A Frequency Domain Model to Predict the Estimation Accuracy of Packet Sampling

Luigi Alfredo Grieco
DEE - Politecnico di Bari - Italy
Email: a.grieco@poliba.it

Chadi Barakat
INRIA - Sophia Antipolis, France
Email: chadi.barakat@sophia.inria.fr

***Abstract*—In network measurement systems, packet sampling techniques are usually adopted to reduce the overall amount of data to collect and process. Being based on a subset of packets, they hence introduce estimation errors that have to be properly counteracted by a fine tuning of the sampling strategy and sophisticated inversion methods. This problem has been deeply investigated in the literature with particular attention to the statistical properties of packet sampling and the recovery of the original network measurements. Herein, we propose a novel approach to predict the energy of the sampling error on the real time traffic volume estimation, based on a spectral analysis in the frequency domain. We start by demonstrating that errors due to packet sampling can be modeled as an aliasing effect in the frequency domain. Then, we exploit this theoretical finding to derive closed-form expressions for the Signal-to-Noise Ratio (SNR), able to predict the distortion of traffic volume estimates over time. The accuracy of the proposed SNR metric is validated by means of real packet traces.**

I. Introduction

In network measurement systems, packet sampling techniques are usually adopted by network operators to reduce the overall amount of packets to capture and process [3]. They simply consist on capturing a subset of packets (randomly or periodically), which are then used to infer the original traffic properties. Obviously, such techniques introduce estimation errors that have to be properly counteracted by a fine tuning of the sampling strategy [5]. Given an upper limit on the tolerated load on routers and a targeted measurement task, there is always an appropriate sampling and inversion method [1] that allows the measurement to be carried out with the best possible accuracy. The mostly known sampling pattern consists of a random selection of packets at the interfaces of routers with some predefined and homogeneous probability p. This probability p is called the *sampling rate* and is often set by operators to low values such as 1/100 or even 1/1000 according to the bitrate of links.

The performance of packet sampling has been deeply investigated in the literature with particular attention to the statistical properties of the sampled measurements and the way they should be inverted to recover the original traffic properties [4]. Several metrics were studied as the traffic volume both in packets and bytes, the volume of the largest flows[2] often called heavy hitters, the number of flows and the distribution of the flow volumes. These previous works, among others, have shed some light on many of the statistical properties of packet sampling. Several inversion methods have followed combining stochastic analysis and statistical inference. Most often, the aim of the inversion was to minimize the variance of the estimation given a set of sampled packets collected during some time interval. However, in practice the traffic is not fixed, but varies over time forming a signal composed of several frequencies. Packet sampling would then have different impact if studied over time rather than over some fixed time interval or a set of well defined packets. Instead of inverting a set of sampled packets with the minimum estimation error variance, one can ask the question of how to infer the spectral density of the original traffic with the minimum signal-to-noise ratio. We are sure this way that the main frequencies in the original traffic are preserved, which is of major importance for applications like anomaly detection and network tomography [2], [6].

In a recent communication [4], we had a look at packet sampling from the viewpoint of the spectral density of the original traffic. Our targeted measurement was the rate of the traffic in packets/s or bytes/s when tracked over time on some router interface and binned over time windows equal to T. We came up with a model in the frequency domain that explains the impact of packet sampling. As in classical signal theory, packet sampling was shown to introduce bias because of the replicas of the baseband component. For network traffic with some well defined maximum cutoff frequency f_M, provided that p is sufficiently high, this bias can be eliminated by a proper low-pass filtering of the sampled traffic signal followed by an upscaling of the sampled signal by the inverse of the sampling rate. In particular, we show in [4] that in order to avoid aliasing effects, the inequality $0.445/T < p/t_0 - f_M$ should hold, where t_0 corresponds to the transmission time of the smallest packet over the monitored link. This theoretically ensures that both the inverted traffic rate and the real one possess the same spectral density when binned over T. Unfortunately, the previous manipulation requires the existing of a maximum cutoff frequency in the network traffic, which

This research work is partially funded by the European Commission through the FP7 ECODE project (INFSO-ICT-223936).

[1] In sampling terminology, *inversion* is the process of estimating original traffic properties from sampled measurements.

