La métrologie sur les réseaux : enjeux et quelques problèmes intéressants

Projet METROPOLIS
METROlogie Pour L'Internet et les Services

Objectifs

- Développer un cadre commun pour la métrologie des réseaux IP
  - *Mesure de la Qualité de Service*
  - *Développement de modèles réalistes*
  - *Analyse des protocoles et du comportement du réseau*
  - *Dimensionnement des réseaux*

Partenaires du projet

- LIP6 (coordinateur)
- France telecom R&D
- Le GET
- INRIA
- Institut Eurecom
- LAAS
- Renater

Organisation du projet

- SP 1 : *Etat de l'art sur la métrologie dans les réseaux*
- SP 2 : *Classification et dimensionnement*
- SP3 : *Analyse du réseau*
- SP4 : *Méthodes pour la mesure et échantillonnage*
- SP5 : *Modélisation*
- SP6 : *Tarification et SLA*
- SP7 : *Plate-formes de mesures*

SP 1 : Etat de l'art sur la métrologie dans les réseaux

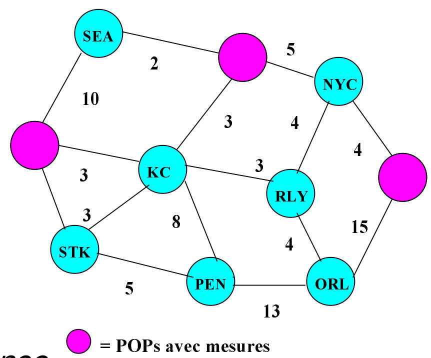
- Leader: LIP6
- Participant : LIP6, FT R&D, LAAS, GET
- Objectifs
  - Effectuer un état de l'art
    - Critères de performances IP
    - Mesures actives/passives
    - Architecture de mesure
    - Modélisation empirique

SP 2 : Classification et dimensionnement

- Leader : France Telecom R&D
- Participant : FT R&D, Eurecom, GET
- Objectifs
  - Effectuer un classification des flots
    - Granularité
    - Type d'application
    - Protocoles
  - Définir des méthodes de dimensionnement et valider celle ci
    - Répartition de charge

Dimensionnement du réseau

- Estimation des matrices de trafic
  - Network tomography
- Determiner la matrice de trafic
- Le trafic de chaque lien est connu
- Objective
  - Estimer la matrice de trafic avec des informations limitées

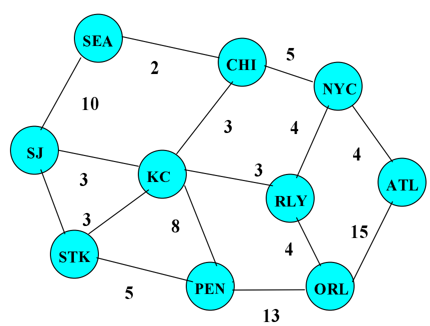


POP = Point of Presence

Notations

- X_j: Demande de trafic pour la paire de POP j
- A: matrice de routage
- Y_i: trafic sur le lien i
- $c = n * (n - 1)$

$$A_{rxc} X_c = Y_r$$



Techniques de résolution

- $A_{rxc} X_c = Y_r$ est un système linéaire fortement sous-déterminé
  - La solution est une optimisation statistique
- Deux directions, 3 techniques :
  - Approche déterministe :
    - Programmation linéaire
  - Approche Statistique :
    - Bayesienne [Sandrine Vaton]
    - Approche EM

Approche EM

- Suppose $X \sim \text{Normal}(\lambda, \Sigma)$.
- Avec $\lambda = (\lambda_1, ..., \lambda_c)$ and $\Sigma \phi \text{diag}(\lambda_1^b, ..., \lambda_c^b)$
- Estimation MLE de θ par Algorithme EM
- L'algorithme nécessite un bon point de départ (A priori nécessaire)
  - Minimum locaux
- Estimation des composants de X par:

$$\hat{X}_j = E[X_j \mid \hat{\theta}, Y]$$

Classification de flots

- Classification comportementale
  - Permettre de détecter des attaques par changement de classe.
  - Étude des éléphants souris.
- Rasoir d'Occam
  - Agréger les données pour une interprétation plus facile.

