On Robust Estimation of Network-wide Packet Loss in 3G Cellular Networks

Angelo Coluccia[†], Fabio Ricciato[†,‡], Peter Romirer-Maierhofer[‡]

[†] *University of Salento, Lecce (Italy)*

[‡] *Forschungszentrum Telekommunikation Wien (FTW), Vienna (Austria)*

***Abstract*—In this work we address the problem of estimating the network-wide packet loss rate across the radio access section of a 3G cellular network. The reference scenario consists of a passive monitoring probe located in the Core Network. The probe counts the number of TCP packets directed to each individual mobile terminal and, from the analysis of (un)acknowledged packets, infers the loss ratio for each individual terminal. The problem is then to derive a synthetic indicator representative of the network-wide packet loss, which can be used to detect large-scale performance drifts and network anomalies. We show that common simplistic indicators like the total rate of lost packets (across all terminals) and the average per-terminal loss rate do not work well in the general case. The key problem is the large disparity of traffic volume across individual terminals. In this contribution we formulate the problem in terms of optimal statistical inference and provide a set of robust near-optimum estimators that are relatively simple to implement. We validate the proposed estimators with simulations in synthetic scenarios and provide results from a real operational 3G network.**

I. Introduction

Complex network infrastructures like third-generation (3G) cellular networks are exposed to errors and attacks. Passive packet-level monitoring can play an important role in supporting the operation and troubleshooting of such systems. Summary statistics of network-layer performance variables (delay, RTT, loss, throughput etc.) can be extracted from passive probes and used as synthetic indicators for the network health status. Sudden deviations in the otherwise regular temporal behavior of such indicators might point to network problems. Change-point detection methods can be used to automatically identify anomalies and raise timely alarms.

In this work we focus on the use of network-wide packet loss statistics obtained from the passive analysis of TCP packets at a single measurement point. The reference scenario consists of a passive monitoring probe located in the Core Network (ref. Fig. 1). The probe counts the number of TCP packets directed to each individual mobile terminal and, from the analysis of (un)acknowledged packets, infers the number of packet loss events. The problem is then to derive a synthetic indicator representative of the network-wide packet loss. The ultimate goal is to use such indicator to detect network problems and large-scale performance drifts.

A common simple approach is to use as the summary indicator the ratio of loss events over the total transmitted packets across all the terminals. Another possible method is to compute the individual per-terminal loss ratios and take the arithmetic average. We show that such simplistic approaches do not always work well in practical scenarios. The key issue is the large disparity found in the distribution of the number of packets per terminal, which is due to the typically large disparity in traffic intensity of individual users: many users generate very low traffic, while a few ones generate very high volumes. For the two simplistic summary indicators introduced above, such disparity translates directly into an increased variance of the estimate. In this work we formalize the problem at hand as an optimal estimation problem. The optimal solution can be found via Bayesian approach [9], but at the cost of a relatively high resolution complexity which might hamper the practical adoption of such schemes. In this contribution we focus on the derivation of simple yet robust near-optimum estimators.

II. Related works

A number of approaches, including Maximum-Likelihood (ML) and Expectation Maximization (EM), have been proposed to estimate the loss rates from passive measurements (e.g., [1]). Caceres *et al.* pioneered the ML method for link loss rates by multicast probing packets [2] [3]. Many existing methods attempt to infer the link loss rates of individual links based on the analysis of the whole network tree (passive tomography). Su *et al.* [4] proposed a low complexity inference approach by dividing the original tree into several sub-trees with minimum depth. This allows to infer the link loss rates in a much simpler way than the original tree. Padmanabhan *et al.* [5] used Bayesian Inference and Gibbs Sampling to identify lossy links from passive end-to-end client-server traffic. With the growth of networks scale, such methods become computationally complex. A more general and somehow simpler technique has also been proposed by Tian *et al.* [6]. The key idea is to infer link loss rates and multicast topology simultaneously.

With respect to such previous works, we tackle the problem from a different angle. We are interested only in obtaining a global synthetic indicator of network-wide loss, rather than measure accurately the loss of individual links. Also, we aim at simplicity as a key requirement, to encourage practical adoption in operational networks. To the best of our knowledge no previous work has addressed the issue of evaluating synthetically the average network-wide loss, and no robust estimator is available in the literature for the problem at hand.

978-1-4244-5626-0/09/$26.00 ©2009 IEEE

III. Problem formulation

A. System model

For a generic measurement time bin (e.g. 1 minute), let I denote the total number of *active* terminals — for which at least one packet was observed in the downlink direction at the monitoring point in the specific time bin. For every terminal i ($i = 1, 2, \dots, I$), we introduce the following variables:

- n_i the total number of downlink packets to terminal i seen at the monitoring point ($n_i \geq 1$);
- m_i the number of loss events for terminal i, i.e. the number of packets losses occurred along the path between the monitoring point and the terminal ($0 \leq m_i \leq n_i$);
- $r_i = \frac{m_i}{n_i}$ the empirical loss ratio for terminal i;
- a_i the underlying (unknown) loss probability of the path to terminal i ($a_i \in [0, 1]$).

