END-TO-END QUALITY OF SERVICE PREDICTION BASED ON FUNCTIONAL REGRESSION

L. ASPIROT, P. BELZARENA, G. PERERA, B. BAZZANO

ARTES *

FACULTAD DE INGENIERÍA, UNIVERSIDAD DE LA REPÚBLICA
MONTEVIDEO, URUGUAY

Abstract. This work focus on quality of service (QoS) estimation based on end-to-end active measurements. The main problem is to continuously monitor the QoS that will receive a multimedia application in a path between two agents on the Internet. We want to estimate the QoS without sending 'heavy' multimedia traffic in order to measure its QoS parameters (delay, losses, jitter, etc.). In this paper we extend a recent work of Ferraty et al. [5] on functional nonparametric regression. We generalize it to include a very important case on the Internet: nonstationary traffic. We apply this result to our problem describing an end-to-end active measurement methodology. We also show by simulations how it predicts the QoS of a multimedia application in a link where its capacity and buffer sizes are unknown. The cross traffic is also unknown and can be a nonstationary process.

Keywords. End-to-end active measurements, nonstationary, QoS, functional regression.

1. Introduction and Motivation

With the new services offered over the Internet (particulary voice and video), the need to measure the performance of the network has increased. Measurements of the Internet performance are necessary for different reasons, for example to advance in understanding the behavior of the Internet or to verify the QoS assured to the new services.

Multimedia services in a packet switched network have some QoS problems like delay, jitter, packet losses and, in some cases, degradation due to low bit rate codification.

An IP network like Internet has some problems to measure the performance parameters since the route of the packets can change, the traffic bit rate is not constant and normally is in burst, the probe packets can be filtered or altered by one ISP (Internet Service Provider) in the path, there is not clock synchronization between routers and end equipments, etc. Normally the internal routers in the path between two points of interest are not under the control of only one user or one ISP. Therefore, it is not useful to have measuring procedures that depend on the information of the internal routers. For this reason end-to-end measures is one of the most developed methodologies during last years.

This work focus on the development of an end-to-end active measurement methodology that allows to estimates the QoS parameters of a multimedia application sending the least possible amount of probe packets.

Our goal is to estimate the QoS of an Internet multimedia service for applications like the following.

(1) to monitor and verify during long time periods a Service Level Agreement between a user and a ISP or between ISPs.
(2) to predict the QoS that will experiment a video stream or one or more voice conversations along an Internet path.

* **ARTES**: Joint Research Group of the Electrical Engineering and Mathematics and Statistics Departments. **Contact:**artes@fing.edu.uy
This research was partially supported by PDT (Programa de Desarrollo Tecnolgico, Prstamo 1293/OC-UR): S/C/OP/ 17/02 and program FCE (Fondo Clemente Estable) 8079.

(3) to monitor the paths between one video or TV Internet service provider and its users in order to take actions such as reducing the bit rate codification or selecting an alternative path.

Some measurement applications on the Internet estimates QoS parameters like delay or losses for any service, sending probe packets with a constant or an exponential interdeparture time. However QoS parameters depend on the statistical behavior of each service, so, in many cases, this type of estimation gives inaccurate estimators.

For the type of applications described above, one way to estimate the QoS parameters is to send during a sufficiently long time period a video or a set of voice conversations and to measure over these packets the selected parameters (mean delay, packet losses, etc.). This solution gives more accurate estimators, but it is unpractical because we are probably overloading the network during long time periods.

We are looking for a methodology to infer the QoS of a multimedia application without sending the multimedia service all the time.

One way to solve this problem could be to have a model of an Internet path with a known function (or functions) that allows to calculate the end-to-end QoS parameters given:

(1) a multimedia traffic model
(2) a cross traffic model
(3) the capacities and buffer sizes of the links along the path.

The first problem of this method is that looking the path from the ends, we do not know the cross traffic, and the links capacities and buffer sizes. Unfortunately, even in the case that we can have good estimations from the ends of the cross traffic, the link capacities and the buffer sizes along the path, there is not a general analytic model of an Internet path that gives us such function (or functions) to calculate the end-to-end QoS parameters.

Our approach to solve this problem is based on:

(1) the estimation of the cross traffic using a set of 'light' probe packets
(2) inferring or learning with some probe multimedia traffic the function that gives the QoS parameter of interest from an estimation of the cross traffic
(3) sending only 'light' probe packets in order to estimate the cross traffic and to predict the QoS of a multimedia application, without sending 'heavy' multimedia traffic during long time periods.

In this work we extend some recent theoretical results about functional nonparametric estimation and we apply these generalizations to solve the explained problem. In that sense this work is a first step in a research where there are many open issues.

