Angelo Coluccia[1], Peter Romirer-Maierhofer[2], Fabio Ricciato[1,2]
[1]Univ. of Salento, Lecce, Italy
[2]FTW, Vienna, Austria

BAYESIAN ESTIMATION OF NETWORK-WIDE MEAN FAILURE PROBABILITY IN 3G CELLULAR NETWORKS

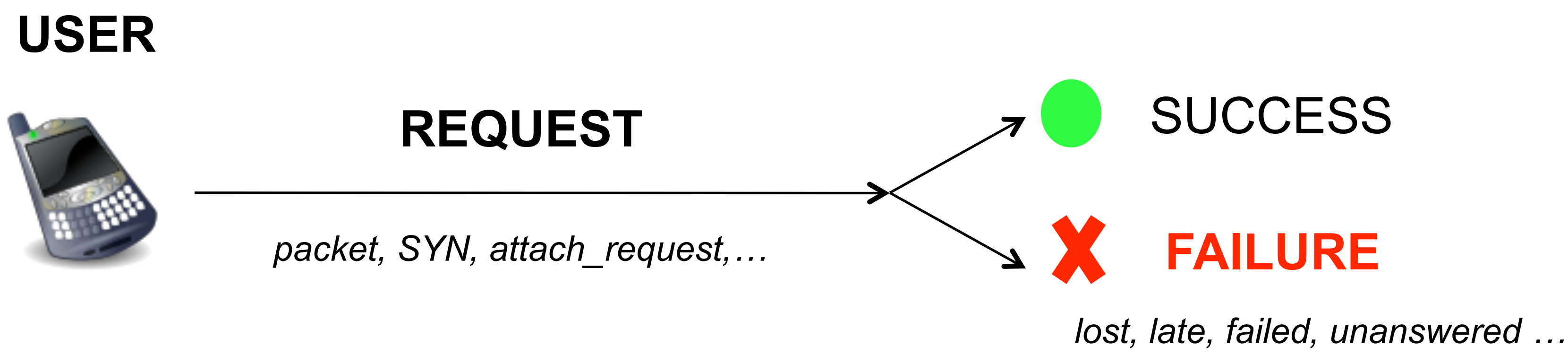
Motivations

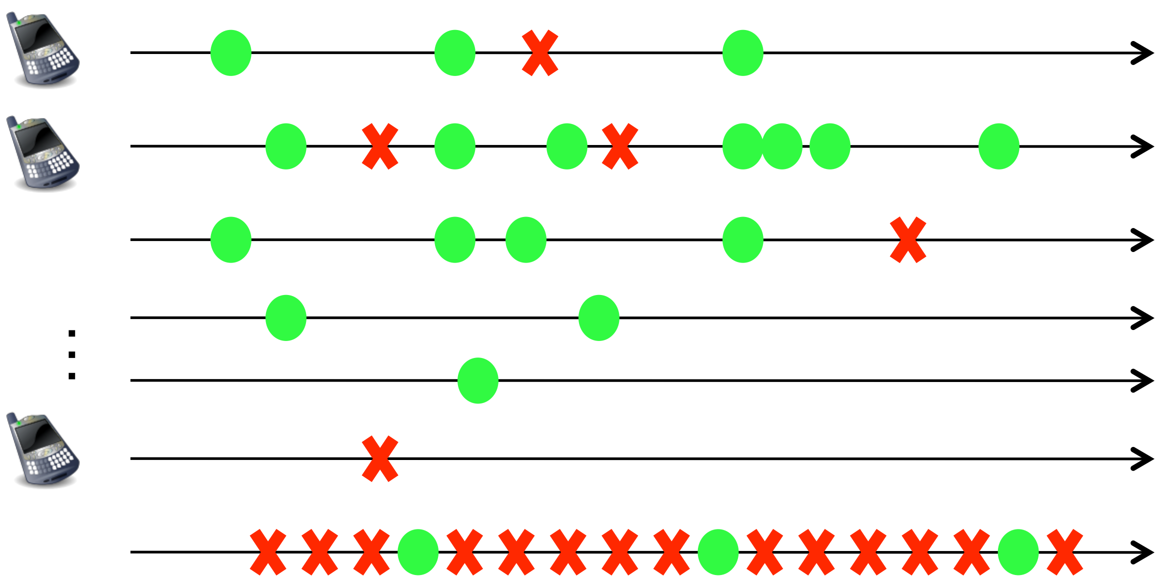
- Third-generation (3G) cellular networks
  - functionally complex, fast changing
  - exposed to errors, failures and attacks
- Early detection of network problems
  - is fundamental to network operation
  - often based on passive traffic monitoring: packet capture, log data
- Key Performance Indicators (KPI)
  - common tool to concisely quantify network performances
  - for comparison/benchmarking between different network areas
  - for detecting drifts and/or sudden changes

Motivations

- Global Percentages of “what goes wrong”
  - most common class of KPI
  - refer to events with binary output: **success** or **failure**
  - Examples
    - % of lost packets (over transmitted ones)
    - % of late packets with delay > *Th* (over received ones)
    - % of unanswered TCP connection attempts (over total SYN openings)
    - % of failed attach_requests (over total attach_requests)
    - …

