult
      - Too many parameters
    - Simulation results do not describe real measurements
    - The approach is open-loop

Objectives

- We want to propose a methodology for
  - Interpreting measurement
    - Relating observations to causes
  - Developing realistic models of real network
    - For controlling the QoS in networks
  - Building scenarios for realistic evaluations
    - By using models fed by realistic parameters calibrated over empirical traces

Plato cavern allegory

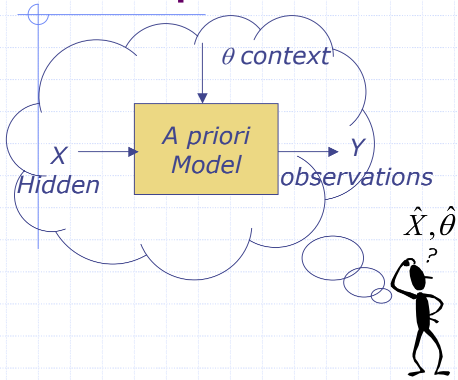
Socrate : « compare our nature in respect of education and its lack to such an experience as this.Picture men dwelling in a sort of cavern ...Picture further the light from fire burning higher up and at a distance behind them, and between the fire and the prisoners and above them ...men carrying past the wall implements of all kinds »

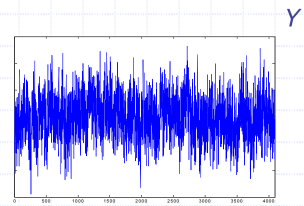
Glaucon : « A strange image you speak of, and strange prisoners. »

Socrate : Like to us, for, to begin with, tell me do you think that these men would have seen anything of themselves or of one another except the shadows cast from the fire on the wall of the cave that fronted them?"

NeXtworking'03 *June 23-25,2003, Chania, Crete, Greece*
The First COST-IST(EU)-NSF(USA) Workshop on EXCHANGES & TRENDS IN NETWORKING Kavé Salamatian

Interpretation framework





- What is the hidden causes (X et θ) that have lead to observing Y
  - *A priori* mode compress all of our understanding of the process involved in the generation of the observation in $Y=M(X,\theta)$

Interpretation

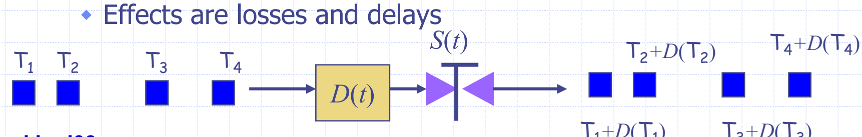
- We have to deal with two main inverse problem
  - Modelling problem
    - What are the best context parameters θ that best describe the environnement
  - Interpretation problem
    - Knowing the context θ what is hidden input X that will best describe observations
- A lot of measurement problem might be expressed in this framework
  - Active measurement interpreting
  - Network tomography

Steps in measurement interpretation

- Choice of the *a priori* model
  - The input-output structure
  - Statistical hypothesis on context θ
  - Statistical hypothesis on input X
- Solving the inverse problems
  - Solving the modelling problem
    - Inferring context parameter θ'
  - Solving the inference problem
    - Inferring input X' that lead to observation of Y
- Evaluating the model
  - Define a quality function $D(Y,Y')$
    - Calculate $Y'=M(X', \theta')$ *et* $D(Y,M(X', \theta'))$

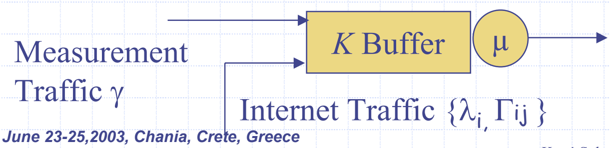
Active measurement

- A probing Agent send packets to destination
  - Each packet is a probe charged by information about the path it crossed
  - At reception loss process and delay are extracted
    - Delays are difficult to measure because of asynchronous clock.
- Underlying model
  - Network is seen by the probing flows through its effects on it.
    - Effects are losses and delays