In section 2 we resume some related works. In section 3 the main problems of this work and the proposed solutions are explained. Further, in section 4 we formalize some theoretical results that are necessary for our work. In chapter 5 the experimental methodology that we use to evaluate our results is explained and also some simulations are shown. Finally in section 6 we discuss the main conclusions and future research directions.

2. Related works

The main research topics on end-to-end Internet metrology are:

(1) Estimation of each link capacity in a network path or the capacity of the bottleneck link. The are many proposed procedures in order to estimate the link capacity. Each technique works better than the others depending on the constraints on the network and the cross traffic. In general, all techniques work fine if there is not cross traffic, but in a heavy loaded network or in a path with cross traffic in many links, the errors in the estimation can be large [11].
(2) Internet tomography, that is inferring some QoS parameter of a network interior link from estimations of the end-to-end value of these parameters [1] [2] [4] [6] [12] [13] [17]. These works are related with our problem, because they contribute with methodologies and ideas to measure some end-to-end parameters, but they do not solve it.
(3) Measure the link or path 'available bandwidth'(ABW). The ABW of a link i in the time interval $(t, t+\tau)$ is $A_i(t, t+\tau) = C_i(1 - u_i(t, t+\tau))$ where C_i is the link capacity and

$u_i(t, t+\tau)$ is the average link utilization in the time interval $(t, t+\tau)$. The minimum A_i in a path is defined as the ABW of the path. There are two main techniques to estimate the ABW. The first one sends a growing volume of probe traffic and analyzes the point where the probe traffic generates congestion in the path. There are different tools that use this methodology, for example Pathload [7] [8] and PathChirp [15]. The second technique is based on sending a packet pair or a packet train to measure the time dispersion at the end of the path. For example Spruce [16] sends packet pairs with interdeparture time D_{in} and measures at the end the interarrival time D_{out} of the packets of each pair. From this values and knowing the link capacity they estimate the ABW. Strauss et al. in [16] compare Spruce with other tools used to estimate the ABW, like Pathload.

In the context of our work, the estimation of the link capacity is important but it is not enough. In order to have a good QoS for a multimedia traffic, it is necessary but not sufficient condition to have enough capacity at the bottleneck link. If the link has enough capacity but the cross traffic at that link introduces delays, or jitter to the multimedia traffic the QoS could be poor. For this reason, we are interested on an estimation of the whole effect of the link characteristics and the cross traffic on a multimedia service. In this sense the works about the estimation of the ABW of a link or a path are more useful to solve our problem because the ABW take into account both effects. But the ABW does not give enough information to evaluate the QoS that a multimedia traffic on that path will receive because:

(1) the ABW gives information about the average over a time interval of the cross traffic, while the QoS of a multimedia traffic depends on all the statistics of the traffic that use the link or the path and not only on its average value.
(2) the ABW depends on the time scale selected for its average, and its value and variability depend on that time scale.
(3) with the information about the ABW, it is not possible to estimate the QoS that will be received by a multimedia traffic on that path. Jain and Dovrolis [9] have recently shown, measuring the ABW over different Internet paths, that this value experiments large variations over the time for each fixed time scale. They also show that these variations depend not only on the time scale but also on the cross traffic type.

3. Problem formulation and solution proposed

In this section we analyze a single link where cross traffic and probe traffic arrive. The cross traffic of this link, the link capacity, the buffer size, etc. are not known. The goal is to monitor the QoS that will experiment a multimedia traffic during long time periods.

We know that the performance metric Y (delay, jitter, losses, etc.) for packets of a multimedia traffic is a function $Y = \Phi(X_t, V_t, C, B)$, where X_t is the cross traffic stochastic process, V_t is the video or voice stochastic process, C is the link capacity, and B is the buffer size. We want to estimate Y without sending video or voice traffic during long time periods.

The link capacity C and the buffer size B are not known but it is assumed that are constants during the monitoring process. We want to evaluate the QoS of the process V_t so this is a known variable in our problem. Therefore, we can say that $Y = \Phi(X_t)$.

The first problem is that the cross traffic process X_t on the Internet is a dependent nonstationary process. This topic has been studied for many authors during the last years. Zhang et al. [18] [19] show that many processes on the Internet (losses for example) can be well modelled as i.i.d. within a 'change free region', where stationarity can be assumed. They describe the overall network behavior as a series of piecewise-stationary intervals. Karagiannis et al. [10] recently have found nonstationarity at different time scales analyzing the traffic of a link belonging to a Tier 1 ISP. They found that the traffic can been considered stationary at small time scales with events that change its stationarity at multi-second scale o larger.