Elephant hunting

- One of the few invariants of Internet Traffic is "the elephants and mice phenomenon"
  - A few percentage of flows contributes to a large proportion of total traffic
  - Results from heavy tails in traffic pdf
- Elephants detection can be a base for differentiated treatment
  - Elephant can be routed differently
  - Load balancing is easier on smaller number of flows
- We want to define a methodology for classifying traffic to elephants and mices (maybe rabbits, foxes, etc...)

Elephant hunting

- The aim is to find a threshold T such that
  - If $R(n) > T$ then flow n is an elephant
- Traffic classification
  - Parametric methods
    - Mixture models + Bayesian classification
  - Non parametric methods
    - Clustering
      - K-means
    - NN
      - learning by back propagation
    - Heuristics
      - definition of threshold by heuristic arguments

Traffic data

- A flow is defined by the BGP prefixes
  - BGP prefix are extracted from the routing table
  - Flows are aggregated over a granularity time step τ
    - SNMP use generally $\tau = 5$ min
    - The optimal granularity is an open problem

Gaussian Mixture Model

- GMM model

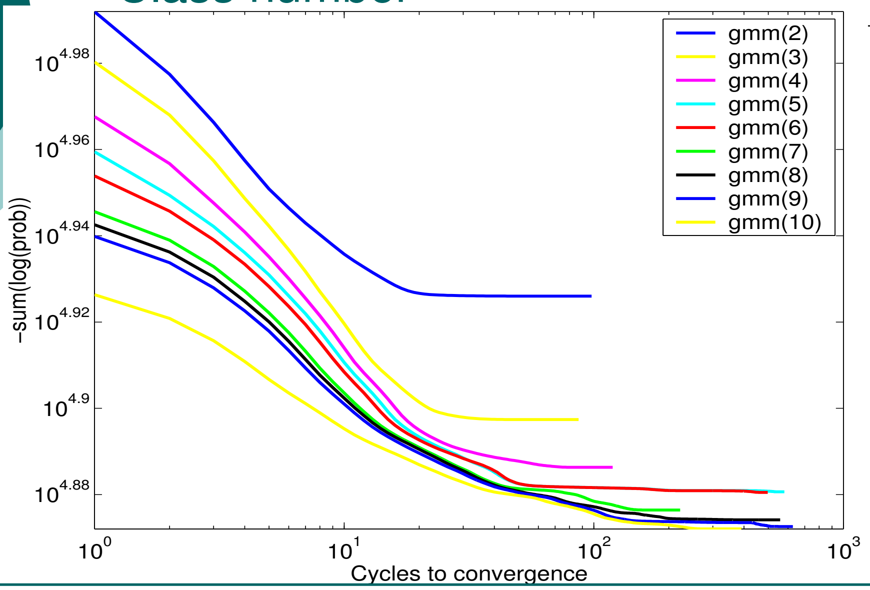
$$f(x)=\sum_{i=1}^{K} p(C_K)N(\mu_i,\sigma_i)$$

- The pdf is approximated by a mixture of gaussian
  - Every (smooth) pdf can be approximated uniformly as a mixture of (maybe infinite) number of gaussian
    - Radial Basis Function
  - Central Limit theorem
    - Every traffic regime will converge to a gaussian pdf over a given granularity

GMM calibration

- GMM parameter can be calibrated by EM methods
- We use a matlab toolbox (*netlab*) for parameter calibration
- An OC-12 link is analysed
  - The East-West and the West-East link are analysed