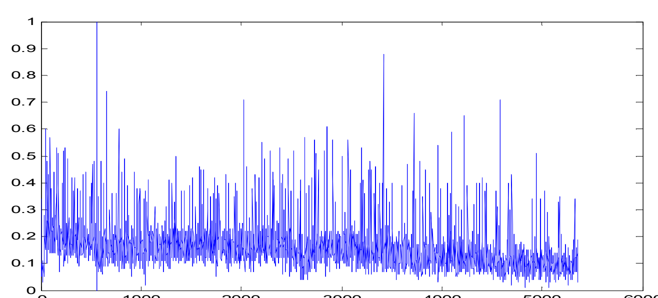
A *priori* model for interpretation

- We assume that the network might be described by a single bottleneck that is fed by an MMPP traffic
  - At each state of the MMPP we have a Poisson traffic of rate λ
  - States follow a markov chain with transition matrix Γ
- Context parameter are $\theta=(\mu, K, \lambda_i, \Gamma_{ij})$
- Input X is the sequence of states of the MMPP



Active Measurement interpretation

- Trace obtained between France and US
  - 50 msec interval, Pkt size = 100 Bytes



- Quality function
  - Mean square error between observed error rate and simulated

Modelling problem

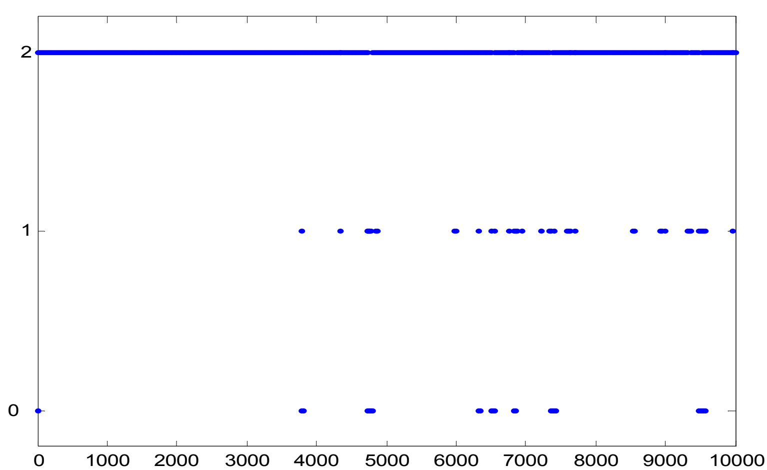
- EM results

ρ=(20, 1.2594,1.07)

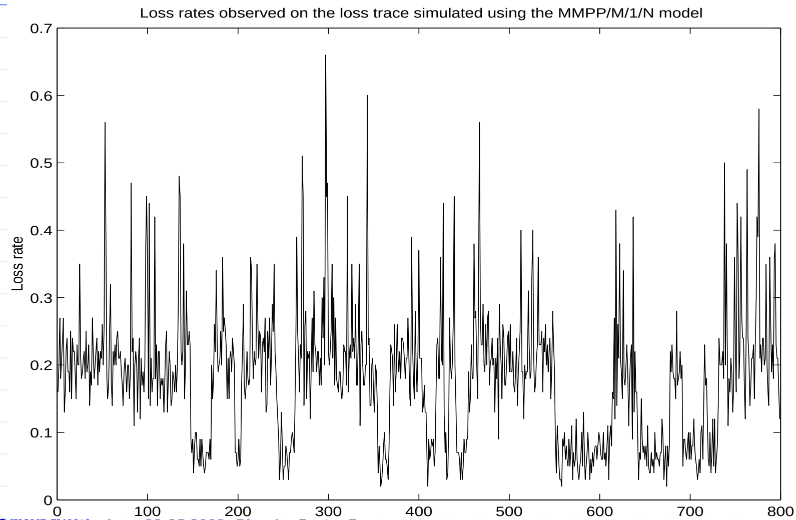
π=(0.03, 0.65, 0.32)

$$Q' = \begin{bmatrix} -0.0651 & 0.0645 & 0.0006 \\ 0.026 & -0.028 & 0.0002 \\ 0.0001 & 0.003 & -0.0004 \end{bmatrix}$$

Interpretation problem

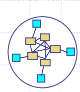


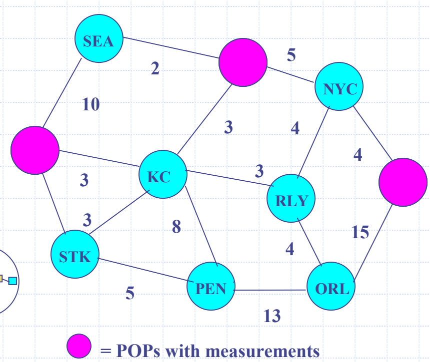
Simulated trace



Network tomography problem

- Goal: Obtain POP-level volume traffic matrices for operational IP networks
- Challenge: Using limited sources of information, "guess" traffic demands
- Network Tomography [Vardi96] : use only link counts