The nonstationarity has different causes at different time scales and the 'stationary' time scale can be different for different paths and performance metrics. In all cases is very important to have measuring methodologies that can be used with nonstationary traffic. This is one of the main contributions of this paper. We will develop a measuring methodology that allows to estimate the QoS parameters in a network with nonstationary cross traffic.

In what follows we describe the procedure to estimate the function Φ. We divide the experiment in two phases:

(1) First, we send a burst of small probe packets (pp) of fixed size K spaced a fixed time t_{in}. Immediately after the burst we send during a short time a video stream. We repeat the previous procedure during some time, sending a new burst and a video probe after an interval time t_1 measuring from the previous end of the video stream. This can be seen in figure 1.

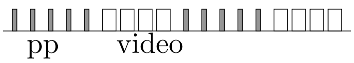

FIGURE 1

With the probe packets burst we infer the cross traffic of the link. We measure at the output of the link the interarrival time t_{out} between consecutive probe packets. This time series is strongly correlated with the cross traffic process that shares the link with the probe traffic. Using this cross traffic estimation and measuring the performance metric of interest over the video we will estimate the function Φ.

(2) In the second phase, sending only the probe packets in order to estimate the cross traffic and using the function $\widehat{\Phi}$ estimated in the first phase we estimate the performance of the QoS parameter $\widehat{Y}$.

One problem is to find the 'best' time scale t^*_{in} to be used as interdeparture time between consecutive probe packets. Later we will discuss further this topic, and for the moment we consider a fixed time scale.

At the output of the link we have a time series t_{out}. We compute the interarrival times and estimate its empirical distribution function $X^{t^*_{in}}$. We write $X^{t^*_{in}}$ to indicate the dependence on the time scale for interdeparture times between probe packets.

From each probe packet burst and video sequence j we have a pair (X_j, Y_j), where X_j is the empirical distribution function estimated from the time series t^j_{out} and Y_j is the performance metric of interest measured from the video stream j.

Therefore, our estimation problem has been transformed to the problem of inferring a function $\Phi : \mathcal{D} \to \mathbb{R}$ where $\mathcal{D}$ is the space of the probability distribution functions and $\mathbb{R}$ is the real line.

After we have obtained the estimation $\widehat{\Phi}$, we only need to send probe packets in order to estimate the empirical distribution $X^{t^*_{in}}$ and the performance metric of interest can be estimated from $\widehat{Y} = \widehat{\Phi}(X^{t^*_{in}})$.

To obtain and estimation of Φ from the pairs (X_j, Y_j) we will use a recent result about functional nonparametric estimation [5]. This result needs to be extended in order to be applied to our problem because in that work the authors suppose that the samples (X_j, Y_j) are equally distributed As we have explained before in the Internet the cross traffic is nonstationary, and for this reason our samples of cross traffic estimation will not be equally distributed. In the next section we will extend the results of [5] to a more general case that includes the nonstationary case.

To estimate the cross traffic suppose that the probe packets are separated a time t_{in} shorter enough to assure that for each two consecutive packets the second one is queued before the first one leaves the queue (see figure 2).

The time t^i_{out} measured at the output of the link between the packets i and $i+1$ of the burst is equal to $\frac{X_i}{C} + \frac{K}{C}$, where X_i is the amount of bits of cross traffic that arrived to the queue between probe packets i and $i+1$. The probe packets size K and the link capacity C are constants and K is small. Then the interarrival times are proportional to the cross traffic volume at least of a small constant.

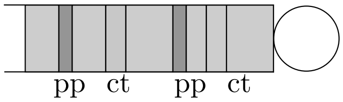

FIGURE 2

Suppose now that the packet $i+1$ is queued after the packet i leaves the queue. In this case as we are inferring the cross traffic volume from the times t^i_{out}, we are concluding that there is a cross traffic volume larger than the real one.

Therefore, it is necessary to send the packets at the highest rate as possible to be sure that for each two consecutive packets the second one is queued before the first one leaves the queue, but without overloading the path.

It is important to note that always between two packets separated a very short time we can have a very small cross traffic volume such that: $\frac{X_i}{C} + \frac{K}{C} < t^i_{in}$ and we are inferring during this time interval a cross traffic volume bigger than the real one. The question is if this small cross traffic volume is relevant or not for our estimation. For example for voice traffic, where consecutive packets are separated twenty miliseconds or more, if we have a cross traffic volume $\frac{X_i}{C}$ in the order of nanoseconds, clearly this traffic volume is not relevant for our estimation. The problem is to find a time scale t^*_{in} large enough to not overload the path but also small enough to give all the information about the cross traffic needed to our estimation problem.

To start, we send a burst of small packets with a time t_{in} smaller or equal than the time between packets of the traffic we are interested to measure (voice, video, etc.). Our aim is to estimate the performance metric Y using larger times between packets.