Class number



Calibration results

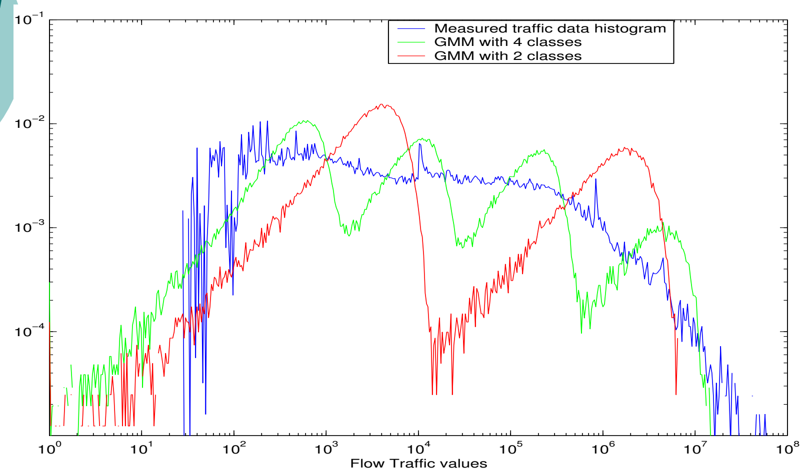
- GMM with 2 classes

$$\begin{cases} P(C_1) = 0.33, \mu_1 = 449 \text{ KBytes/sec}, \sigma_1 = 1.63 \text{Mbytes/sec} \\ P(C_2) = 0.33, \mu_2 = 2 \text{ KBytes/sec}, \sigma_2 = 3 \text{Kbytes/sec} \end{cases}$$

- GMM with 4 classes

$$\begin{cases} P(C_1) = 0.44, \mu_1 = 398 \text{ Bytes/sec}, \sigma_1 = 345 \text{Bytes/sec} \\ P(C_2) = 0.28, \mu_2 = 7.41 \text{ KBytes/sec}, \sigma_2 = 6.42 \text{Kbytes/sec} \\ P(C_3) = 0.22, \mu_3 = 135.7 \text{ KBytes/sec}, \sigma_3 = 12 \text{KBytes/sec} \\ P(C_4) = 0.06, \mu_4 = 2.02 \text{ MBytes/sec}, \sigma_2 = 3.47 \text{Mbytes/sec} \end{cases}$$

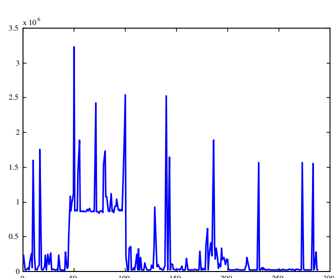
Pdf approximation



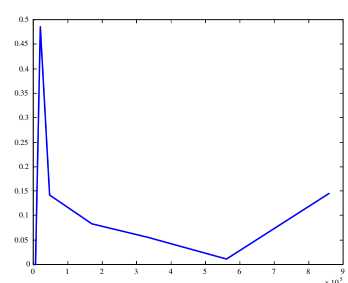
Quoi classifier ?

Une distribution

Utilisation d'une distribution de valeurs.