Finally we denote by N the total number of packets across all terminals, i.e. $N \overset{\text{def}}{=} \sum_i n_i$. To simplify the notation we will occasionally use the following vectorial notation : $\boldsymbol{n} \overset{\text{def}}{=} [n_1 n_2 \cdots n_I]^T$, $\boldsymbol{m} \overset{\text{def}}{=} [m_1 m_2 \cdots m_I]^T$, $\boldsymbol{r} \overset{\text{def}}{=} [r_1 r_2 \cdots r_I]^T$ and $\boldsymbol{a} \overset{\text{def}}{=} [a_1 a_2 \cdots a_I]^T$. For the sake of mathematical treatment we assume independence between losses: each packet directed to terminal i is lost with probability a_i independently from other packets. Therefore the generic variable m_i is the sum of n_i trials with success probability a_i, and all m_i's are independent Binomial random variables:

$$m_i \sim \mathcal{B}(n_i, a_i) \Longrightarrow \begin{cases} \mathrm{E}[m_i] = n_i a_i \\ \mathrm{VAR}[m_i] = n_i a_i (1 - a_i). \end{cases} \qquad (1)$$

In practice the estimation of loss events at a single monitoring point can be done for TCP traffic based on the analysis of packet/acknowledgment pairs: SYN/SYNACK, SYNACK/ACK and DATA/ACK. Unacknowledged packets, i.e. incomplete pairs, signal a loss event, either of the front packet or of its associated ACK. The technicalities involved in the passive measurement of TCP losses, including the sources of possible measurements errors, are left outside the focus of this paper — refer e.g. to [7] for more material on this topic. Here we assume that the vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ have been measured and serve as input for the problem at hand.

B. Goal definition

A central component of the model is that the (unknown) loss probabilities a_i's are considered as i.i.d. random variables generated from a common underlying distribution with mean value $\bar{a} \overset{\text{def}}{=} \mathrm{E}[a_i]$. Therefore it is natural to take $\bar{a}$ as the summary indicator representative of "network-wide mean loss". In other words, we are interested in estimating the mean value $\bar{a}$ out of the measured vectors $\boldsymbol{n}$ and $\boldsymbol{m}$.

The goodness of an estimator can be evaluated against the following criteria:

- **Optimality**: We consider only unbiased estimators. Therefore the optimality criterion can be expressed in terms of minimum variance: the "optimal" estimator is the one that minimizes the uncertainty (variance) around the estimated value. Lower variance allows for better discrimination of change-points, i.e. deviations of the underlying mean loss $\bar{a}$, from statistical fluctuations.

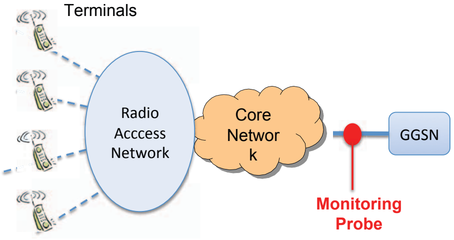

Fig. 1. Reference measurement scenario.

- **Generality**: The central difficulty of the estimation at hand lies in the fact that the vector $\boldsymbol{n}$ is given and can not be controlled. Typically the individual traffic volume — hence the set of n_i's — is distributed very unevenly across the terminals. Moreover, the traffic distribution can change across time, following the daily/weekly changes in the traffic mix and in user activity, not to mention long term trends in user behavior [8, §VI-A]. Therefore we are interested in obtaining a robust estimation procedure that does not rely on any specific assumption about the distribution of the n_i's, i.e. can perform well for any distribution, and does not require manual re-tuning when the traffic distribution changes.
- **Simplicity**: in many practical cases, implementation simplicity is a key requirement, especially when real-time computation is required. More in general, simple estimation methods are much more likely to be implemented and adopted in practical systems.

In the present contribution we focus on the derivation of a "best compromise" solution among these criteria. In particular, we are set to trade-off optimality for simplicity: we seek estimators that are reasonably simple to implement and that perform near-optimum in a wide range of operational conditions, i.e. for arbitrary distributions of n_i's.

IV. Problem Analysis

Since $\boldsymbol{a}$ is regarded as a random vector, the problem can be attacked with Bayesian inference techniques. Such techniques are powerful but often also pretty complex, and require *a priori* information on $p(\boldsymbol{a})$. In a parallel work we are investigating different Bayesian methods for the problem at hand. Preliminary results with Bayesian hierarchical models [9] show that one can obtain an optimal estimator that attains the theoretical bound, but with a relatively involved numerical computation. This is in contrast with the goal of attaining an estimator that is simple to implement — a key requirement for practical adoption in online monitoring systems. Therefore we leave the treatment of Bayesian techniques out of this work and instead focus on more handy sub-optimal approaches.

A. Empirical Mean Ratio (EMR)

The ratio r_i is an unbiased estimator of the mean loss probability $\bar{a}$, formally:

$$\mathrm{E}[r_i] = \mathrm{E}\big[\mathrm{E}[r_i | a_i]\big] = \mathrm{E}[a_i] = \bar{a}. \qquad (2)$$

It can be easily seen that r_i is the Maximum Likelihood (ML) estimate of a_i — taken as an unknown parameter — and also the Minimum Variance Unbiased Estimator (MVUE), i.e. it attains the Cramer-Rao lower bound for the estimation of a_i. An intuitive summary indicator for the network-wide loss can be obtained as the arithmetic mean of the ratios r_i's:

$$S_M \overset{\text{def}}{=} \frac{1}{I}\sum_{i=1}^{I} r_i = \frac{1}{I}\sum_{i=1}^{I}\frac{m_i}{n_i}. \tag{3}$$

We refer to such indicator as the *Empirical Mean Ratio* (EMR). Despite the optimality properties of the r_i's as estimators for the individual loss probabilities a_i's, the EMR is not always the best choice for the estimation of the underlying mean probability $\bar{a}$. The problem lies in the fact that the variance of r_i — considered as the best available estimate for a_i — is inversely proportional to the number of packets n_i, i.e. $\mathrm{VAR}[r_i] = a_i(1-a_i)/n_i$. Therefore the (usually large) variability of n_i's maps into a large variability of the uncertainty (variance) of the individual estimates — a case of heteroscedasticity. In the simple arithmetic mean, the more accurate estimates (for large n_i) are weighted in the same way as very poor ones (for very small n_i), hence if the number of thin-traffic terminals is considerable, the resulting overall variance is dominated by the latter ones.