Our performance metric Y depends on the cross traffic process X at all time scales, but for each application it is not necessary to measure below a specific time scale, that depends on the statistical behavior of the traffic.

To find this time scale we can say that $Y = \Phi^{t_{in}}(X^{t_{in}}) + \varepsilon^{t_{in}}$ where t_{in} is the time scale for the interdeparture times between probes to estimate the interarrival times empirical distribution function $X^{t_{in}}$ and $\varepsilon^{t_{in}}$ is the estimation error by considering only t_{in}. The idea is to find a time scale t^*_{in} such that the estimation $\Phi^{t^*_{in}}$ gives the estimation of the parameter of interest with the minimum error, by minimizing the empirical variance of the estimation. We will explain in more detail in the following section how to calculate this time scale t^*_{in}.

4. Theoretical results

In this section we present a brief description of previous works on functional regression, and we especially summarize the results on functional regression of Ferraty et al. Our main theoretical results is the complete convergence of the estimator in the case of a nonstationary mixture of random variables. Finally we present our approach to the problem of time scales and how to choose an accurate one.

4.1. **Previous results.** Our approach to the problem of functional nonparametric regression is based on the work of Ferraty Goia and Vieu [5]. The model presented is a regression

$$Y = \Phi(X) + \varepsilon \tag{1}$$

where the regressor X is a function in a seminormed vector space with seminorm $|| \; ||$, the response Y is a real random variable and ε is a real, centered and independent of X random variable. Their estimator for Φ, obtained from a sequence of observations (X_i, Y_i), is a generalization of the Nadaraya-Watson Kernel estimator

$$\widehat{\Phi}_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{||x-X_i||}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{||x-X_i||}{h_n}\right)} = \frac{\sum_{i=1}^{n} Y_i K_n(X_i)}{\sum_{i=1}^{n} K_n(X_i)} \tag{2}$$

In [5] the authors proved the complete convergence of the estimator, the rate of convergence and the uniform complete convergence, when the observation is a sequence (X_i, Y_i) of stationary weakly dependent (α-mixing) random variables. One of the main topics discussed in [5] is the problem of finding good estimators when the observations come from an infinite dimensional vector space, where finding enough samples near x is crucial. The same problem in a different context is also treated in the work of Cuevas, Febrero, and Fraiman [3]. Ferraty, Goia and Vieu introduce the fractal dimension of the random variable X, and the convergence results depend on this dimension.

4.2. **Nonstationary mixture model.** We will consider a nonstationary model where X is a mixture of stationary variables. Consider the regression model

$$Y = \Phi(X) + \varepsilon \tag{3}$$

but instead of having the random variables X equally distributed let

$$X_i = \varphi(\xi_i, Z_i) \tag{4}$$

where ξ_i takes values in a seminormed vector space with a seminorm $\|\ \|$, and Z_i is a real random variable that takes values in a finite set $\{z_1, z_2, \dots, z_m\}$. The sequence $\varphi(\xi_i, z_k)$ is weakly dependent and equally distributed for every $1 \leqslant k \leqslant m$, but the sequence Z_i may be nonstationary as in the work of Perera [14]. Under these hypotheses in this section we will obtain the asymptotic behavior of the estimator defined by (2). Our proof is a generalization of the proof of Theorem 4.1 in [5].

We assume some mild hypotheses for the kernel, the mixing coefficients of the sequence (X_i, Y_i), and the distribution of Y that are the same of Ferraty et al. [5] and for the sequence X_i we assume different hypotheses as follows.

There exist the conditional moments

$$E\left\{\left|\Phi\left(\varphi(\xi_i, z_k)\right), \Phi\left(\varphi(\xi_j, z_l)\right)\right|\right\} = h(|i-j|) < \infty \tag{5}$$

For each $1 \leqslant k \leqslant m$ exists $\delta^k(x) > 0$ such that

$$\lim_{\alpha \to 0} \frac{1}{\alpha^{\delta^k(x)}} P\left(\varphi(\xi_i, z_k) \in B(x, \alpha)\right) = c^k(x) > 0 \forall i \tag{6}$$

For each $1 \leqslant k \leqslant m$ and $\forall i \neq j$ exists $\delta^{kl}_{|i-j|}(x)$ such that

$$\lim_{\alpha \to 0} \frac{1}{\alpha^{\delta^{kl}_{|i-j|}(x)}} P\left(\varphi(\xi_i, z_k), \varphi(\xi_j, z_l) \in B(x, \alpha) \times B(x, \alpha)\right) = c^{kl}_{|i-j|}(x) > 0 \tag{7}$$