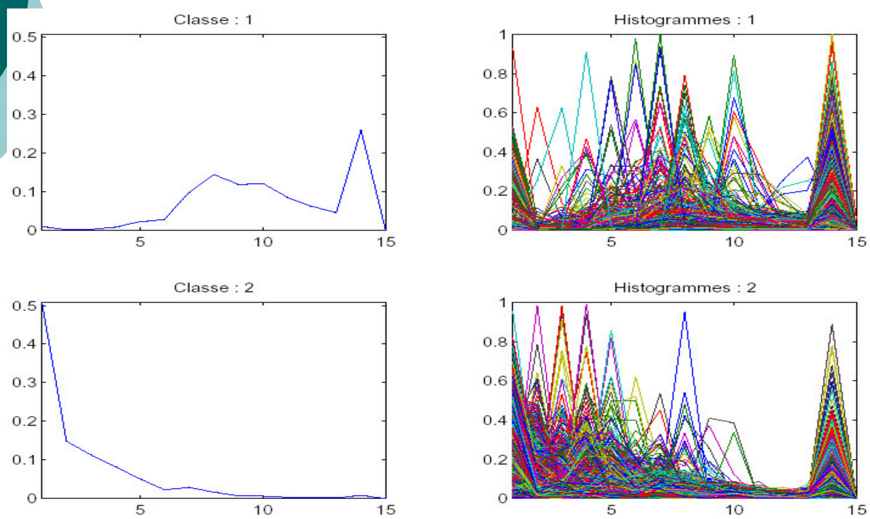
Flot originel



Distribution

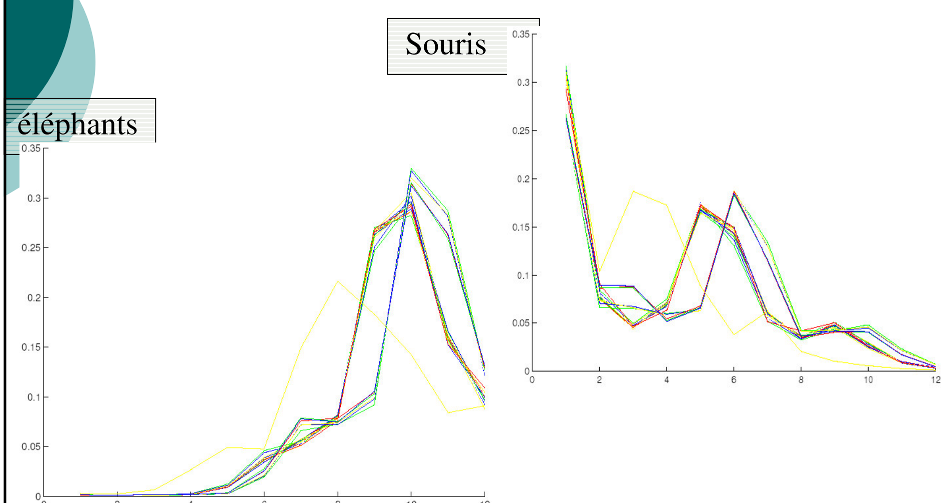
Qu’est-ce que cela donne ? (1/2)

La chasse aux éléphants

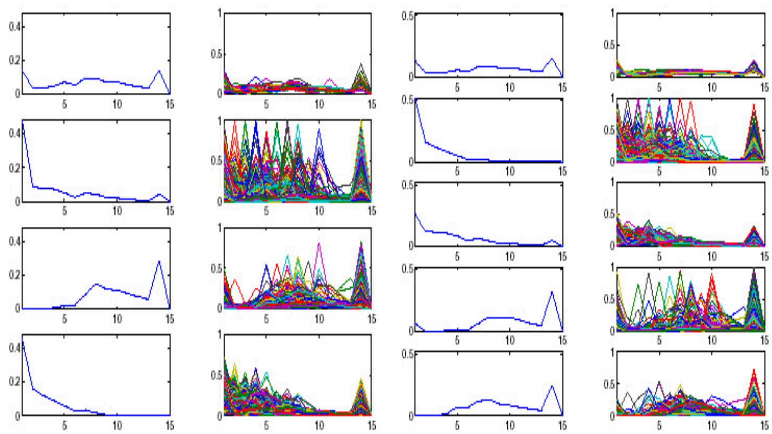


Qu’est-ce que cela donne ? (2/2)

Résiste au temps



Nombre de classes



4 classes 5 classes

SP 3 : Analyse du réseau

- Leader : LAAS
- Participant : LIP6, FT R&D, Eurecom, GET
- Objectifs
  - Analyser les mesures passives
    - Analyse des flots TCP
    - Analyse des délais dans les routeurs
    - Etude des attaques