One possible "correction" to the EMR is to ignore (discard) all the observations associated to terminals with n_i below a threshold γ, and consider only the subset of measurement points with $n_i \geq \gamma$ (e.g. $\gamma = 10$). We refer to such estimator as $S_M|_\gamma$. The drawback of such strategy is that the value of γ needs to be manually tuned in some way. Another weakness of $S_M|_\gamma$ is that a certain amount of information available in the data is simply discarded, which is clearly a grossly suboptimal approach to the problem. We will discuss the performance of $S_M|_\gamma$ with reference to some numerical results in Sec. VI-B.

B. Empirical Global Ratio (EGR)

Another simple indicator often used in practice is obtained as the total share of lost packets across the whole network:

$$S_G \overset{\text{def}}{=} \frac{\sum_{i=1}^{I} m_i}{\sum_{i=1}^{I} n_i} = \frac{\sum_{i=1}^{I} m_i}{N}. \tag{4}$$

We refer to such estimator as the *Empirical Global Ratio* (EGR). Contrary to S_M, which suffered from the presence of many terminals with thin traffic (small n_i), the problem with S_G is the presence of a few terminals with very high traffic (large n_i) which can dominate the overall estimator. This aspect is discussed later in Sec. VI-B with reference to some illustrative numerical results.

V. Weighted estimators

A. Definition of Empirical Weighted Ratio (EWR)

The simple estimators EMR and EGR seen above can be considered as particular cases of a wider class of estimators obtained as the weighted average of the individual empirical loss ratios r_i, formally:

$$S(\boldsymbol{w}) \overset{\text{def}}{=} \sum_{i=1}^{I} w_i r_i = \boldsymbol{w}^T \boldsymbol{r} \tag{5}$$

where the weight vector $\boldsymbol{w} \overset{\text{def}}{=} [w_1 w_2 \cdots w_I]^T$ has non-negative components $w_i \geq 0 \ \forall i$ and $\sum_{i=1}^{I} w_i = 1$ — the latter condition is justified below. We will refer to such class of estimators as *Empirical Weighted Ratio* (EWR).

When all weights are equal, $w_i = 1/I \ \forall i$, the EWR leads directly to the EMR, formally:

$$S(\boldsymbol{w})|_{w_i=\frac{1}{I}} = \frac{1}{I}\sum_{i=1}^{I} r_i = S_M. \tag{6}$$

Conversely, when the weights are proportional to the number of packets for each terminal, $w_i = n_i/N \ \forall i$, the EWR coincides with the EGR, i.e.

$$S(\boldsymbol{w})|_{w_i=\frac{n_i}{N}} = \frac{1}{N}\sum_{i=1}^{I} n_i r_i = \frac{1}{N}\sum_{i=1}^{I} m_i = S_G. \tag{7}$$

B. Derivation of the optimal weighted estimator

Given the EWR estimation structure in eq. (5), we are interested in finding the optimal weight vector which minimizes the variance of the estimate. Recalling that the a_i's are taken as i.i.d. random variables with mean $\bar{a}$ and variance σ_a^2, we first require that an estimator of the form (5) is unbiased:

$$\mathrm{E}[S(\boldsymbol{w})] = \bar{a}, \tag{8}$$

which develops as:

$$\mathrm{E}\left[\sum_{i=1}^{I} w_i r_i\right] = \sum_{i=1}^{I} w_i \mathrm{E}\big[\mathrm{E}[r_i|a_i]\big] = \sum_{i=1}^{I} w_i \mathrm{E}[a_i] = \bar{a}\sum_{i=1}^{I} w_i \tag{9}$$

which leads to the constraint:

$$\sum_{i=1}^{I} w_i = 1. \tag{10}$$

The variance of the estimator $S(\boldsymbol{w})$ is derived by resorting to the law of total variance, as follows:

$$\begin{aligned}\mathrm{VAR}[S(\boldsymbol{w})] &= \mathrm{VAR}\big[\mathrm{E}[S(\boldsymbol{w})|a_i]\big] + \mathrm{E}\big[\mathrm{VAR}[S(\boldsymbol{w})|a_i]\big] \\ &= \mathrm{VAR}\left[\sum_{i=1}^{I} w_i a_i\right] + \mathrm{E}\left[\sum_{i=1}^{I} w_i^2 \frac{a_i(1-a_i)}{n_i}\right] \\ &= \sigma_a^2 \sum_{i=1}^{I} w_i^2 + (\bar{a} - \sigma_a^2 - \bar{a}^2)\sum_{i=1}^{I}\frac{w_i^2}{n_i}.\end{aligned} \tag{11}$$