The bandwidth h_n is a positive sequence such that

$$\lim_{n \to \infty} h_n = 0, \quad \lim_{n \to \infty} \frac{n h_n^{\delta(x)}}{\log n} = \infty, \quad \delta(x) = \min\{\delta^k(x) : 1 \leqslant k \leqslant m\} \tag{8}$$

For each $1 \leqslant k \leqslant m$ exists $p_k \geqslant 0$ such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} P(Z_i = z_k) = p_k \tag{9}$$

For each $1 \leqslant k, l \leqslant m$, $h \geqslant l$ exists $p^h_{kl} \geqslant 0$ such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} P(Z_i = z_k, Z_{i+h} = z_l) = p^h_{kl} \tag{10}$$

Remark. Hypothesis (6) is about the distribution of each component of the mixture and (7) is about the joint distribution. As it is noticed in [5] $\delta(x)$ is the analogous to the dimension d when the random variable $X \in \mathbb{R}^d$. In the case of the mixture, if $\delta^k(x)$ is different for each k, for the convergence theorems the component of the mixture that gives the properties of the estimator is $\delta(x) = \min\{\delta^k(x) : 1 \leqslant k \leqslant m\}$. Then, the properties of the estimator depend on the component of the mixture that has more samples in a neighborhood of x.

Remark. Hypothesis (9) means that the sequence Z_i only verifies stationarity conditions in average, and (10) implies stationarity in average for the joint distribution. (9) is a very general hypothesis, a counterexample can be constructed as in example 2.2 in [14].

In order to prove the main result that is the complete convergence of the estimator $\widehat{\Phi}_n(x)$ to $\Phi(x)$ we will prove some previous lemmas, analogous to lemmas 4.1, 4.2 and 4.3 of [5] under our hypotheses. Let us write

$$\widehat{\Phi}_n(x) = \frac{g_n(x)}{f_n(x)} \tag{11}$$

with $g_n(x) = \frac{1}{nh_n^{\delta(x)}} \sum_{i=1}^{n} Y_i K_n(X_i)$ and $f_n(x) = \frac{1}{nh_n^{\delta(x)}} \sum_{i=1}^{n} K_n(X_i)$

Lemma 1. *Under general hypotheses about the kernel, (6), (8) and (9)*

$$\lim_{n\to\infty} E\left(f_n(x)\right) = C_\delta(x) = \sum_{k=1}^{m} p_k C_{\delta^k}(x) 1_{\{\delta^k(x)=\delta(x)\}} \tag{12}$$

where $1_{\{\ \}}$ *is the indicator function and* $C_{\delta^k}(x) = c^k(x)\delta^k(x)\int_0^\theta K(v)v^{\delta^k(x)-1}dv$

Further, if Φ *is a continuous function then*

$$\lim_{n\to\infty} E\left(g_n(n)\right) = \Phi(x)C_\delta(x) \tag{13}$$

Proof. Lemma 4.1 of [5] assures that for each $1 \leqslant k \leqslant m$

$$\lim_{n\to\infty} \frac{1}{h_n^{\delta^k(x)}} E\left\{K_n\left(\varphi(\xi, z_k)\right)\right\} = C_{\delta^k}(x) \tag{14}$$

$$E\left(f_n(x)\right) = \frac{1}{nh_n^{\delta(x)}} \sum_{i=1}^{n} E\left(K_n\left(X_i\right)\right)$$

$$E\left(K_n\left(X_i\right)\right) = E\left\{E\left(K_n\left(X_i\right)|Z_i\right)\right\} = \sum_{k=1}^{m} E\left\{K_n\left(\varphi(\xi_i, z_k)\right)\right\} P\left(Z_i = z_k\right)$$

As $E\left\{K_n\left(\varphi(\xi_i, z_k)\right)\right\}$ does not depends on i

$$E\left(f_n(x)\right) = \sum_{k=1}^{m} \left(\frac{1}{h_n^{\delta(x)}} E\left\{K_n\left(\varphi(\xi, z_k)\right)\right\} \frac{1}{n} \sum_{i=1}^{n} P\left(Z_i = z_k\right) \right)$$

We have that

$$\lim_{n\to\infty} \frac{1}{h_n^{\delta(x)}} E\left\{K\left(\varphi(\xi, z_k)\right)\right\} = \begin{cases} C_{\delta^k}(x) & \text{if } \delta^k(x) = \delta(x) \\ 0 & \text{if } \delta^k(x) < \delta(x) \end{cases} \tag{15}$$

and using (9) and (14)

$$\lim_{n\to\infty} E\left(f_n(x)\right) = \sum_{k=1}^{m} p_k C_{\delta^k}(x) 1_{\{\delta^k(x)=\delta(x)\}} = C_\delta(x)$$