Challenges existant

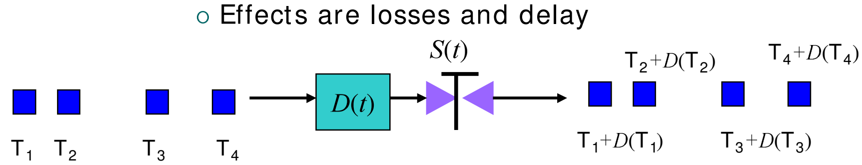
- Ingénierie de trafic
  - Comment utiliser la matrice de trafic pour concevoir un réseau
  - Comment prédire à moyen et long terme l'évolution du réseau
    - Estimation de la demande de trafic
- Classification applicative et comportementale
  - Détecter les comportements déviants
    - Attaque
    - SLA violés
    - Kazaa et quoi d'autres ???

SP 4 : Méthodes pour la mesure et échantillonnage

- Leader : GET
- Participant : LIP6, FT R&D, GET
- Objectifs
  - Développement d'une théorie de l'échantillonnage
    - Similaire à la théorie classique dans le traitement de signal
  - Echantillonnage
    - Spatial
    - Temporel
    - Applicatif

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
- Underlying model
  - Network is seen by the probing flows through its effects on it.
    - Effects are losses and delay



Sampling

- The active probe send at time T_i give two samples $D(T'_i)$ and $S(T'_i)$
  - T_i- T'_i is the time needed to reach the bottleneck
  - We will suppose that $T_i = T'_I$
  - The process $\{ T_i \}$ is a renewal process with live distribution $G(\tau)$

$$G(\tau) = \Pr\{T_{i+1} - T_i < \tau\}$$

- Inverse problem
  - Inverse problem: Based on the observed samples processes $S(T_i)$ and $D(T_i)$ infer about statistical properties of underlying network effect $S(t)$ and $D(t)$

State estimation

- Main Hypothesis
  - The process $S(t)$ has reached to a steady and stationary distribution
- The steady state distribution is function of
  - the background traffic sharing the path
  - the application traffic
- Active measurement try to reduce the effect of probing traffic by maintaining a low sending rate

Generalisation

- How to relate inference made on the probing flow to other flow ?
- Probing flows samples $S(t)$ at time $\{T_i\}$
  - A concurrent flows see $S(t)$ at time $\{T_j\}$
- If the process $S(t)$ is stationary and ergodic
  - Temporal mean $\bar{S} = \frac{1}{n}\sum_{k=1}^{n} S(T_{j+k})$ is an unbiased estimator for $\mu = E\{S(t)\}$
  - Variance of this estimator is $\mathrm{var}\{\bar{S}\} = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} C(|k-l|)$ where $C(|k-l|) = E\{(S(T_k)-\mu)(S(T_l)-\mu)\}$ is the autocorrelation function of $S(t)$

Generalisation

- Estimation variance can be expressed as a function of inter packet time distribution

$$\operatorname{var}\{\overline{S}\} = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\int C(\tau)\,dG^{(|k-l|)}(\tau)$$

Where $G^{(n)}(\tau)$ is the convolution of n $G(\tau)$ distribution

- For the simple (and general) case of exponential live time distribution we have

$$G(\tau) = 1 - e^{-\tau/\bar{\tau}} \qquad G^{(n)}(\tau) = erl_n(\tau)$$

Challenges existant

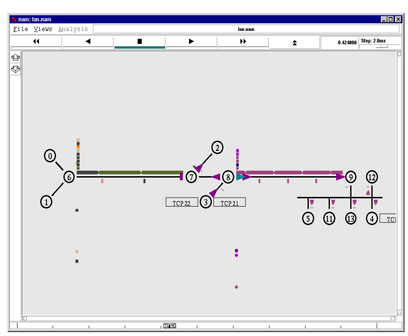
- Développer une théorie de l'échantillonnage pour la mesure dans le réseau
  - PASTA ou PIZZA ????
- Comment échantillonner dans un graphe
  - Quelle est la taille d'Internet ?
  - Quelle sont les performances d'un algorithme de routage
- Comment générer une topologie de réseau