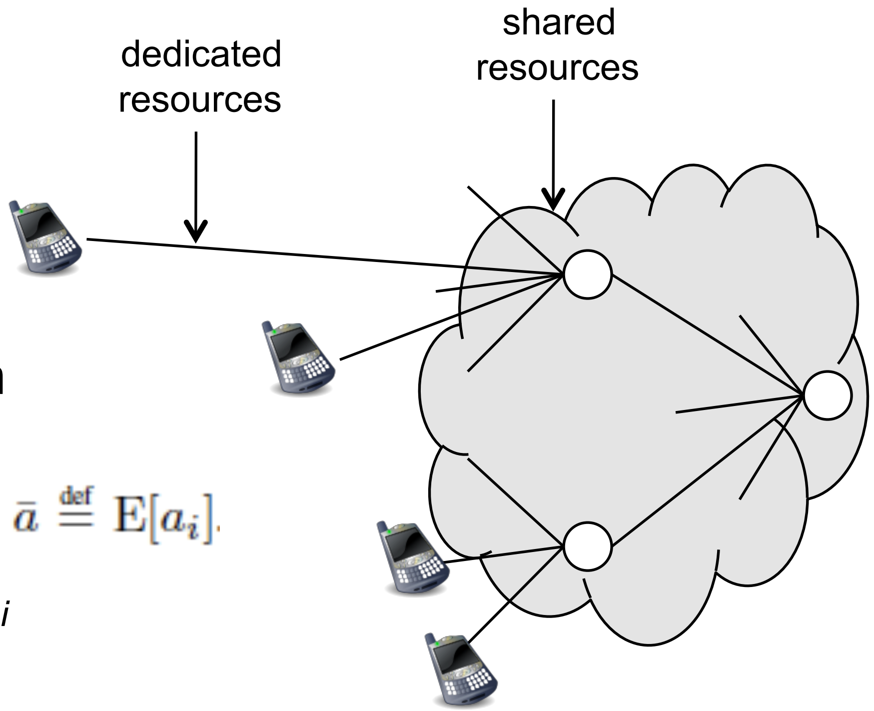
System Modeling





System Modeling

- Notation
  - in a generic timebin *t* user *i* (*i=1...I*)
  - generates n_i "requests" (e.g. packets)
  - out of which m_i "fail" (e.g. lost)

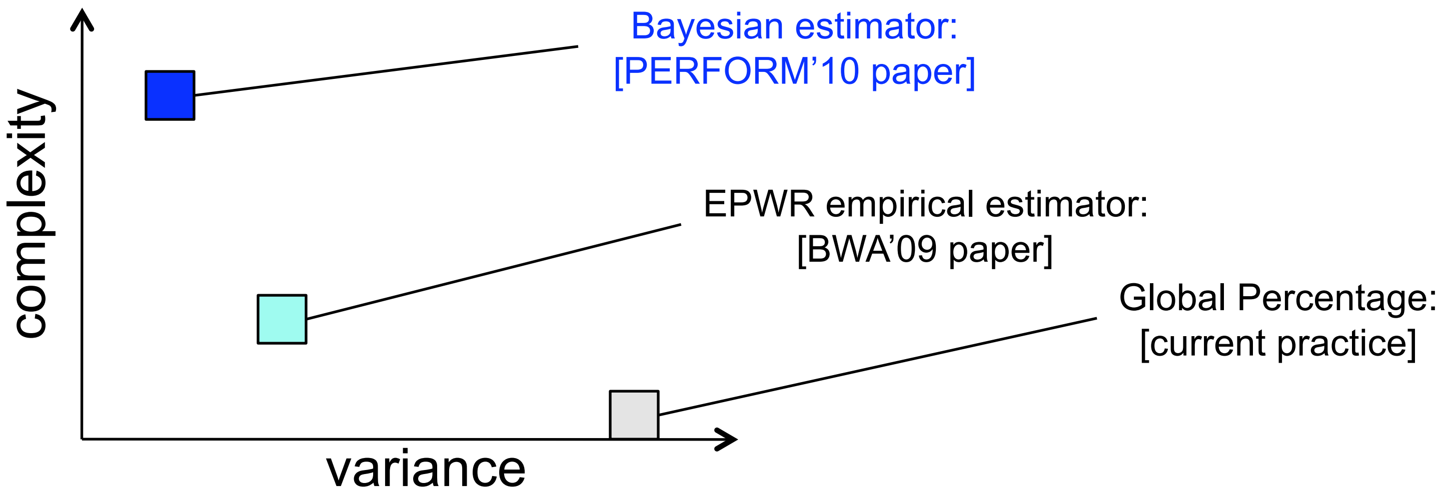


- Assumptions
  - failures are <u>independent</u> and occur with (unknown) probability a_i
  - a_i's are iid random variables with mean $\bar{a} \stackrel{\text{def}}{=} \mathrm{E}[a_i]$
  - independency between traffic volume n_i and failure prob. a_i