To prove (13) lemma 4.2 of [5] assures that

$$\lim_{n\to\infty} \frac{1}{h_n^{\delta^k(x)}} E\left\{\Phi\left(\varphi(\xi, z_k)\right) K_n\left(\varphi(\xi, z_k)\right)\right\} = \Phi(x) C_{\delta^k}(x) \tag{16}$$

We have

$$E\left(g_n(x)\right) = E\left(\frac{1}{nh_n^{\delta(x)}} \sum_{i=1}^{n} Y_i K_n\left(X_i\right) \right) = \frac{1}{nh_n^{\delta(x)}} \sum_{i=1}^{n} E\left(Y_i K_n\left(X_i\right)\right)$$

Then taking

$$E\left(Y_i K_n\left(X_i\right)\right) = E\left(\Phi(X_i) K_n\left(X_i\right)\right) = E\left\{E\left(\Phi(X_i) K_n\left(X_i\right)|Z_i\right)\right\}$$

with calculus analogous to the previous ones

$$E\left(g_n(x)\right) = \sum_{k=1}^{m} \frac{1}{h_n^{\delta(x)}} E\left\{\Phi(\varphi(\xi, z_k)) K_n\left(\varphi(\xi, z_k)\right)\right\} \frac{1}{n} \sum_{i=1}^{n} P\left(Z_i = z_k\right)$$

From (9) and (16)$E\left(g_n(x)\right)$ converges to $\Phi(x) C_{\delta(x)}$ □

Lemma 2. *Under hypotheses about the kernel, the mixing coefficients, the distribution of Y and (5) to (10), with general assumptions for the relationship between $\delta(x)$, h_n and conditions on the mixing coefficients and the distribution of Y, there exists $\varepsilon > 0$ such that*

$$\sum_{i=1}^{\infty} P\left(\left|f_n(x) - E(f_n(x))\right| > \varepsilon \sqrt{\frac{\log n}{n h_n^{\delta(x)}}}\right) < \infty$$

$$\sum_{i=1}^{\infty} P\left(\left|g_n(x) - E(g_n(x))\right| > \varepsilon \sqrt{\frac{\log n}{n h_n^{\delta(x)}}}\right) < \infty$$

Proof. This lemma is the analogous of lemma 4.3 of [5]. The proof is based on computing $s_n^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} |Cov(\Delta_i, \Delta_j)|$ with $\Delta_i = K_n(X_i) - E\left(K_n(X_i)\right)$ and proving that there exists $\varepsilon > 0$ such that $\sum_{i=1}^{\infty} P\left(s_n^2 > \varepsilon n h_n^{\delta(x)}\right) < \infty$. The rate $n h_n^{\delta(x)}$ is obtained by imposing some relationships between a, p, $\delta(x)$ and h_n. In our case the thesis is obtained by assuming the extra hypothesis (10) that allow us to make the same calculus as in [5] but conditional to Z as in the previous lemma, and to prove the thesis for the stationary sequences $\varphi(\xi, z_k)$ and then taking the limit of (10). □

The main result is the complete convergence stated in the next theorem.

Theorem 1. *Under the assumptions of lemma (2), for Φ continuous, $\widehat{\Phi}_n(x)$ converges completely to $\Phi(x)$.*

Proof. The proof of this theorem is the same as in [5], based on lemmas 1 and 2. □

With this theorem we assure that we can compute the function $\widehat{\Phi}_n$ with n samples of the sequence (X_i, Y_i) and then use this function to predict the QoS based only on new samples X_i. In what follows we will address the problem of time scales for the samples X_i.

4.3. **Choosing the time scale.** In our experiment the probe traffic is sent with fixed time t between consecutive probe packets. The aim is to find some criterion for choosing the best time scale in order to infer as accurately as possible the performance metric Y. We consider different sequences of observations for a finite number of time scales $\{t_1, t_2, \ldots, t_r\}$. In practice, as we send bursts of probe traffic with fixed time t between packets we will have observations with time scales in the set $\{t, 2t, \ldots, rt\}$. Consider $n+m$ observations for each time scale

$$\left\{(X_i^{t_j}, Y_i^{t_j}) : 1 \leqslant i \leqslant n+m,\ 1 \leqslant j \leqslant r\right\}$$

By dividing the sequence for a fixed time scale in two we can estimate the function Φ^{t_j} (for the time scale t_j) by $\widehat{\Phi}_n^{t_j}$ with the first n samples. With the remaining data we compute the difference

$$\sigma^2{}_{t_j}(n,m) = \frac{1}{m} \sum_{i=1}^{m} \left(\widehat{\Phi}_n^{t_j}(X_{n+i}^{t_j}) - Y_{n+i}^{t_j}\right)^2$$

that gives a measure of how good is the estimator at time scale t_j. If we compute $\sigma^2{}_{t_j}(n,m)$ for $1 \leqslant j \leqslant r$ we could choose $t_{n,m}^*$ such that

$$\sigma^2{}_{t_{n,m}^*}(n,m) = \min\{\sigma^2{}_{t_j}(n,m) : 1 \leqslant j \leqslant r\}$$