The problem is to find the weight vector $\dot{\boldsymbol{w}}$ that minimizes the variance in eq. (11) subject to constraint (10), formally:

$$\dot{\boldsymbol{w}} = \underset{\substack{\boldsymbol{w}>0 \\ \sum_i w_i = 1}}{\arg\min}\, \mathrm{VAR}[S(\boldsymbol{w})]. \tag{12}$$

The problem can be solved by Lagrange multipliers, yielding:

$$\dot{w}_i = \frac{\dot{n}_i}{\sum_{j=1}^{I} \dot{n}_j} \quad \text{with} \quad \dot{n}_i \overset{\text{def}}{=} \frac{1}{\sigma_a^2 + \frac{(\bar{a} - \sigma_a^2 - \bar{a}^2)}{n_i}}. \tag{13}$$

We denote by $S_C \overset{\text{def}}{=} S(\dot{w})$ the EWR estimator obtained with the optimal weight vector given by eq. (13). This solution can not be used in practice because the values of $\bar{a}$ and σ_a^2 are unknown — $\bar{a}$ is the desired output of the estimation

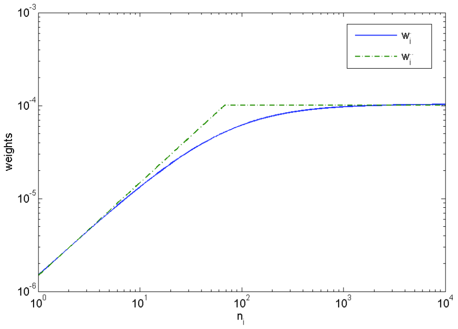

Fig. 2. Plot of the optimal weights $\dot{w}_i$'s and the approximated piecewise linear weights $\ddot{w}_i$. Here the optimal knee-point is $\dot{\theta} \approx 68$.

procedure, not its input! Remarkably, preliminary results from a parallel work show that S_C attains the same performances of significantly more complex Bayesian estimation procedures. Hereafter S_C will provide a reference theoretical bound to assess the goodness of other estimators.

A graphical representation of the optimal weight vector $\dot{\boldsymbol{w}}$ is plotted in Fig. 2. We notice that $\dot{n}_i$ in eq. (13) can be approximated as follows:

$$\dot{n}_i \approx \begin{cases} \frac{n_i}{\bar{a}-\sigma_a^2-\bar{a}^2} & n_i \ll \dot{\theta} \\ \frac{1}{\sigma_a^2} & n_i \gg \dot{\theta} \end{cases} \tag{14}$$

where $\dot{\theta} \stackrel{\text{def}}{=} \frac{\bar{a}-\sigma_a^2-\bar{a}^2}{\sigma_a^2}$. In other words, for $n_i \ll \dot{\theta}$ the weights tend to be proportional to n_i — as in the EGR (ref. eq. (7)) — while for $n_i \gg \dot{\theta}$ the weights tend to be constant — as in the EMR (ref. eq. (6)). Therefore the optimal EWR estimator can be considered as a mixture of the two basic estimators EMR and EGR, which dominate respectively the upper and lower range of the n_i's distribution, with a cross-over point in $\dot{\theta}$. A key point to be remarked from eq. (14) is that the knee-point $\dot{\theta}$ depends only on (the first two moments of) the distribution of loss probabilities a_i's and not on the actual values of n_i's.

C. Piecewise linear approximation

The approximation in eq. (14) is very good one decade before and after the knee-point $\dot{\theta}$ (ref. to Fig. 2). We can therefore approximate the optimal weight vector $\dot{w}_i$ with the following piecewise linear (near-optimum) vector $\ddot{w}_i$:

$$\dot{w}_i \approx \frac{\ddot{n}_i}{\sum_{j=1}^{I} \ddot{n}_j} \stackrel{\text{def}}{=} \ddot{w}_i \quad \text{with} \quad \ddot{n}_i \stackrel{\text{def}}{=} \min(n_i, \dot{\theta}). \tag{15}$$

We denote by $S_L \stackrel{\text{def}}{=} S(\ddot{w})$ the EWR estimator obtained with the piecewise linear weight vector given by eq. (15). In the simulations we found that S_L performs extremely close to the theoretical optimum S_C in all the considered scenarios (see Sec. VI-B). Similarly to S_C, S_L can not be computed in practice since $\dot{\theta}$ depends on the unknown parameters $\bar{a}$ and σ_a^2, but it serves as the basis to build practical approximated estimators. The advantage of adopting a piecewise linear approximation is simplicity: S_L has only one parameter to set ($\dot{\theta}$) rather than two ($\bar{a}, \sigma_a^2$), and also the numerical computation of the weights is slightly faster.

D. Approximate near-optimum estimators

Since the real values of $\bar{a}$ and σ_a^2 are unknown, the optimum weights can not be calculated exactly but only approximated. One possibility is to resort to some initial (coarse) estimates of $\bar{a}$ and σ_a^2 for computing the weights from eq. (13) or eq. (15). This might appear as a critical step, especially because the estimation of the variance is an hard task (see e.g. [10]) which contrasts with the goal of achieving a simple estimation solution. However, it turns out from the simulations that the final variance of the EWR estimator is relatively insensitive to the exact location of the knee-point. In other words, relatively large deviations from the optimal value $\dot{\theta}$ do not cause large differences in the final performance of the estimator as far as the knee-point stay within a reasonable range. This point is discussed later in Sec. VI-C.