- Goal: estimate $\bar{a} \stackrel{\text{def}}{=} \mathrm{E}[a_i]$ given a set of measurements $\{n_i,m_i\}$

Estimators

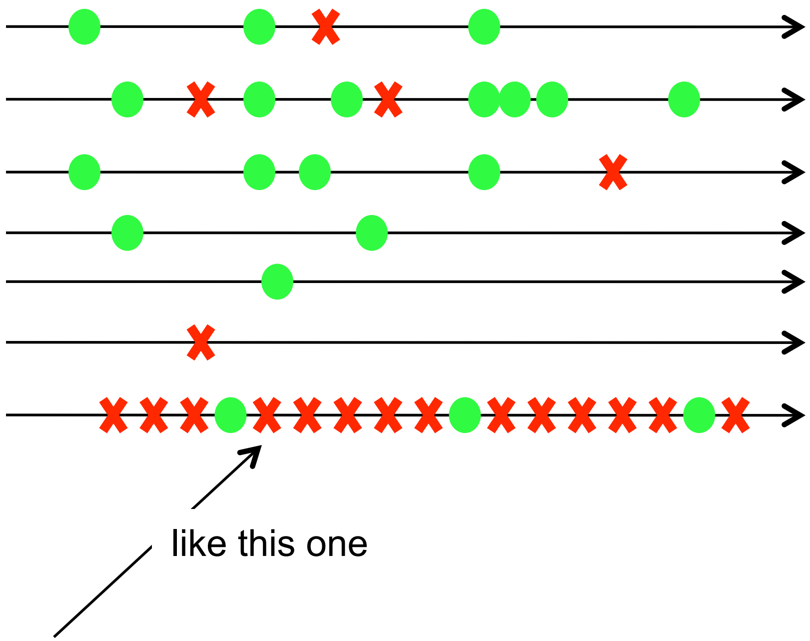
- Goal: estimate $\bar{a} \stackrel{\text{def}}{=} \mathrm{E}[a_i]$ given $\{n_i, m_i\}$
- The Ideal Estimator
  - is <u>unbiased</u>
  - has <u>minimum variance</u>
  - is <u>simple</u>: fast to compute, easy to implement (and easy to understand by practitioners!)
  - is <u>general</u>: not bound to a specific (class of) traffic distribution (n_i's)



EGR

- Empirical Global Ratio
  - aka Global Percentage

$$EGR = \frac{\#\ \text{of total failures}}{\#\ \text{of total requests}} = \frac{\sum_i m_i}{\sum_i n_i}$$



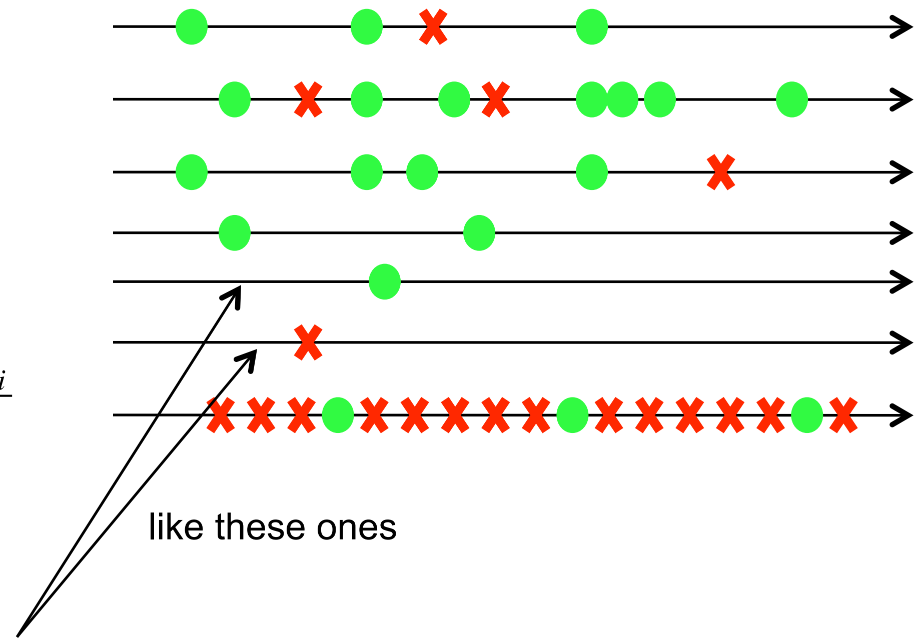
- Problem:
  - few “big” users with many failures
  - sporadically inflate the EGR, increase its variance
  - <u>*problem when traffic volume (the n_i’s) are heavy tailed !*</u>

EMR

- Empirical Mean Ratio
  - = Mean of individual ratios r_i

$$EMR = \frac{1}{I}\sum_i r_i = \frac{1}{I}\sum_i \frac{m_i}{n_i}$$

with $r_i = \frac{m_i}{n_i}$



- Problem:
  - “small” users with only one or few packets
  - bring very inaccurate estimates of their a_i’s
  - “weighting” as much as larger users with more reliable estimations
  - *problem when traffic volume (the n_i’s) are heavy tailed !*

NB: $r_i = m_i / n_i$ is the minimum variance unbiased estimator (MVUE) for a_i

EWR - definition

- Both EGR and EMR can be “corrected” by filtering away very big (for EGR) or very small (for EMR) users
  - discarding lots of data, especially for long-tailed n_i’s
- Is there a more clever alternative to data discarding ?
  - → data weighting !