Hereafter we choose this time scale for the measurement methodology. $\sigma^2{}_t(n,m)$ converges a.s. to $Var(\varepsilon^t)$ when $n, m \to \infty$. Then, if there exists an optimum time scale t^* such that $Var(\varepsilon^{t^*})$ is minimum, and the estimator $\widehat{\Phi}_n$ has good properties of convergence, the estimator $t_{n,m}^*$ converges to t^*.

5. Simulations

We have simulated cross traffic using the function $X = \varphi(\xi, Z)$, where each ξ is equally distributed in one of two sets as an ON-OFF Markovian traffic and Z is a random variable that selects periodically between this two sets.

The first set (SET 0) is a set of Markovian ON-OFF traffic that has average bit rate varying from 150 Mb/s to 450 Mbs and average time Ton in the ON state and Toff in the OFF state Toff varying from 100 to 300 ms.

The second set (SET 1) of Markovian ON-OFF traffic has average bit rate varying from 600 Mb/s to 900 Mb/s and average time Ton in the ON state and Toff in the OFF state varying from 200 to 500 ms.

After we have selected for the next period the SET 0 or SET 1, an independent random variable is sampled to select the average bit rate, Ton and Toff.

We send this cross traffic to a network link. We also send to that link the probe traffic that is composed of a number of tests. Each test has a burst with fixed interdeparture time t_{in}^*, and after it a simulated multimedia traffic. For each test j we estimate the empirical distribution function X_j of the cross traffic at the time scale t_{in}^*.

With the simulated multimedia traffic we measure for each test j the average delay Y_j of this traffic.

We generated 51 tests. We use 50 tests to estimate the function $\widehat{\Phi}$ and we leave one test out to verify our estimation. The kernel used to estimate Φ is $K(x) = \begin{cases} (x^2-1)^2 & x \in [-1,1] \\ 0 & x \notin [-1,1] \end{cases}$ and we use the L^1 norm for the distance between the empirical distribution functions. The selection of the Kernel and the bandwidth must be studied in more detail in the future.

In figure 3 the estimated and the measured value for the average delay and the relative error of the estimated and the measured value are plotted. Each value corresponds to a estimation for this point taking as training sample all the samples without this value. The results are relatively accurate for that small amount of tests.