The above finding motivates the use of "reasonably inaccurate" estimates of $\bar{a}$ and σ_a^2 to locate the knee-point. A simple and natural solution is to select only the more reliable r_i's for the initial estimation of mean and variance of the a_i's, discarding the most "noisy" observations. For example, one can select the subset of observations with $n_i \geq \gamma$, as already done for introducing $S_M|_\gamma$. Therefore $S_M|_\gamma \stackrel{\text{def}}{=} \widehat{\bar{a}}|_\gamma$ can be used as initial estimate for $\bar{a}$, while the *sample variance* calculated on the selected subset — denoted by $\widehat{\sigma}_a^2|_\gamma$ — is taken as a rough estimate for σ_a^2. By replacing the unknown parameters $(\bar{a}, \sigma_a^2)$ in eq. (13) with such initial estimates $(\widehat{\bar{a}}|_\gamma, \widehat{\sigma}_a^2|_\gamma)$, we obtain a new weight vector $\widehat{\dot{w}}|_\gamma$. Similarly, they can be used in eq. (15) to derive an estimate of the piecewise linear vector $\widehat{\ddot{w}}|_\gamma$ after computing the estimate of the knee-point $\widehat{\dot{\theta}}|_\gamma$. From these vectors we finally obtain the estimators $S_{C,\gamma} \stackrel{\text{def}}{=} S(\widehat{\dot{w}}|_\gamma)$ and $S_{L,\gamma} \stackrel{\text{def}}{=} S(\widehat{\ddot{w}}|_\gamma)$.

An alternative (simpler) strategy for implementing S_L is to skip the initial estimation phase and to set directly the position of the knee-point at a heuristically chosen value θ. Any external information about the phenomenon at hand — e.g. the order of expected level and degree of variability of the individual loss probabilities a_i's — would help in taking an "educated guess" about a reasonable location of the optimal knee-point $\dot{\theta}$. We denote the estimator built in this way as $S_L(\theta)$. Although at a first glance it might appear as a gross simplification of the problem, such heuristic approach is motivated by the fact that the sensitivity of $S_L(\theta)$ to the location of the knee-point is surprisingly contained, provided that the setting is "reasonable" (ref. Secs. VI-B and VI-C). Besides the simulation results, our preliminary trials with measurements from a real network suggest that such simple approach might still attain reasonably good performance in practical scenarios (ref. Sec. VI-D).

VI. Numerical results

A. Requirements on the simulation scenario

The performances of the the proposed estimators are evaluated here by MATLAB simulations. The goal is to investigate the dependency between the packet distribution $\boldsymbol{n}$ and the estimation variance: a good estimator would yield near-minimum variance (optimality) for any choice of $\boldsymbol{n}$ (generality).

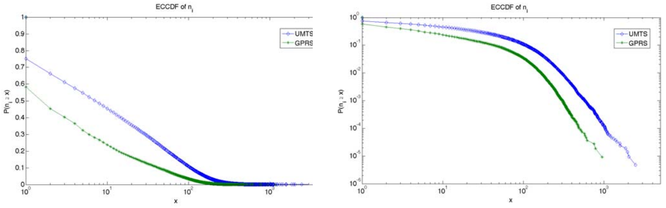

Fig. 3. Distribution of downlink SYNACK to terminal in 5 min bins, GPRS and UMTS (left: linear scale; right: logarithmic)

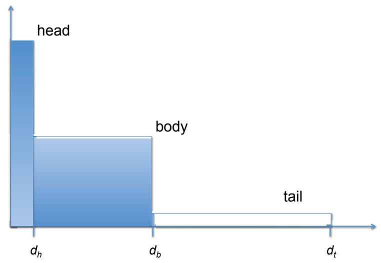

Fig. 4. Uniform-mixture distribution used in simulations for $\boldsymbol{n}$.

The choice of vector $\boldsymbol{n}$ is a key point in the design of the simulations. On one hand, it is desirable to use vectors that are representative of the real packet distributions found in real-world scenarios. On the other hand, in order to assess the generality of the proposed scheme — also for the prospective adoption in other different application contexts — it is important to explore the range of possible distributions at wide, including extreme "worst-case" scenarios. In order to address both such requirements, a possible strategy is to use a parametrized family of synthetic distributions that (i) contains the same ingredients of real distributions and (ii) whose parameters can be varied to explore a sufficiently broad set of synthetic scenarios, also outside the range of what is "typical" in a particular application context.

As a preliminary step towards the design of a suitable family of synthetic distributions, it is instructive to look at some real data set. In Fig. 3 we plot the Empirical Complementary Cumulative Distribution Function (ECCDF) for the number of TCP handshake packets seen in downlink (SYNACK) in a real operational 3G network, counted in time bins of 5 min, separately for GPRS/EDGE and UMTS/HSDPA terminals. The data are obtained with the METAWIN monitoring system [11] from the operational network of a major mobile operator in Austria. The same data were used for the analysis of Round-Trip-Times for TCP handshaking packets in a previous work [7]. Three different components are found in such empirical distributions — head, body and tail — associated to three different user segments (ref. Fig. 3):

- Thin users: a considerable fraction of terminals open only one or a few connections within the monitoring interval, therefore n_i is close to 1. This is the "head" of the empirical distribution.
- Intermediate users: the central part ("body") of the distribution is made of terminals generating an intermediate number of connections, n_i takes values up to several tens.
- Fat users: the "tail" of the distribution consists of a minority of heavy users that generate a huge amount of connections, with n_i ranging to several hundreds or even thousands in this case (see logscale plot in Fig. 3).