[2] A flow is a set of packets sharing common fields in their headers as the IP source address prefix, the IP destination address prefix and the port number.

might not be the case as our measurements show (for now see Fig. 3). Second, outside the no aliasing region, the previous analysis does not give any idea on the energy of the noise introduced by packet sampling and on how much the inverted sampled traffic differs from the original one.

In this paper, we exploit the theoretical finding in [4] and derive closed-form expressions for the Signal-to-Noise Ratio (SNR) that are able to predict the distortion of the traffic rate estimate. The proposed SNR model calculates for a given packet sampling probability p and averaging time window T, the amount of error in each frequency band of the traffic signal, thus allowing to trade off sampling overhead with frequency resolution of the measurements. Indeed, for a fixed p, increasing T allows smaller estimation errors to be obtained at the expense of a worse time resolution. On the opposite, to achieve a small estimation error using a small value of T, one needs very high values of p, with a consequent increase in the monitoring overhead. With our expressions of the SNR, network operators can tune their monitoring system, i.e., T and p, in order to achieve the desired accuracy at the desired time resolution. Moreover, being expressed in closed-form, our results can be easily implemented in real network monitoring systems. Finally, we demonstrate the accuracy of the proposed SNR expressions against real packet traces collected by the MAWI project over trans-pacific links [1].

II. Problem formulation in the frequency domain

We outline in this section our model for the analysis of packet sampling in the frequency domain. This model has been introduced for the first time in [4] and is exploited in this paper to calculate the signal-to-noise ratio while inverting the sampled traffic. Our targeted measurement is the amount of data sent from a sender node (S) to a receiver node (R) averaged over time intervals of duration T, which will be referred to as bins. Our objective is to capture correctly the spectrum, and hence the amplitude and oscillations, of this time varying signal. In our analysis, a node can be a net or a subnet with some IP address prefix, a domain, an edge router, or even a single host. The estimation of the binned values of the traffic is carried out using packet sampling, i.e., each packet is captured with a uniform probability p, then the number of captured packets per bin is divided by the sampling rate p to infer the rate of the traffic for that bin. The monitor moves then to the following bin and so on. To model the spectral density of the traffic signal, we divide the time axis into small time slots with size t_0. In each slot, no more than one packet can be transmitted. In practice, this t_0 corresponds to the transmission time of the smallest packet over the monitored link. Next, we define $d(k)$ as the amount of data sent by S during the time interval $[kt_0, (k+1)t_0)$, where $k \in N$. To be more precise, if the transmission of an entire packet is accomplished during the time interval $[kt_0, (k+1)t_0)$, $d(k)$ will be equal to the size of the sent packet, otherwise $d(k)$ will be equal to 0. Moreover, we take the bin size T to be an integer multiple of t_0, i.e., T is made by T/t_0 slots. The expected Fourier Transform of $d(k)$ can be expressed as follows [7]:

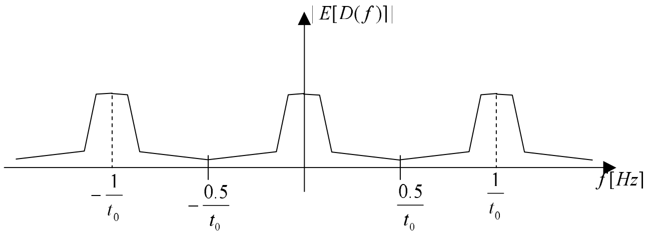


Fig. 1. Expected spectrum of original packet stream $d(k)$.

$$E[D(f)] = \sum_{k=-\infty}^{+\infty} E[d_k] \cdot e^{-j2\pi k f t_0} = \sum_{n=-\infty}^{+\infty} D_0(f - \frac{n}{t_0}), \quad (1)$$

where $D_0(f) = 0$, for $|f| > \frac{0.5}{t_0}$. This expression has a general validity because the spectrum of any discrete-time signal is periodic with period equal to $1/t_0$, if the time between two subsequent samples is equal to t_0. Basically, $D_0(f)$ is a function that we introduce and that includes all frequencies of the signal $d(k)$ in the interval $[-0.5/t_0, +0.5/t_0]$. To better clarify the meaning of our notation, Fig. 1 pictures a typical example of $E[D(f)]$.