- Empirical Weighted Ratio

$$EWR = \sum_i w_i \frac{m_i}{n_i} = \sum_i w_i r_i = \underline{w}^T \underline{r} \qquad \text{with} \quad \mathrm{w_i} > 0, \ \sum_i \mathrm{w_i} = 1$$

- Generalization of EGR, EMR
  - *w_i constant → $w_i = 1/I$ → EWR=EMR*
  - *w_i proportional to n_i → $w_i = n_i/N$ → EWR=EGR*

EWR - optimization

$$EWR = \underline{w}^T \underline{r} \qquad \text{with} \quad |\underline{\mathrm{w}}|_1 = 1$$

- Problem: Find the *optimal* weights w_i's
  - that minimize the variance of the estimator

- Methodology:
  - compute variance of estimator as function of weights var(***w***)
  - constrained optimization: minimize var(***w***) under constraint |***w***|***=1***
  - solve by Lagrangian multipliers

$$\dot{\mathbf{w}} = \arg\min_{\substack{\mathbf{w}>0 \\ \sum_i w_i = 1}} \mathrm{VAR}[S(\mathbf{w})]$$

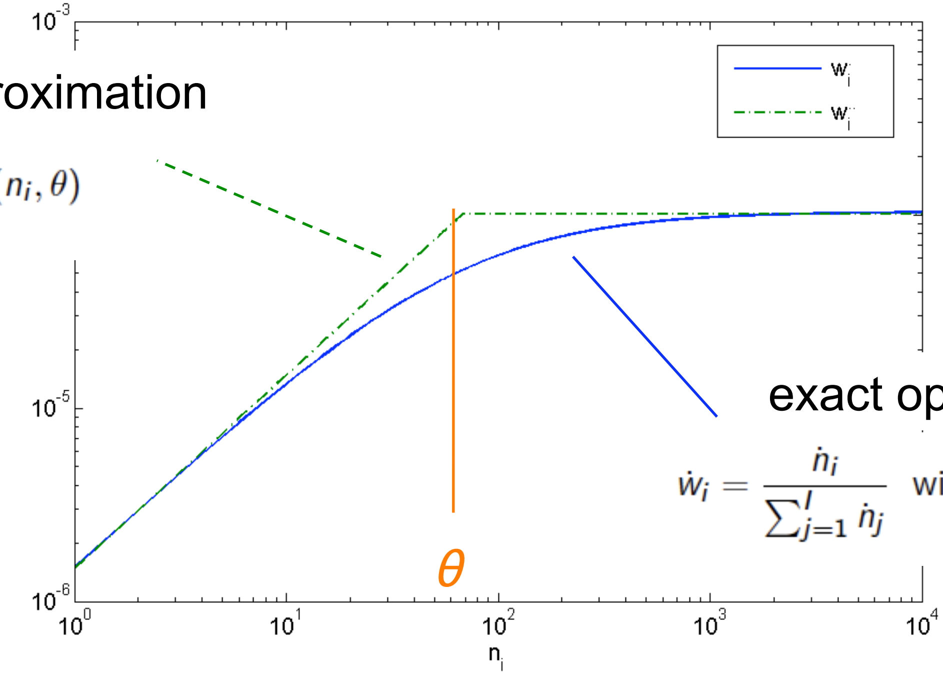
[BWA'09] A. Coluccia, F. Ricciato, Peter Romirer, On Robust Estimation of Network-wide Packet Loss in 3G Cellular Networks, IEEE BWA 2009, Honolulu, 30 November 2009

EWR - approximation

piece-wise linear approximation

$$\ddot{w}_i = \frac{\ddot{n}_i}{\sum_{j=1}^{I} \ddot{n}_j} \qquad \ddot{n}_i \overset{\text{def}}{=} \min(n_i, \theta)$$

$$\text{where } \dot{\theta} \overset{\text{def}}{=} \frac{\bar{a} - \sigma_a^2 - \bar{a}^2}{\sigma_a^2}$$



exact optimal solution

$$\dot{w}_i = \frac{\dot{n}_i}{\sum_{j=1}^{I} \dot{n}_j} \quad \text{with} \quad \dot{n}_i \overset{\text{def}}{=} \frac{1}{\sigma_a^2 + \frac{(\bar{a} - \sigma_a^2 - \bar{a}^2)}{n_i}}$$

- optimal knee-point θ depends on mean, variance of the a_i's
  - no dependency on the distribution of **n** → generality ☺
  - depends from moments of *p(a)*: → remains unknown ☹
- workaround: heuristic setting ("educated guess")
  - final estimator performance are weakly sensitive to exact location of knee-point, as far as "extreme" setting (very low, very high) is avoided