SP 5 : Modélisation

- Leader : INRIA
- Participant : LIP6, FT R&D, INRIA, LAAS
- Objectifs
  - Développer des modèles réalistes du réseau
    - Macroscopique
      - Modèle de flots
    - Microscopique
      - Modèles de TCP
  - Comment passer du micro au macro
    - Vers une théorie macroéconomique du réseau

Modeling approaches

- Constructive approach
  - Classical approach
  - derivate 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 scenarii and and find performance measures
- Drawbacks
  - Generalization is difficult
    - Too much parameters
  - Simulation results do not describe real measurements

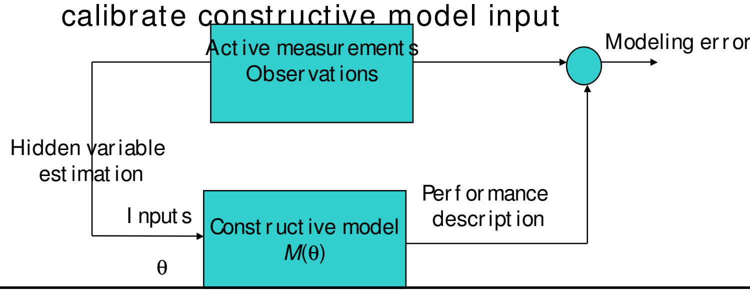


Modeling approaches

- Descriptive approach
  - Much more used in measurement papers
  - Network is a black box with unknown structure
    - describe observations only through descriptive statistical
      - mean, variance, Hurst or multi-fractal parameters, etc…
  - Top-down approach
    - Begin with observations and derivate descriptive parameters
  - Drawbacks
    - It does not explain why?
    - It does not answer what if ?
    - It is difficult to interpret them
      - Interpretation need *a priori* and *a priori* needs constructive model
    - It does not use all the available information
      - We have some *a priori* about the process generating the observations

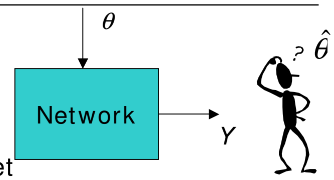
Modeling of observations based on *a priori* constructive model

- Mixing up the constructive and descriptive approach
  - Interpreting the measurement using an *a priori* constructive model
  - Inverse model approach
    - We begin with the observations and we calibrate constructive model input



How to interpret measurement ?

- Interpretation?
  - Be able to relate effect to causes
  - Be able to use proactive control
  - Be able to predict the behavior
    - In different time scale
- Active measures are difficult to interpret
  - They reflect other phenomenon
  - They are very dynamic
  - They need models to be interpreted
  - It is difficult to evaluate the effect of measurement
- Passive measurement are also difficult to interpret
  - They are local
  - They are huge
    - 500 Gb per day for micro measure
  - Should be related to macro and active measurement
- Active and passive measurement should be related to make sense



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

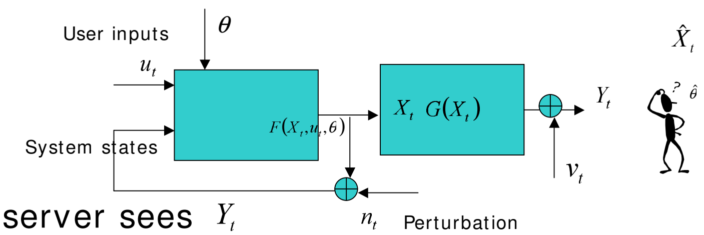
Statistical inference

- A parameter θ
- A sequence of observation *X* related to θ by a stochastic function $f(\theta)$ (a pdf)
  - Observations may be independent or correlated
  - Function $f(\theta)$ give a description of the observation *X*
- Statistical inference consist of using the information in *X* to guess the unknown (and unobservable) θ
  - θ is a hidden variable
- Optimal statistical inference will try to minimize an error criteria
  - Least square, Maximum Likelihood, …
    - Trying to find the input parameter θ that will generate the closer outcome to observations