As first step, we model the spectrum of the traffic signal under the ideal assumption of capturing all packets, i.e., $p = 1$. The spectrum of the sampled traffic is derived in the next section. Given that the measurement bin lasts T/t_0 slots, summing the data received in a bin time can be seen as filtering $d(k)$ using a discrete-time filter with pulse response $h(k) = 1$ for $k = 0 \ldots \frac{T}{t_0} - 1$, and $h(k) = 0$ for $k \geq \frac{T}{t_0}$. The corresponding transfer function is:

$$H(f) = e^{-j\pi f(\frac{T}{t_0}-1)t_0} \cdot sin(\pi f T)/sin(\pi f t_0). \quad (2)$$

$H(f)$ is a low-pass filter with cutoff bandwidth $B \approx \frac{0.445}{T}$ and static gain equal to T/t_0 [4]. The expected Fourier Transform of $\bar{d}(k)$, i.e., the filtered version of $d(k)$, is:

$$E[\bar{D}(f)] = H(f)E[D(f)] = \frac{T}{t_0} \sum_{n=-\infty}^{+\infty} \bar{D}_0(f - n/t_0), \quad (3)$$

where $\bar{D}_0(f) = t_0 \frac{H(f)D_0(f)}{T}$. The last equality in Eq. (3) holds because both $H(f)$ and $E[D(f)]$ are periodic functions with the same period $1/t_0$. Now, we present our approach to move from a discrete-time signal representation to a continuous-time one. The signal $\bar{d}(k)$ is decimated by a factor T/t_0, i.e., one sample of $\bar{d}(k)$ is taken every bin, then the resulting signal $\bar{d}_T(k)$ is processed with Zero Order Holder (ZOH), which is a device that keeps the output $\hat{d}(t)$ equal to the last received sample. Using the Poisson summation formula [7], the expected spectrum of $\bar{d}_T(k)$, i.e., the decimated version of $\bar{d}(k)$, is:

$$E[\bar{D}_T(f)] = \sum_{n=-\infty}^{+\infty} \bar{D}_0(f - n/T). \quad (4)$$

It is worth noting that the spectrum $E[\bar{D}_T(f)]$ is the sum of the functions $\bar{D}_0(f - \frac{n}{T})$, which are obtained by translating

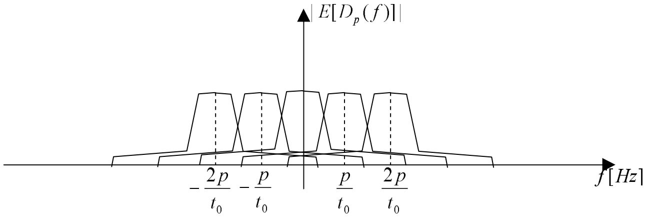


Fig. 2. Expected spectrum of sampled packet stream $d_p(k)$.

$\frac{T}{t_0} \cdot \bar{D}_0(f)$ by integer multiples of $\frac{1}{T}$ and by dividing the result by T/t_0. As a consequence, and given that the bandwidth of $\bar{D}_0(f)$ is $B \approx \frac{0.445}{T}$ [4], the decimation does not introduce aliasing. Moreover, the transfer function of the ZOH, namely $G_{ZOH}(f)$, is a low-pass filter with unitary static gain and bandwidth equal to that of $H(f)$. With respect to $H(f)$, the ZOH is able to filter out all high frequency components of the input signal, so that, the expected spectrum of the continuous-time signal $\hat{d}(t)$ is no more periodic and can be expressed as follows:

$$E[\hat{D}(f)] = G_{ZOH}(f)E[\bar{D}_T(f)] \approx G_{ZOH}(f)\bar{D}_0(f). \tag{5}$$

The signal $\hat{d}(t)$ is what network operators track over time on their router interfaces. Our aim is to evaluate the impact of packet sampling on the spectrum of this signal and hence propose conservative values for T and p to be used. Such values should ensure that the estimated binned traffic and the real binned traffic do not differ from each other by more than the error margin defined by the operator. Note that most of the difficulty comes from the fact that the spectrum of the original signal $d(k)$ is unknown from sampled traffic, so one has to estimate it jointly with the optimization of p and T.