POP = Point of Presence = 



A priori model for tomography

- X_j: Traffic demand for pair of POP j
- A: routing matrix
- Y_i: traffic over link i
  - c = n * (n - 1)

$$\boxed{A_{rxc}X_c = Y_r}$$

- It is highly underdetermined system
- Different hypothesis for X

$X \sim Normal\left(\mu_i, \sigma_i^2\right)$ Cao & coll.

$X \sim Poisson(\lambda_i)$ Tebaldi & West

$X \sim MMPP(\lambda_i, \Gamma_{ij})$ Vaton

A priori model for tomography

- Model

$$\boxed{A_{rxc}X_c = Y_r}$$

  - Context parameters
    - For MMPP example
      - $\theta=(A, \lambda_i, \Gamma ij)$
    - Modelling problem
      - Infer θ
  - Input
    - X: the particular values of X that validate the link constraint $A_{rxc}(t)X_c(t) = Y_r(t)$
- Quality criteria
  - Related to the application of TM estimation

How to solve it ?

- The problems are of statistical inference problem
  - Two main classes of resolution
    - Maximum likelihood
      - θ is an unknown parameter that is to be derived by optimisation
      - EM (*Expectation Maximisation*) method
    - Bayesian
      - θ is a random variable with a known prior distribution
      - MCMC method (Monte Carlo Markov Chain)

Maximum Likelihood

- The a priori model $M(X,\theta)$ give the conditional probability $\Pr\{Y \mid X;\theta\}$ and $\Pr\{X;\theta\}$
- Log-likelihood is defined as

$$L(\theta \mid X, Y=y) = \log \Pr\{X, Y=y;\theta\}$$

- Maximum likelihood criteria is

$$\hat{\theta} = \max_{\theta} L(\theta \mid X, Y=y)$$

  - X is hidden and unknown
- EM method can be applied

Interpretation problem

- EM method give you as a side result the *a posteriori* probability

$$\gamma^{(i)}\left(X\right)=\Pr\left\{X\mid Y=y;\theta=\theta^{(i)}\right\}$$

  - *X* is defined using the maximum a posteriori criteria $\hat{X}=\max_{X}\Pr\left\{X\mid Y=y;\theta=\hat{\theta}\right\}$
- Sometimes other criteria as Maximal entropy might be used
- Morale
  - EM method is powerful
    - Local minima problem
      - Extension TO SAEM
    - Initial point is important $\theta^{(0)}$
      - Clever choice is mandatory

Bayesian approach

- Bayes law

$$\Pr\left\{A\mid B\right\}=\frac{\Pr\left\{A\mid B\right\}\Pr\left\{B\right\}}{\sum_{E}\Pr\left\{A\mid E\right\}\Pr\left\{E\right\}}$$

- We assume that θ is a RV
  - θ has a distribution Pr{ θ }
- Modelling and interpretation problem are dealt by a maximum *a posteriori criteria*

$$\Pr\left\{\theta\mid Y=y\right\}=\frac{\sum_{X}\Pr\left\{Y=y\mid X,\theta\right\}\Pr\left\{X\mid\theta\right\}\Pr\left\{\theta\right\}}{\sum_{X,v}\Pr\left\{Y=y\mid X,v\right\}\Pr\left\{X\mid v\right\}\Pr\left\{v\right\}}$$

$$\Pr\left\{X\mid Y=y\right\}=\frac{\sum_{\theta}\Pr\left\{Y=y\mid X,\theta\right\}\Pr\left\{X\mid\theta\right\}\Pr\left\{\theta\right\}}{\sum_{X,v}\Pr\left\{Y=y\mid X,v\right\}\Pr\left\{X\mid v\right\}\Pr\left\{v\right\}}$$

Conclusion

- We are all in the cavern
  - Who is running the shadows !!!!
  - The main challenges of the coming years in Internet measurement will be to develop interpretative model and to solve the interpretation problems
- A lot of models have been defined without validation by application
  - Another challenge will be evaluate empirical models on real application to see which one are good enough