Estimation theory

- If the θ to guess is the value of X_{i+1} based on previous value $X_1,\ldots,X_i$
  - If X_i's are independent
    - Best estimate is $\theta = E\{X_{i+1}\}$
  - If X_i's are correlated
    - Best estimate is $\theta = E\{X_{i+1} | X_i, \ldots, X_1\}$
- If we know or have an *a priori* about the way X_i's are generated we can do even better
  - *A priori* constructive model
  - Kalman filtering

General framework



- Observer sees Y_t
- He uses an *a priori* model of the network process *F*(.)
  - Classical state based model that may be non-linear
    - Linearization around equilibrium ?
  - n_t may represent modeling error
- Objectives
  - To infer about X_t and θ

Asymptotic observer

- Let $e_t = X_t - \hat{X}_t$ the estimation error
- Strictly constructive model
  - Without any feedback reuse of observations estimation error will propagate
- Feedback based model

$$\hat{X}_{t+1} = F\left(\hat{X}_t, u_t, \theta\right) + \overline{K}\left(y_i - \hat{y}_i\right)$$
$$e_{t+1} = F'(e_t, \theta) - \overline{K}G(e_t)$$

  - K should be chosen so that $\lim_{t \to \infty} e_t = 0$ even in presence of disturbance and measurement noise
- System theory derives condition of existence of such K

Inference on parameter θ

- System identification framework
  - Find parameter θ that generate outcomes as close as possible to observations
  - MSE criteria

    $$\hat{\theta} = \arg\min \left(\hat{Y}(X(\theta)) - Y\right)^2$$

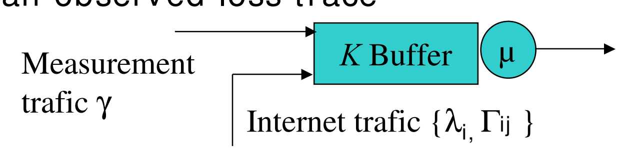
    - May be difficult to implement in general framework
  - Maximum likelihood criteria

    $$\hat{\theta} = \arg\max L(X, Y; \theta)$$

    - Based on Expectation-Maximization algorithm
    - The sequence of states X_t is supposed as missing
      - EM methods find also the most likely sequence of states

Case study: modeling of losses on an Internet Path

- We suppose as an *a priori* constructive model that the network is a single bottleneck queue fed by an MMPP trafic
  - Each state of the MMPP trafic is a Poisson trafic with rate λ
  - State transition follows a Markov Chain
- We want to calibrate the set of rate λ_i and the transition matrix such that it follows an observed loss trace



Losses with MMPP trafic

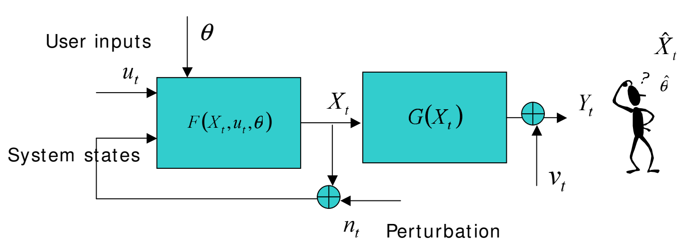
- Losses of measurement flow in an M/M/1/K queue

$$P = \frac{(1-\rho)\rho^{M}}{1-\rho^{M+1}} \approx 1 - \frac{1}{\rho} \qquad \rho = \frac{\gamma + \lambda}{\mu}$$

- Losses of measurement flow in an MMPP queue
  - If measurement packet interval T is larger than the queue time constant $\frac{K}{\mu}$
    - Losses will follows a HMM with transition Γij and loss probability in each state

$$P_i \approx 1 - \frac{1}{\rho_i} \qquad \rho_i = \frac{\gamma + \lambda_i}{\mu}$$