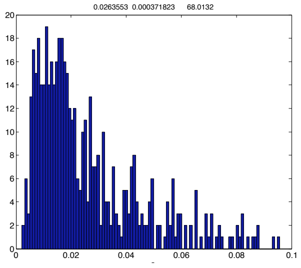

Fig. 5. Empirical histogram of a sample set of a_i's used in simulations.

	ϕ_1	ϕ_2	ϕ_3	ψ_1	ψ_2	ψ_3
Fig. 6	$[0, 0.9]$	$1 - \phi_1 - \phi_3$	0.05	8	100	3000
Fig. 7	0.3	$1 - \phi_1 - \phi_3$	0.05	8	100	$[500, 5000]$
Fig. 8	0.3	$1 - \phi_1 - \phi_3$	0.05	8	100	3000

TABLE I
PARAMETERS USED IN THE SIMULATIONS FOR THE DISTRIBUTION OF $\boldsymbol{n}$.

One simple approach is to model each component separately. Users can be assigned to one of the three classes according to some fixed shares — denote by ϕ_1, ϕ_2, ϕ_3 for head, body and tail respectively, with $\phi_1 + \phi_2 + \phi_3 = 1$. The problem is then how to model each class. The Gamma, Weibull and Lognormal distributions have often been used in the past literature to model traffic volumes and rates, at both packet and connection level, at least for the body and tail. The use of truncated Pareto distributions is also quite common to model the tail behavior. The head can be easily modeled with a highly concentrated distribution located at low values, or even as a fixed share of users with $n_i = 1$.

One possible investigation strategy is to test with different combinations of such well-known distributions — or better, their discrete variants or approximations — and for each combination explore a reasonably wide range of the parameter space. An alternative simpler approach is to adopt the same distribution type for all the three classes, with different parameters. The simplest and most "neutral" choice is obviously the (discrete) Uniform distribution $U(x, y)$, where any integer value in the interval $[x, y]$ has the same probability. Notably, the Uniform distribution is the one which maximizes the entropy for any limited range of values, therefore this choice can be considered a sort of "worst case" with respect to statistical dispersion. Following this approach, one can model each component with a discrete Uniform distribution with different ranges. We denote by ψ_1, ψ_2, ψ_3 the upper range boundary for head, body and tail respectively, with $\psi_1 \ll \psi_2 \ll \psi_3$. Together with the per-class shares, we end up with six parameters in total $\{\phi_k, \psi_k, \ k = 1, 2, 3\}$. A schematic representation of the resulting Uniform-mixture distribution is sketched in Fig. 4. The default values of the simulation parameters are given in Table I.

The generic terminal loss rate is obtained as $a_i = 10^v$, where v is a random variable extracted from a truncated normal distribution defined in the range $[-3, -1]$ with parameters $(\mu_v, \sigma_v^2) = (-1.67, 0.35)$. In this way we obtain a distribution that is conceptually similar to a lognormal but constrained within a finite range, i.e. $a_i \in [10^{-3}, 10^{-1}]$. It results that

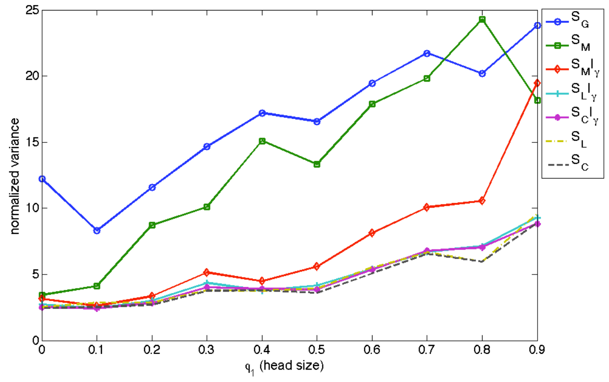

Fig. 6. Simulation varying the head size. Parameters are given in Table I.

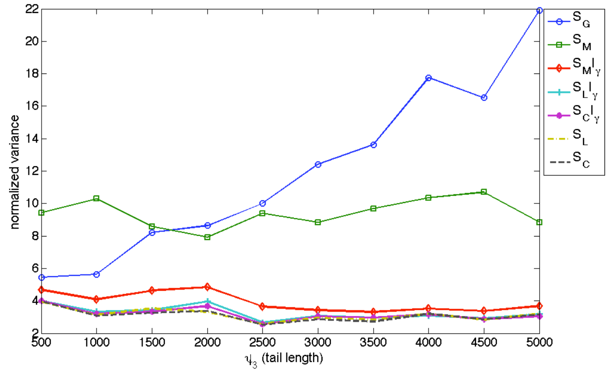

Fig. 7. Simulation varying the tail length. Parameters are given in Table I.

$\bar{a} \approx 2.6\%$ and $\sigma_a^2 \approx 3.7 \cdot 10^{-4}$ (hence, $\theta \approx 68$). An empirical histogram of a sample set of a_i's is given in Fig. 5.