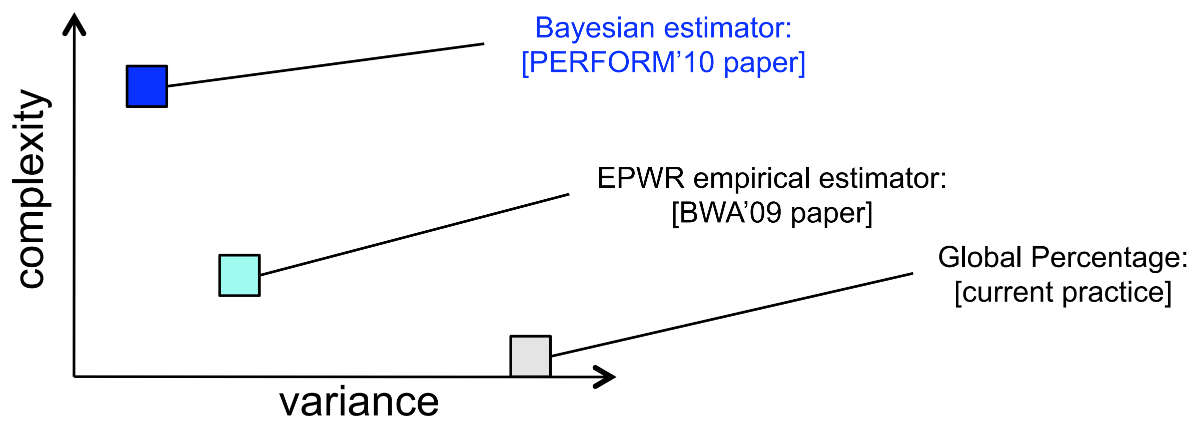
Bayesian Estimators

- Empirical Piece-wise Linearly Weighted Ratio (EPWR)
  - single parameter θ (to be set heuristically)
  - very simple conceptually and computationally
  - requires individual per-user counters

$$S(\mathbf{w}) \stackrel{\text{def}}{=} \sum_{i=1}^{I} w_i r_i = \mathbf{w}^T \mathbf{r}$$

$$\ddot{w}_i = \frac{\ddot{n}_i}{\sum_{j=1}^{I} \ddot{n}_j} \qquad \ddot{n}_i \stackrel{\text{def}}{=} \min(n_i, \theta)$$

- Why then going for Bayesian estimators ?
  - provides optimal reference for EPWR
  - preferred in applications where accuracy is critical, not simplicity



Bayesian approach

- Bayesian hierarchical model
  - $p(\mathbf{a}) \rightarrow \{a_i\} \rightarrow \{m_i\}$

$$\underbrace{p(a|m)}_{\text{posterior}} = \frac{p(m,a)}{p(m)} = \frac{p(m,a)}{\int p(m,a)\,\mathrm{d}a} \qquad \underbrace{p(m,a)}_{\text{joint}} = \underbrace{p(m|a)}_{\text{data}}\underbrace{p(a)}_{\text{prior}}.$$

- Approach: empirical parametric Bayes + conjugate prior
  - chose a parametric distribution family for *p(a)*
  - such that prior and posterior *p(a)*, *p(m|a)* belong to the same family (conjugate prior)
  - and estimate its (hyper-)parameters

Bayesian approach

Beta D.
posterior

$$p(a|m) = \frac{p(m,a)}{p(m)} = \frac{p(m,a)}{\int p(m,a)\,\mathrm{d}a}$$

Beta-Bin D. joint; Beta D. prior

$$p(m,a) = p(m|a)p(a).$$

data
Binomial D.

- Conjugate prior for Binomial → Beta Distribution
  - very versatile
  - two (hyper-)parameters α and β

$$p(a_i) = \frac{1}{\mathrm{B}(\alpha,\beta)} a_i^{\alpha-1}(1-a_i)^{\beta-1}$$

Beta function: $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$

Bayesian approach

- ML estimates of hyper-parameters:

$$\hat{\alpha}, \hat{\beta} \stackrel{\text{def}}{=} \arg\max_{\alpha,\beta} p(\boldsymbol{m}|\alpha, \beta)$$

- with p(m|α,β) = product of BetaBin(n_i,α,β)

$$p(\boldsymbol{m}|\alpha, \beta) = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{I} p(m_i, a_i|\alpha, \beta)\, \mathrm{d}a_i = \prod_{i=1}^{I} \binom{n_i}{m_i} \frac{\mathrm{B}(\alpha + m_i, \beta + n_i - m_i)}{\mathrm{B}(\alpha, \beta)}$$

- numerically solution straightforward
  - p(m|α,β) is unimodal and log-convex ☺
  - Well-known numerical procedures lead to accurate and fast resolution (e.g. simplex)
  - Example: resolution time with MATLAB for I=10^5 users: <10 sec (for EPWS a few msec)

Bayesian Estimators

- ok we got $\hat{\alpha}, \hat{\beta}$ but how to get $\bar{a}$?