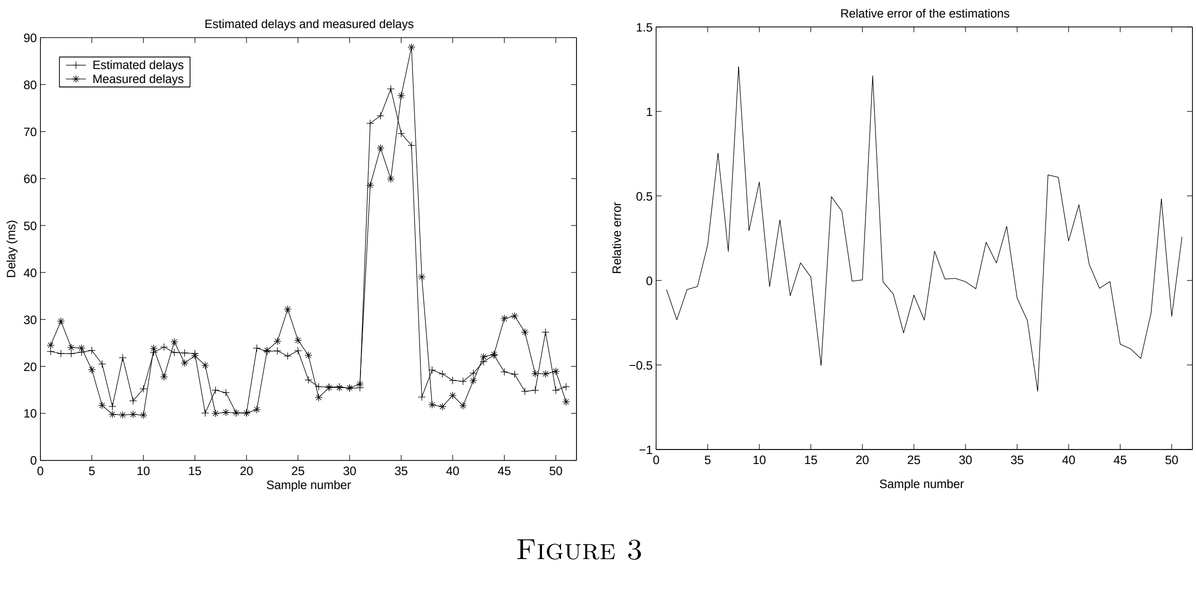

Figure 3

6. Conclusions

In this work we have developed a methodology to estimate the end-to-end QoS of a multimedia service. This methodology allows to monitor the QoS that a multimedia application along an Internet path during long periods of time would receive. We have extended some recent results on functional nonparametric regression that enable us to apply our measuring methodology to nonstationary traffic. This is very important for practical applications because the Internet has nonstationary traffic as has been shown by many works in last years. We tested our methodology with simulated nonstationary traffic and we obtained good accurate estimations. There are many open issues in this research. Our methodology must be tested with real Internet traffic and can be extended to other QoS estimation problems. We will also work in other two directions: to find confidence intervals for the estimators and to study the extension of the results that are valid for weak dependence to other dependence models.

References

[1] Adler M., Bu T., Sitaraman R. and Towsley D. (2001). Tree Layout for Internal Network Characterizations in Multicast Networks. *Proceedings of 3rd International Workshop on Networked Group Communication.*

[2] Cáceres R., Duffield N.G., Horowitz, J. and Towsley, D. (1999). Multicast-based inference of network-internal loss characteristics. *IEEE Transactions on Information Theory* Vol. 45, pp. 2462-2480.

[3] Cuevas, H., Febrero, M. and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *Canadian Journal of Statistics* Vol. 30, No. 2, pp. 285-300.

[4] Downey A. Using pathchar to Estimate Internet Link Characteristics. *ACM SIGCOMM.*

[5] Ferraty, F., Goia, A. and Vieu P. (2002). Functional nonparametric model for time series: a fractal approach for dimension reduction. *Test* Vol. 11, No. 2, pp. 317-344.

[6] Jacobson V.(1997). Pathchar- a tool to infer characteristics of Internet path. [ftp://ee.lbl.gov/pathchar/, 1997].

[7] M. Jain and C. Dovrolis (2002). Pathload: A Measurement Tool for End-to-End Available Bandwidth, *Proceedings of Passive and Active Measurements (PAM) 2002* pp. 14-25.

[8] M. Jain and C. Dovrolis (2003) End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput. *IEEE/ACM Transactions in Networking.*

[9] M. Jain and C. Dovrolis (2005) End-to-End Estimation of the Available Bandwidth Variation Range. To appear in the *Proceedings of ACM SIGMETRICS Conference*, June 2005, Banff, Canada.

[10] Karagiannis T., Molle M., Faloutsos M. and Broido A. (2004) A Nonstationary Poisson View of Internet Traffic. *IEEE INFOCOM 2004.*

[11] Lai K. and Baker M. (2001). Nettimer: A Tool for Measuring Bottleneck Link Bandwidth. *Proceedings of the USENIX Symposium on Internet Technologies and Systems.*

[12] Lawrence, E., Michailidis, G., and Nair, V. N. (2003). Inference of Network Delay Distributions Using the EM Algorithm. *Technical Report, University of Michigan.*

[13] Lo Presti F., Duffield N. G., Horowitz J. and Towsley D. (2002). Multicast-Based Inference of Network-Internal Delay Distributions. *ACM/IEEE Transactions on Networking* Vol 10, pp. 761-775.

[14] Perera, G. (2001). Random fields on $\mathbb{Z}^d$, limit theorems and irregular sets. *Spatial statistics: methoodlogical aspects and applications.* Lecture notes in statistics, Springer.

[15] Ribeiro V., Riedi R., Baraniuk R., J. Navratil J. and Cotrell L. (2003). PathChirp: Efficient Available Bandwidth Estimation for Network Paths. *Passive and Active Measurement Workshop, 2003.*

[16] Strauss J., Katabi D., and Kaashoek F. (2003). A measurement study of available bandwidth estimation tools. *Internet Measurement Workshop, Proceedings of the 2003 ACM SIGCOMM conference on Internet measurement.* pp. 39-44.

[17] Tsang Y. and Coates M. and Nowak R. (2003) Network Delay Tomography. *IEEE Transactions on Signal Processing.* Vol. 51, pp. 2125-2136.

[18] Zhang Y., Paxson V. and Shenker S. (2000). The Stationarity of Internet Path Properties: Routing, Loss, and Throughput *ACIRI Technical Report.*

[19] Zhang Y., Duffield N., Paxson V. and Shenker S. (2001). On The Constancy of Internet Path Properties. *ACM SIGCOMM Internet Measurement Workshop.*

ARTES (Análisis de Redes, Tráfico y Estadística de Servicios)
Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
Contact: *artes@fing.edu.uy.*