III. Spectrum of sampled traffic

Following the same methodology, we derive the spectrum of the binned traffic rate estimated from sampled packets. Suppose that packets are sampled with some uniform probability $0 < p < 1$ and denote by $d_p(k)$ the volume of sampled data in the time slot $[kt_0, (k+1)t_0)$, $k \in N$. The signals $d(k)$ and $d_p(k)$ are related to each other, as for each k, $d_p(k)$ is equal to $d(k)$ with probability p and to 0 with probability $1-p$. Since we are interested in low-frequency components, with $|f| < \frac{1}{T}$, we can safely assume that the time bin size is much larger than the jitter of the inter-arrival time between sampled packets. Thus, in the frequency band of interest, the spectrum of the sampled traffic $E[D_p(f)]$ can be viewed as the spectrum of the original traffic $E[d_k]$ sub-sampled with frequency $\frac{p}{t_0}$ (see also [4] for more details). Recalling the spectrum of the signal $E[d_k]$ reported in Eq. (1), it holds that [7]:

$$E[D_p(f)] \approx p \sum_{n=-\infty}^{+\infty} D_0(f - n \cdot p/t_0). \tag{6}$$

An example of this spectrum is plotted in Fig. 2 where we can see the aliasing introduced by packet sampling. The effect of the aliasing cannot be fully filtered out given the overlap of baseband replicas. However, the amount of noise due to aliasing can be reduced by low-pass filtering $d_p(k)$. That is exactly what the binning of the sampled traffic ensures, $H(f)$ defined in Eq. (2) being the transfer function of the resulting low-pass filter. Thus, after packet sampling, the reduced traffic $d_p(k)$ is filtered using $H(f)$ to obtain the signal $\bar{d}_p(k)$. Finally, in order to move to a continuous time representation that models the averaging of the sampled traffic over bins of T/t_0 slots, the signal $\bar{d}_p(k)$ has to be decimated by a factor T/t_0 before being interpolated by a ZOH. Using the Poisson summation formula as done to derive Eq. (4), and provided that the filter $H(f)$ has made negligible the aliasing due to the sampling, the expected spectrum of $\bar{d}_{(p,T)}(k)$, i.e., the decimated version of $\bar{d}_p(k)$, can be written as:

$$E[\bar{D}_{(p,T)}(f)] \approx p \sum_{n=-\infty}^{+\infty} \bar{D}_0(f - n/T). \tag{7}$$

Next, by applying the ZOH, one can extract a continuous time reconstruction of the sampled traffic whose spectrum is $pG_{ZOH}(f)\bar{D}_0(f)$, i.e., a low-pass filtered version of the base-band component of the average spectrum of $d(k)$ scaled down by p. Compared to Eq (5), this confirms indeed that the signal $d_p(k)$ modeling the sampled traffic has to be divided by p to compensate the scaling due to the aliasing and to obtain the same spectrum as the time averaged reconstruction of the original traffic. From now on, we will always consider the inverted signal $d_p(k)/p$ in our analysis.

IV. Signal-to-Noise Ratio

In this section we propose robust metrics that evaluate the reliability of the traffic rate estimation from sampled traffic. Our metrics are function of the sampling rate p, the bin size T, the probability to find a busy slot in the original traffic p_L, i.e., $p_L = P(d(k) > 0)$, and the first and second order moments of the packet size, which will be referred to as α and β, respectively. All these parameters can be calculated from the sampled traffic without having access to the original traffic, hence the interest of our approach. To clarify this approach, we will first derive an expression for the Signal-to-Noise Ratio (SNR) assuming that all packets have the same size. Then, we will extend this finding to the most general case of having packets with different sizes. The effectiveness of both SNR expressions will be proved using real packet traces. They will be computed by exploiting the theoretical findings presented in the previous sections. Moreover, the analysis will be made even more realistic by using the Discrete Fourier Transform (DFT) [7], which works with finite sets of collected packets. The regular Fourier Transform from its side requires an infinite set to be estimated.