- MAP $\hat{a}_i^{\text{MAP}} \stackrel{\text{def}}{=} \arg\max_{a_i} p(a_i | m_i, \hat{\alpha}, \hat{\beta}) = \frac{\hat{\alpha} + m_i - 1}{\hat{\alpha} + \hat{\beta} + n_i - 2}$
  - not good when small n_i's are present, as in real traffic

- Arithmetic mean of MMSE
  - MMSE of a_i coincides with conditional mean: $\hat{a}_i^{\text{MMSE}} \stackrel{\text{def}}{=} \mathrm{E}\left[a_i \middle| m_i, \hat{\alpha}, \hat{\beta}\right] = \frac{\hat{\alpha} + m_i}{\hat{\alpha} + \hat{\beta} + n_i}$

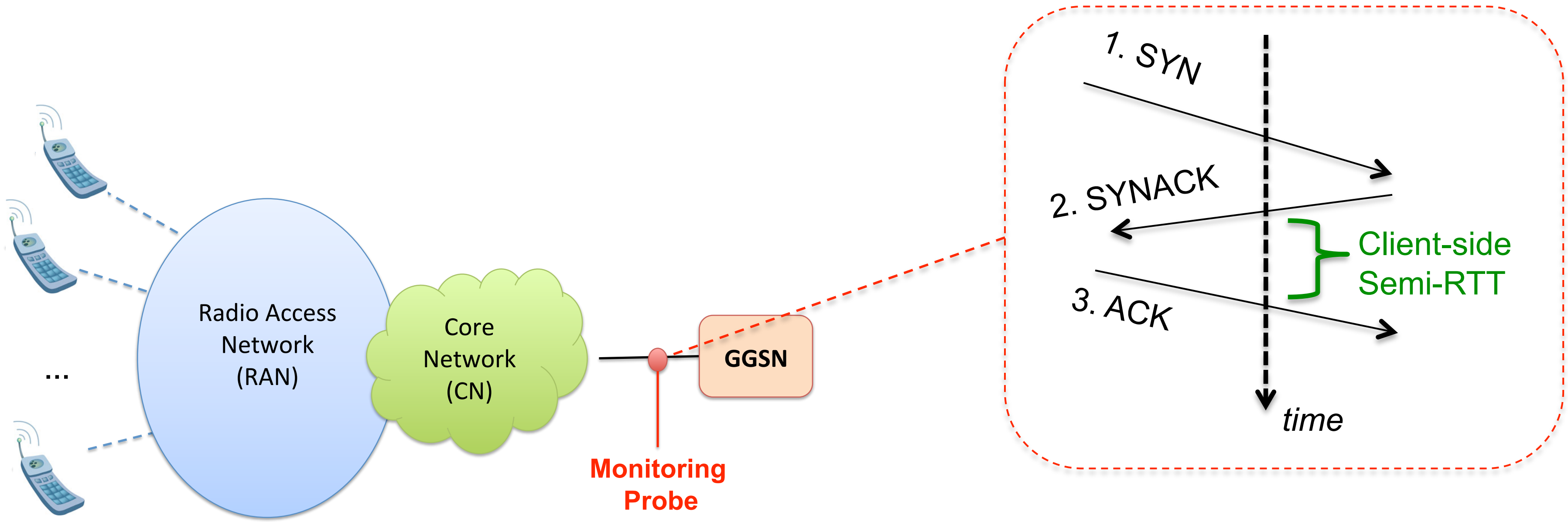
$$\boxed{Q_{\text{MMSE}} \stackrel{\text{def}}{=} \frac{1}{I} \sum_{i=1}^{I} \hat{a}_i^{\text{MMSE}} = \frac{1}{I} \sum_{i=1}^{I} \frac{\hat{\alpha} + m_i}{\hat{\alpha} + \hat{\beta} + n_i}.}$$

- Directly from the hyper-parameters

$$\boxed{Q_{\text{HYP}} \stackrel{\text{def}}{=} \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}}}$$

Results from a real dataset

- Datasets extracted with METAWIN, on a real UMTS/HSPA network
- DATA:INV
  - REQUEST := every SYNACK in DL | SUCCESS := unambiguous ACK in UL
- DATA:RTT
  - REQUEST := unambiguous SYNACK/ACK pair | SUCCESS := RTT < 500 ms