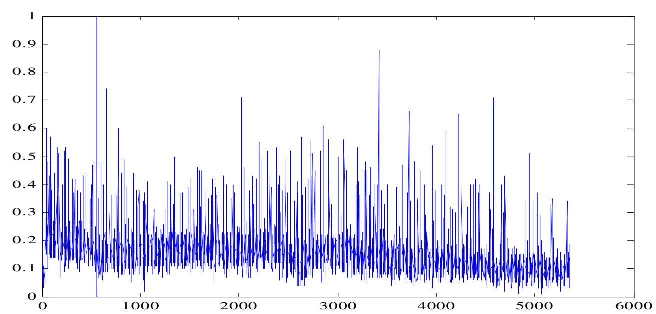
Application of the framework



- X_t is the state of the queue (full or not)
- θ is the parameter of the MMPP
- Function F(.) represent the dynamic of the queue
- Function G(.) is a mean calculator
- X_t represent the modeling error

Experimental results

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



Modeling results

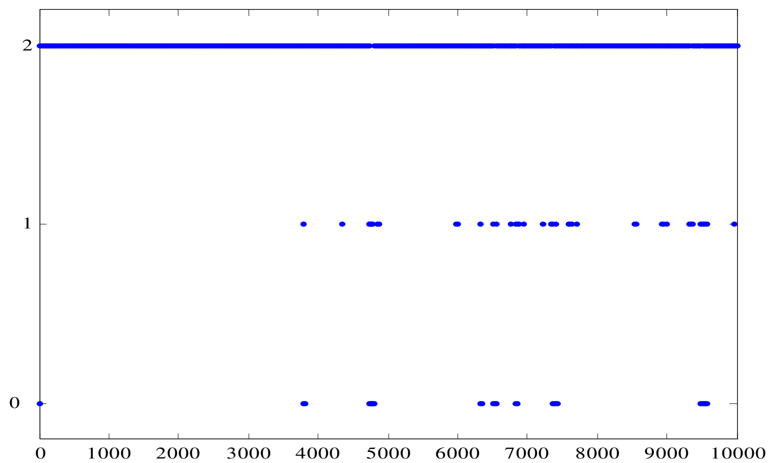
- The EM results in the following

  $\rho = (20, 1.2594, 1.07)$

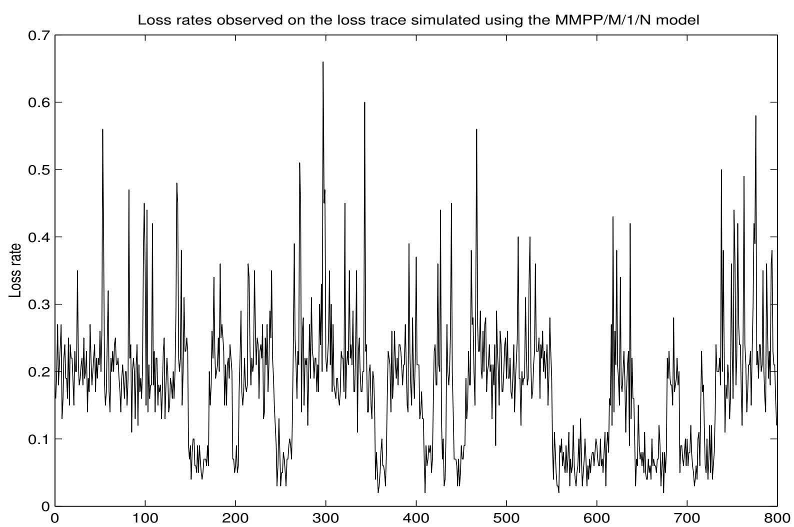
  $\pi = (0.03, 0.65, 0.32)$

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

State transitions



Simulated loss trace



Conclusion

- Forte interaction entre le domaine du traitement de signal et du réseau
  - Théorie de l’échantillonnage
  - Traitement statistique
    - Modélisation
    - Inférence
- Problèmes intéressants à investiguer