A. SNR under a Constant Packet Size

Given a discrete signal $d(k), k = 0 \cdots N-1$, modeling packet sizes accross N time slots, its DFT coefficients

$D_{dft}(n), n = 0 \cdots N-1$, can be expressed as follows:

$$D_{dft}(n) = \sum_{k=0}^{N-1} d(k) e^{-\frac{2\pi j}{N} kn}, n = 0 \cdots N-1. \qquad (8)$$

In this analysis we consider that all packets have the same size d_0, an assumption that we relax later. This implies that $d(k)$ can be either equal to d_0 or 0. Without loss of generality, we take $d_0 = 1$. [3] From the Parseval theorem, we can write that:

$$\sum_{k=0}^{N-1} d^2(k) = \frac{1}{N} \sum_{n=0}^{N-1} |D_{dft}(n)|^2. \qquad (9)$$

The summation on the left-hand side of the equation is no other than the energy carried by the original packet stream. It follows that the n-th DFT coefficient carries an amount of energy equal to $|D_{dft}(n)|^2/N$. On the other hand, one can write the total energy of the signal in this particular case of packet of size equal to one as:

$$E^T = \sum_{k=0}^{N-1} d^2(k) = p_L \cdot N. \qquad (10)$$

Recalling Eq. (8), we can write that $D_{dft}(0) = p_L \cdot N$. As a consequence, the coefficient $D_{dft}(0)$, which models the continuous component of the traffic signal, carries an amount of energy equal to $|D_{dft}(0)|^2/N = p_L E^T$. Motivated by real traffic spectrum measurements reported in Sec. V, we suppose that the remaining energy of the signal $(1 - p_L)E^T$ is uniformly spread over the $D_{dft}(n)$ coefficients having $1 \leq n < N$. Thus, we can derive that the amount of energy carried by the generic $D_{dft}(n), 1 \leq n < N$, is equal to $p_L(1-p_L)\frac{N}{N-1} \approx p_L(1-p_L)$, for sufficiently large values of N. Given this spectrum description, it follows that the energy of the signal $\bar{d}(k)$, i.e., a filtered version of $d(k)$ obtained using a low-pass filter with a bilateral bandwidth equal to $\frac{N_B}{Nt_0}$, is given by:

$$E^S = p_L^2 N + p_L(1-p_L)N_B. \qquad (11)$$

N_B can be seen as the number of frequency components around the continuous one that are allowed by the low-pass filter. In a similar way, we can write the energy of the tail of $d(k)$ comprised in a bandwidth of size $\frac{N_B}{Nt_0}$ as:

$$E_{tail} = p_L(1-p_L)N_B. \qquad (12)$$

Now, we have to recall that the noise in the sampled traffic signal $d_p(k)$ caused by aliasing is due to the tails of the spectrum of the original traffic signal $d(k)$ translated and folded together in the bandwidth of interest. If the sampling probability is p, we expect to have a number of replicas equal to $\frac{1-p}{p}$. Each replica is scaled down by p because of sampling as stated by Eq. (7). To compensate for this, we amplify the signal $d_p(k)$ by $1/p$. The energy of the noise E^N becomes:

$$E^N = (1-p)E_{tail}/p = p_L(1-p)(1-p_L)N_B/p. \qquad (13)$$

The SNR value we are looking for can be computed as $SNR = E^S/E^N$. This metric is a function of p_L, which can be accurately estimated by dividing the number of sampled packets by pN. As for N_B, it is related to the time bin over which the traffic is monitored. A binning over T seconds is equivalent to filtering the traffic signal using a low-pass filter of bilateral bandwidth equal to $\frac{0.89}{T} = \frac{N_B}{Nt_0}$. By replacing N_B by its expression as a function of T, we get the final expression for the SNR metric for constant packet sizes:

$$SNR = \frac{p_L + (1-p_L)0.89t_0/T}{\frac{1-p}{p}(1-p_L)0.89t_0/T}. \qquad (14)$$

B. SNR in a realistic scenario

Here, we relax the assumption of having a constant packet size in order to provide a SNR metric that better reflects realistic scenarios. Our main finding can be summarized as follows:

Proposition 1: Let α and β be respectively the first and second order moments of the packet size. Let p_L be the probability to find a busy slot in the original traffic, i.e., $p_L = P(d(k) > 0)$, p the sampling rate, and T the time bin length. The Signal-to-Noise Ratio (SNR) caused by sampling can be approximated by

$$SNR = \frac{p_L\alpha^2 + (\beta - p_L\alpha^2)0.89t_0/T}{\frac{1-p}{p}(\beta - p_L\alpha^2)0.89t_0/T}. \qquad (15)$$

Proof: In this case of variable packet sizes, the expected value q of the total energy carried by the original traffic can be defined and estimated as follows:

$$q = E[E^T] = \sum_{k=0}^{N-1} E[d^2(k)] \approx N \cdot p_L \cdot \beta. \qquad (16)$$

Moreover, the expected value for $D_{dft}(0)$ can be estimated as:

$$E[D_{dft}(0)] = \sum_{k=0}^{N-1} E[d(k)] \approx N \cdot p_L \cdot \alpha. \qquad (17)$$

This gives the following approximation for the energy t associated to the first DFT coefficient:

$$t = |E[D_{dft}(0)]|^2/N \approx N \cdot (p_L \cdot \alpha)^2. \qquad (18)$$

Now, assuming that the remaining energy of the signal $q - t$ is uniformly spread over the other $D_{dft}(n)$ coefficients having $1 \leq n < N$, we can derive that the amount of energy carried by the generic $D_{dft}(n)$, $1 \leq n < N$, is equal to $\frac{q-t}{N-1}$. The energy of the signal $\bar{d}(k)$, i.e., the filtered version of $d(k)$ obtained using a low-pass filter with a bilateral bandwidth equal to $\frac{N_B}{Nt_0}$, becomes equal to:

$$E^S = t + \frac{q-t}{N-1} N_B. \qquad (19)$$

Moreover, the energy of the tail of d comprised in a bandwidth of size $\frac{N_B}{Nt_0}$ is equal to:

$$E_{tail} = (q-t)N_B/(N-1). \qquad (20)$$

[3] We are computing a ratio between the energy of the signal and the energy of the noise. Any value assigned to d_0 would disappear from our expressions.

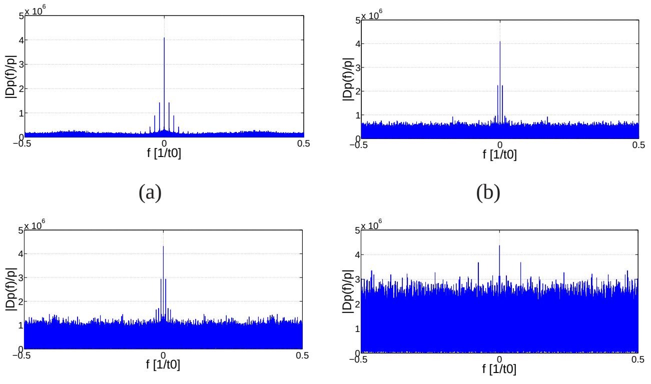
As for the constant packet size case, the noise in the signal $d_p(k)$ caused by aliasing is due to the tail of the spectrum of the signal $d(k)$ translated and folded together in the bandwidth of interest. Having in total $\frac{1-p}{p}$ replicas that overlap with the baseband component, the energy of the noise E^N can be written as $E^N = (1-p)E_{tail}/p$. Now, by considering that: (i) $SNR = E^S/E^N$; (ii) $N/(N-1) \approx 1$ for a sufficiently large N; and (iii) $\frac{N_B}{Nt_0} = \frac{0.89}{T}$ for time bins of length T, one can find the result (15) stated in the proposition. □

V. Experimental Results

Here we validate the effectiveness of the SNR metrics proposed in Eqs. (14) and (15). For that purpose, we have analyzed three distinct packet traces from the MAWI project collected at two trans-pacific 150 Mbps links during January 2009 and December 2005 [4]. The traces, of duration 15 minutes each, have been sampled with probabilities ranging in the interval $[10^{-4}, 1]$. For each sampling probability, 5 distinct experiments have been repeated using different seeds for the random number generator. Sampled traces have been averaged over time bins of length T, ranging in the interval $[1s, 400s]$. For those experiments where at least one packet was sampled, we have compared the measured SNR with respect to the estimated one, using Eqs. (14) and (15). The first equation supposes all packets of equal size, whereas the second one considers packet size variability.