Results DATA:INV

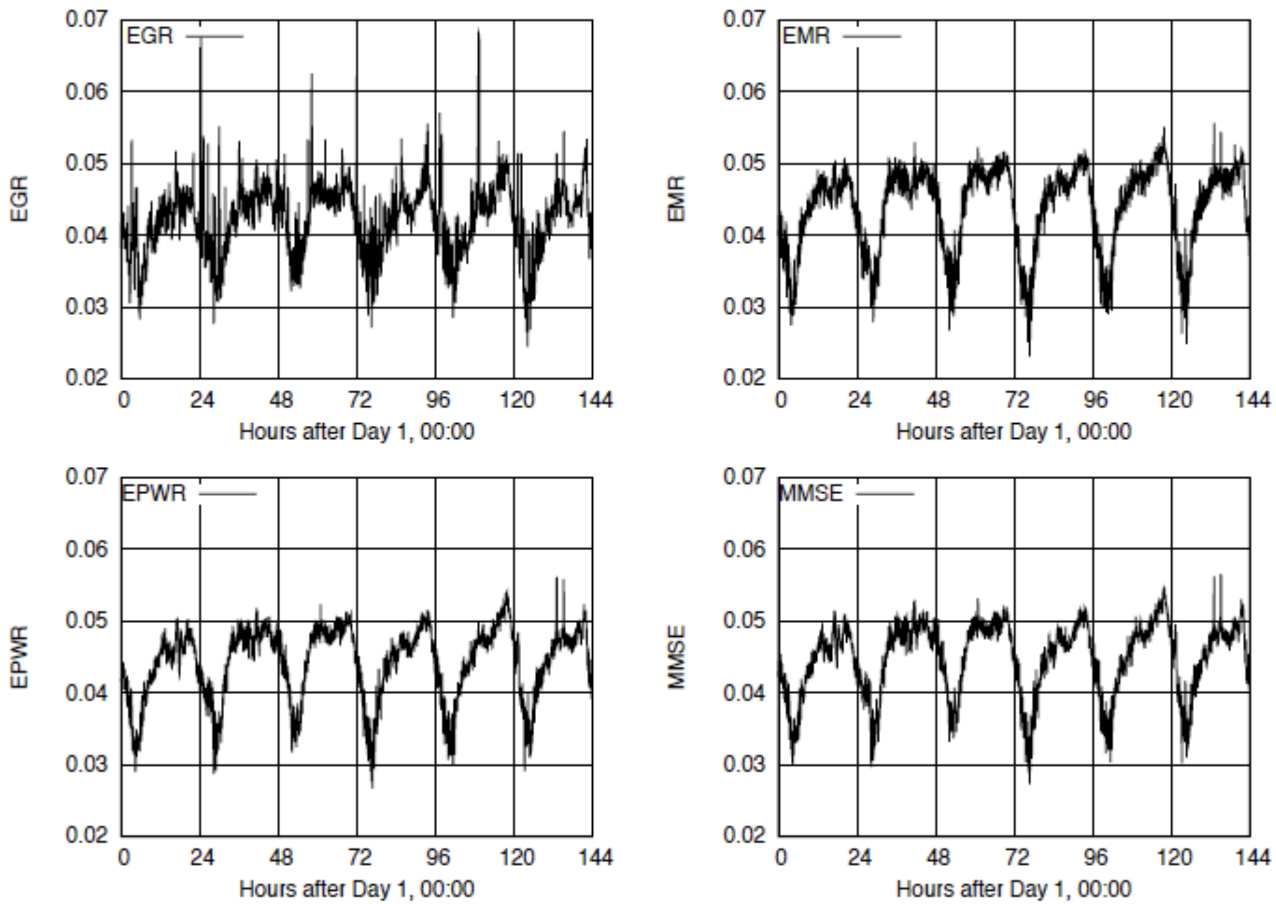


Fig. 2. Estimated mean failure probability for DATA:INV dataset (missing or ambiguous SYNACK/ACK associations).