B. Simulation results

In this section we provide some illustrative MATLAB simulation results with the Uniform-mixture family of distribution for the n_i's. The number of terminals is set to $I = 500$. For each simulation setting (single point in the parameter space) we performed 100 simulations and computed the Mean Squared Value of the estimation error, i.e. its (empirical) variance. In order to take into account for the finite size of the terminal set ($I < \infty$) we rescale the empirical variance by σ_a^2/I which represents the theoretical limit of the estimate uncertainty for $n_i \to \infty, \forall i$, i.e. for an infinite number of per-terminal observations — but finite number of terminals. The resulting quantity, which we refer to as "normalized variance", is plotted in the following graphs for a set of different estimators. Unless differently specified, the cutoff threshold was set to $\gamma = 10$.

In a first set of simulations (ref. Fig. 6) we investigated the impact of thin users: we varied the size (height) of the distribution head by varying the parameter $\phi_1 \in [0, 0.9]$ in steps of 0.1. The tail share was kept fixed $\phi_3 = 0.05$ while the body share was adjusted as $\phi_2 = 1 - \phi_1 - \phi_3$.

In a second set of simulations (ref. Fig. 7) we investigated the impact of fat users: we varied the size (length) of the distribution tail by varying the parameter $\psi_3 \in [500, 5000]$ in steps of 500. All other parameters were kept at the default values shown in Table I.

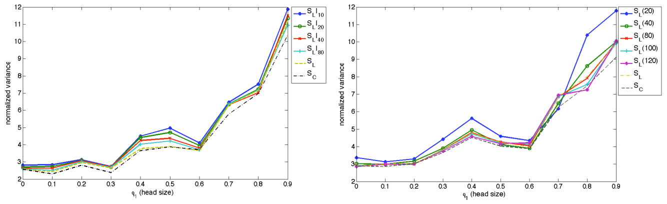

Fig. 8. Sensitivity analysis of $S_L|_\gamma$ (left) and $S_L(\theta)$ (right).

It can be seen that the EMR and EGR do not perform well in all scenarios. In particular, Fig. 6 shows that the EMR suffers when the number of thin users grows. This was expected since each thin user brings a relatively inaccurate estimate of their individual loss level a_i into the overall estimate S_M — recall that $\mathrm{VAR}[r_i] \propto n_i^{-1}$ — and since all r_i's are weighted the same the more reliable estimates provided by other users with larger n_i are overridden. As expected, the mean ratio computed from the partial set $S_M|_\gamma$ (with $\gamma = 10$) performs better than S_M, but still its performance degrades appreciably when the head share is high because the number of residual samples used for the estimation becomes dramatically low.

Regarding EGR, Fig. 7 shows that it suffers when the size of the few fattest users grows. Again this was expected: while fat users bring very accurate estimates of their individual loss levels a_i (large n_i), the whole estimator S_G is dominated by them and this causes large fluctuations in the estimate when the number of fat users is low. In other words, while EMR suffers from the inaccuracy of thin users, the EGR suffers from the scarcity of fat users. One possible approach to "correct" the EGR is to treat fat users as "outliers" and filter them — similarly to the filtering of thin users done for $S_M|_\gamma$. Such strategy has two drawbacks: first it requires an heuristic method to identify which terminals must be considered "outliers", and such heuristic must be tailored on the particular distribution of the n_i's (lack of generality). Second, by filtering out fat users one discards a certain number of very good estimates of a_i's — actually the most reliable ones — and if the number of fat users grows this translates into a non negligible loss of information.

Figs. 6 and 7 show that all the proposed weighted estimators $S_L|_\gamma$ and $S_C|_\gamma$ perform very close to the theoretical optimum (S_C): the reason for such good performance is that the observations from thin and fat users are never discarded but only weighted differently.

In order to investigate the sensitivity of $S_L|_\gamma$ to the parameter γ, i.e. the cutoff threshold for the initial estimate $\widehat{\theta}|_\gamma$ of $\dot{\theta}$, we have performed tests with different values of γ in a sample scenario. The results reported in Fig. 8 (left) show only small variations in the performance between different values of γ (note the different vertical range from previous figures). This indicates that the performances of the piecewise-linear EWR estimator are not too sensitive to the exact location of the knee-point. This claim is further confirmed in Fig. 8 (right), where we tested the $S_L(\theta)$ using different static settings of θ.

C. Summary of findings

The above results collectively indicate that, although the "optimum" weights $\dot{\boldsymbol{w}}$, $\ddot{\boldsymbol{w}}$ can not be attained in practice

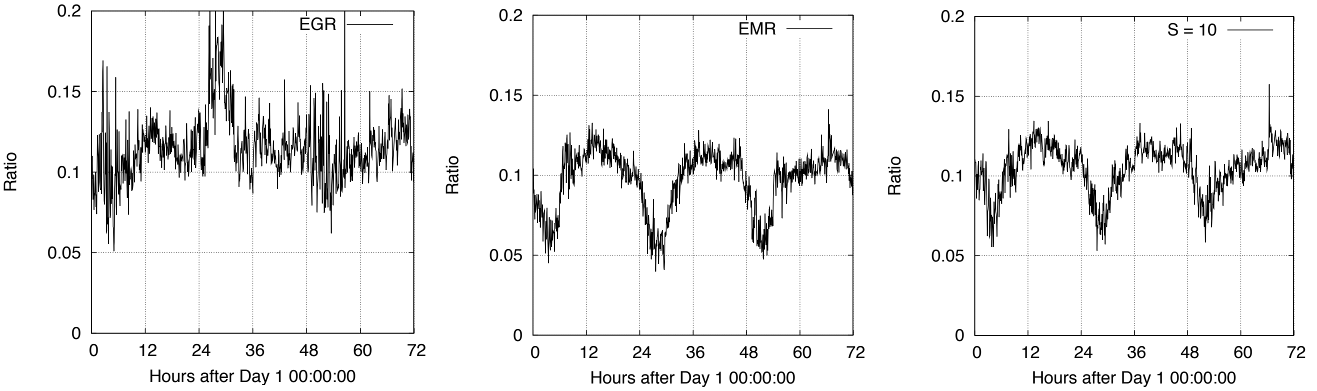

Fig. 9. Ratio of unacknowledged downlink SYNACK in a real GPRS network, 5 min bins. Left to right: EGR, EMR, $S_L(\theta)$ with static setting $\theta = 10$.