Fig. 3 shows the module of the spectrum of the inverted sampled traffic $d_p(k)/p$, obtained for several values of p from the Trace 1. The time slot is set to the minimum packet size available in the trace divided by the link speed, and sampled packets are assigned to time slots using their timestamps. By comparing the plot obtained for $p = 1$ with respect to the other ones, it is straightforward to note that: (i) only low frequencies of the original traffic can be recovered, even using a very high sampling probability as $p = 0.1$; (ii) the harmonic tones of the original traffic, i.e., those obtained for $p = 1$, appear translated in the frequency spectrum of the sampled traffic signals as expected by the Poisson summation formula; (iii) the noise across the continuous component of the traffic signal grows with $1/p$ as expected by Eqs. (14) and (15).

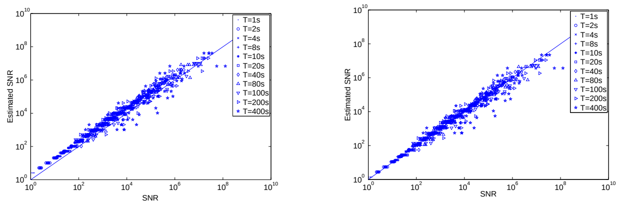
Furthermore, in Fig. 4, we report in a scatter plot both the measured and estimated SNRs on the Trace 1, using Eqs. (14) and (15). Analogous results obtained for Traces 2 and 3 have not been reported due to lack of space. Each dot in the figures corresponds to one experiment with one p and T. These plots show that, for the considered trace, the proposed SNR models, and in particular the one in Eq. (15), lead to dots around the diagonal, which is an indication of high model accuracy. Scenarios with low SNR values as when p is low or T is small enjoy from an even better match of the results. Indeed, for these challenging scenarios, the aliasing is important and our assumption of equal distribution of the energy over frequency components other than the continuous one better holds.



(a) (b)

(c) (d)

Fig. 3. Baseband component of $D_p(f)/p$: (a) $p = 1$; (b) $p = 0.1$; (c) $p = 0.03$; (d) $p = 0.005$.



(a) (b)

Fig. 4. SNR estimated using: (a) Eq. (14); (b) Eq. (15).

As a conclusive consideration, we can say that the proposed SNR models are accurate enough to be exploited to rank the reliability of traffic rate estimates when p sampling is applied by routers at low layers and when the rate is monitored over a finite time window of duration T. This evaluation can be done without having access to the original traffic itself. All what one has to do is to define some SNR threshold, then tune T using Eq. (14) or Eq. (15), accordingly. If some time resolution is required in addition, the sampling rate has to be set high enough using our close expressions for the SNR. Diverse network applications as anomaly detection and network tomography can profit from such optimal and fast setting of the sampling rate according to the required time resolution of the monitored traffic.

References

[1] Mawi working group traffic archive. http://tracer.csl.sony.co.jp/mawi/.

[2] D. Brauckoff, B. Tellenbach, A. Wagner, M. May, and A. Lakhina. Impact of packet sampling on anomaly detection metrics. In *Proc. of ACM SIGCOMM IMC 2006*.

[3] K. C. Claffy, G. C. Polyzos., and K. W. Braun. Application of sampling methodologies to network traffic characterization. *ACM SIGCOMM Comput. Commun. Rev.*, 23(4), 1993.

[4] L. A. Grieco and C. Barakat. An analysis of packet sampling in the frequency domain. In *Proc. of ACM SIGCOMM IMC 2009*.

[5] N. Hohn and D. Veitch. Inverting sampled traffic. *IEEE/ACM Trans. on Networking*, 14(1):68–80, 2006.

[6] P. Kanuparthy, C. Dovrolis, and M. Ammar. Spectral probing, crosstalk and frequency multiplexing in internet paths. In *Proc. of ACM SIGCOMM IMC 2008*.

[7] J. Proakis and D. G. Manolakis. *Digital Signal Processing*. Prentice Hall, Int. Eds., 3 edition, 1996.

[4] The traces are available at http://mawi.wide.ad.jp/mawi/samplepoint-F/2009/ and http://tracer.csl.sony.co.jp/mawi/samplepoint-B/2006/