Results DATA:RTT

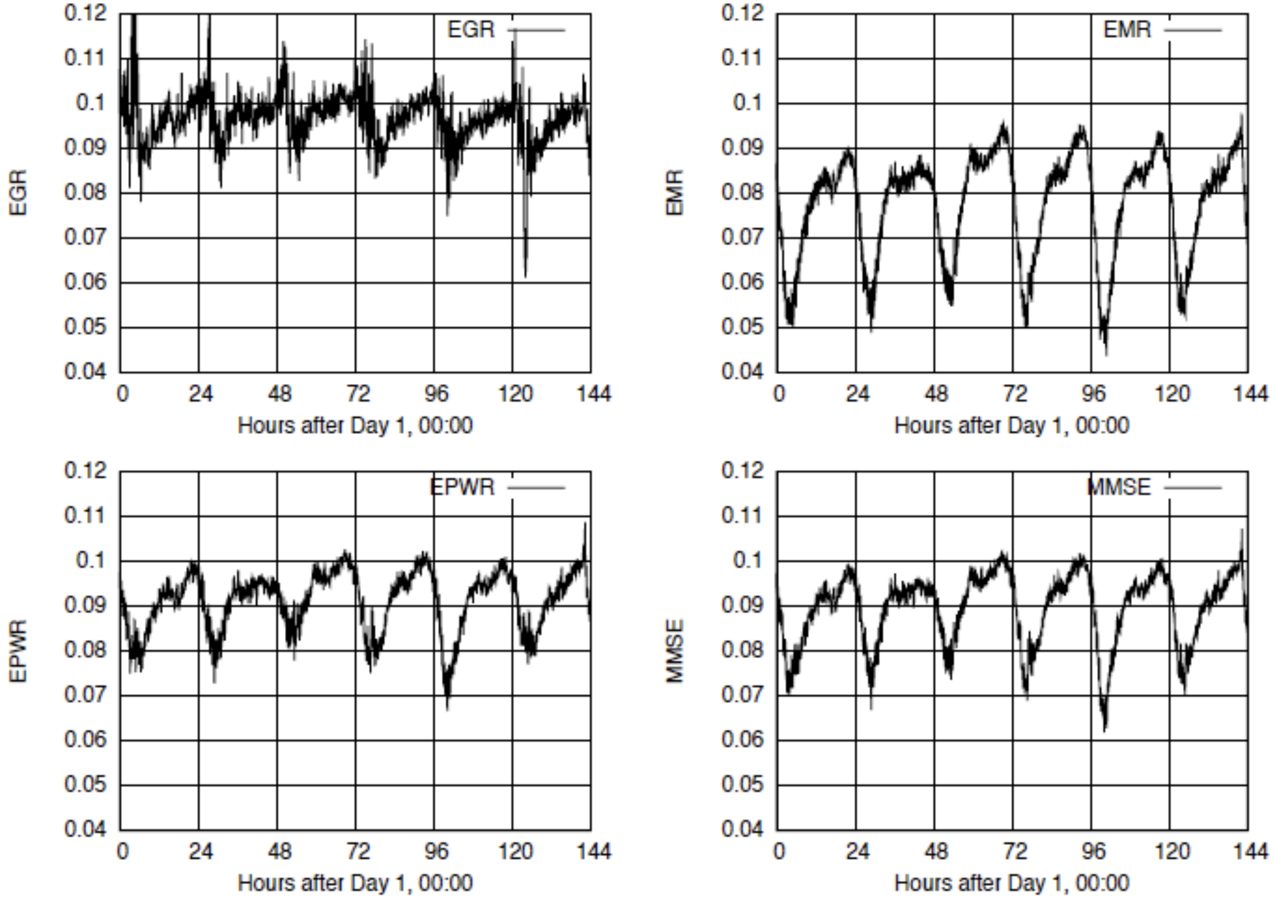


Fig. 3. Estimated mean failure probability for DATA:RTT dataset (unambiguous SYNACK/ACK pairs with semi-RTT exceeding 500 ms).

Summary of Results

- for these datasets…
- Bayesian Estimators Q_{MMSE} and Q_{HYP} almost equivalent
- EGR is a poor choice: high noise, bias
- EPWR very close to Q_{MMSE} (but much simpler)
- EMR not bad, only slightly worse than EPWR (but same complexity)

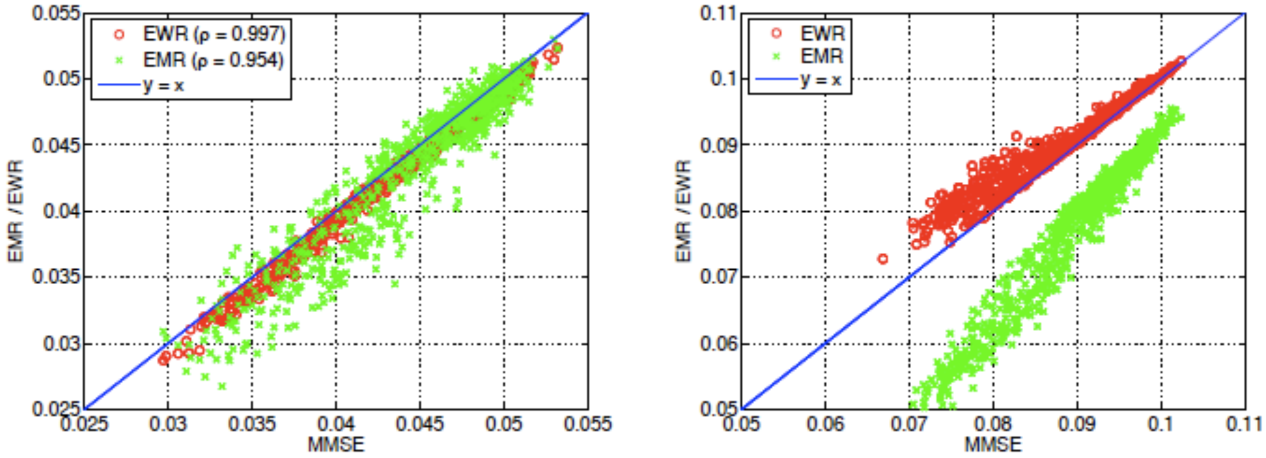


Fig. 4. Scatterplots of S_{EMR} and S_{EPWR} versus Q_{MMSE} for DATA:INV (left) and DATA:RTT (right).

Conclusions

- Mean failure estimation: a very general problem
  - apply to any binary outcome event
  - generated by a user population
- Default practice: Global Percentages
  - ok if traffic distribution across users has low-variance
  - problems with heavy-tailed distributions

- Bayesian estimator
  - provides the optimal reference
  - solvable in reasonable time also for large dataset (seconds)
- EPWR estimator
  - extremely simple (computationally and conceptually)
  - near-optimal performance
  - one parameter to be set heuristically
  - → could become a standard KPI !

- Thanks for your attention !