— recall from eqs. (13) and (15) that they depend on the unknowns $\bar{a}, \sigma_a^2$ — any reasonable approximation seems to work surprisingly good in a wide range of scenarios. This indicates that the geometry of the problem is such that differential weighting for low/high n_i is the key feature of a good solution, while the detailed shape of the weight vector plays a secondary role. This justifies the use of the piecewise linear approximation $S_L(\theta)$ which is more handy to compute than eq. (13). Moreover, it suggests that in practical scenarios the fine-tuning of the knee-point θ is a secondary problem: any "reasonable" value of θ would do a good job and deliver better performance than EMR/EGR also in extreme conditions. In other words, the sensitivity of $S_L(\theta)$ to such parameter is moderate, provided that the choice is reasonable. This point can motivate in practical applications the adoption of a static setting for θ, thus avoiding the initial estimation of $\bar{a}, \sigma_a^2$.

D. Measurements from a real network

Here we report some preliminary results from the application of EWR to real network data. Data were collected with the measurement procedure described in an earlier work [7] (see also Sec. VI-A and Fig. 3) at the beginning of July '09. Here the n_i's count the number of SYNACK packets originated by Internet servers and directed to GPRS users, in timebins of 5 min. The m_i's count the number of unacknowledged SYNACKS (note that the early retransmitting servers were dynamically identified and pre-filtered from the dataset, see [7] for a discussion on this issue).

In Fig. 9 we report the values estimated with S_G, S_M and $S_L(10)$, i.e. EWR with a static threshold setting $\theta = 10$. It can be seen that the temporal profile of S_G is very noisy, likely due to the presence of fat users. The profile of $S_L(10)$ is only slightly more clean (less fluctuating) than S_M, which seems to perform rather well in this case. A short spike is clearly visible in $S_L(10)$ in the last day, which might correspond to a transient anomaly an is currently being investigated.

VII. Conclusions and ongoing works

In this contribution we have addressed the problem of inferring the network-wide mean level of packet loss from passive measurements, which we have formulated as an estimation optimization problem. We have derived a near-optimum weighted estimator and identified a simple piecewise-linear form for the weights, with a single parameter θ. We have found analytically that the optimal value of θ does not depend on the size of the individual observations (n_i's) but only on the first two moments of the individual per-terminal loss probabilities. Furthermore, we have observed that the performance of such estimators are relatively insensitive to the particular setting of θ, at least for a wide range of values around the theoretical optimum. Such findings motivate the use of such estimators in dynamic contexts where the distribution of the n_i's is time-variant, as typically found in cellular networks due to the daily/weekly activity cycles of the users. Due to the extreme simplicity, better performance and generality, such class of estimators lend themselves very well to be adopted in practical scenarios, for example as global Key Quality Indicators (KQI). In the progress of this work we are seeking to identify a set of practical guidelines and a "rule of thumb" to guide the setting of the parameter θ in real-world scenarios. Furthermore, we are investigating the behavior of the estimator in the presence of measurement correlations, an aspect that can not be disregarded in real-world data sets.

References

[1] Y. Tsang, M. Coates, R. Nowak: Passive Network Tomography using EM algorithms, *Proc. of the IEEE ICASSP '01*

[2] R. Cáceres, N.G. Duffield, J. Horowitz, D. Towsley: Multicast-based inference of network-internal loss characteristics, *IEEE Trans. on Information Theory*, 1999

[3] R. Cáceres, N.G. Duffield, S.B. Moon, D. Towsley: Inference of internal loss rates in the MBone, *IEEE/ISOC Global Internet '99*

[4] H. Su, W. Chen, S. Lin, D. Jin, L. Zeng: The inference of link loss rates with internal monitors, *Proc. of the GLOBECOM '08*

[5] V.N. Padmanabhan, L. Qiu, H.J. Wang: Server-based Inference of Internet Link Lossiness, *Proc. of the IEEE INFOCOM '03*

[6] H. Tian, H. Shen: Multicast-Based Inference of Network-Internal Loss Performance, *Proc. of 7th Int'l Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04)*

[7] Peter Romirer *et al.* Network-Wide Measurements of TCP RTT in 3G, *Proc. of TMA'09 workshop, Aachen, May 2009, in LNCS vol. 5537.*

[8] A. D'Alconzo, A. Coluccia, F. Ricciato, P. Romirer-Maierhofer: A Distribution-Based Approach to Anomaly Detection for 3G Mobile Network, *to appear in IEEE Globecom '09*

[9] E.L. Lehmann, G. Casella: Theory of Point Estimation, *Springer*, 1998

[10] K.M. Wolter: Introduction to Variance Estimation, *Springer Series in Statistics*, 2007

[11] METAWIN and DARWIN projects http://userver.ftw.at/~ricciato/darwin/