Telecommun Syst (2010) 43: 265–277
DOI 10.1007/s11235-009-9201-x

A new structure-preserving method of sampling for predicting self-similar traffic

Halima Elbiaze · Mohamed Faten Zhani · Omar Cherkaoui · Farouk Kamoun

Published online: 17 October 2009

Abstract The paper presents a *structure-preserving* method of sampling *self-similar* traffic with an application to network monitoring and resource provisioning. Based on the observation of the self-similarity of Internet traffic, we propose a new sampling technique (so-called *the maximum-based sampling*). We show that the resulting data suits perfectly for predicting the bandwidth required by upcoming traffic so that the resource provisioning can be done efficiently and intelligently especially for the context of IP over WDM networks.

Hence, we prove mathematically that the proposed technique preserves the self-similarity of the traffic. Besides, experimental results using real Internet traffic show that unlike other sampling techniques (*systematic sampling* and *stratified random sampling*), the maximum-based sampling capture faithfully the traffic self-similarity. In order to assess the effect of the sampling technique impact on the performance of the traffic prediction,we undertake a series of prediction experiments using sampled traffic with the proposed technique and the other sampling techniques. A *neurofuzzy model* (α_SNF), *the AutoRegressive Integrated Moving Average model* (ARIMA) and *the Linear Minimum Mean Square Error* (LMMSE) are considered in this study for bandwidth prediction. Our experiments results show that the *maximum-based* sampled traffic—used for the identification of the prediction model—is the most suitable for predicting the traffic for different time scales.

Keywords Self-similarity · Traffic analysis and prediction · Sampling · Network monitoring · Resource provisioning · IP over WDM

H. Elbiaze · M.F. Zhani (✉) · O. Cherkaoui
Department of Computer Science, University of Québec in Montréal, C.P. 8888, Succursale centre-ville Montréal, QC H3C 3P8, Canada
e-mail: zhani.mohamed_faten@courrier.uqam.ca

H. Elbiaze
e-mail: elbiaze.halima@uqam.ca

O. Cherkaoui
e-mail: cherkaoui.omar@uqam.ca

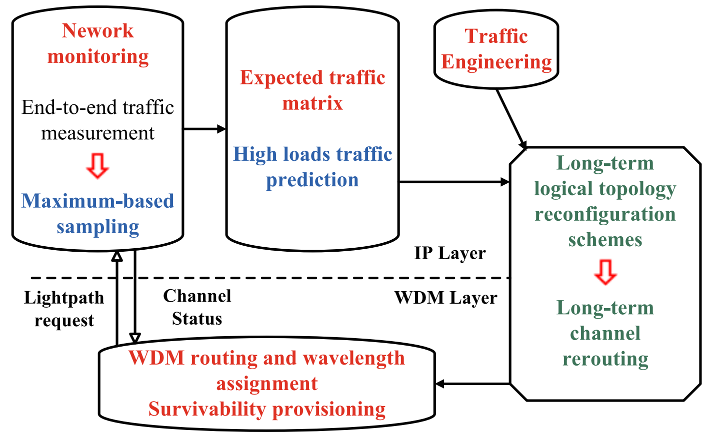
F. Kamoun
National School of Computer Sciences, Manouba 4010, Tunisia
e-mail: frk.kamoun@planet.tn

1 Introduction

A large number of researches was devoted to the design of Internet traffic models [3, 19] in LAN and WAN networks. However, it is still not clear what models or pertinent metrics to use when characterizing Internet traffic. Three main factors amplified this phenomena (i) the overstated reports on Internet traffic exponential growth, (ii) the complexity of Internet traffic measurements (iii) and the inability to find the exact applicability—resulted from the lack of understanding—of those recently discovered peculiar traffic phenomena such as self-similarity. In this context a lot of work has investigated techniques able to improve the understanding and the exploitation of the Internet traffic. Thus, traffic sampling and prediction techniques have constituted an important research topics.

The foremost and fundamental question regarding sampling is its accuracy. In other words sampled traffic must reflect faithfully the characteristics of the original Internet traffic. This is especially pertinent due to the existence of concentrated periods of hight activities (peaks) and low activities i.e. burstiness. The self-similarity nature of the traffic is also widely accepted since the Bellcore studies [19].

Fig. 1 Interaction between IP layer and WDM layer



Hence, inaccurate sampling can lead to wrong analysis and interpretation of the traffic. Consequently, it could involve wrong decisions by the network operators.

Several research efforts have been made to investigate the effectiveness of sampling techniques in measuring network traffic [5, 10, 13, 21, 25, 27]. Three commonly used sampling techniques, i.e., static systematic, stratified random and simple random, have been studied by Claffy et al. [5]. In a latest work, He and Hou [10] have shown that while all of these three sampling techniques can capture the Hurst parameter (second order statistics) of Internet traffic they fail to capture the mean (first order statistics) faithfully. Based on the observation of the traffic—especially its self-similarity—they have also proposed a new variation of static sampling called *biased systematic sampling* (BSS). The authors showed that their technique gives better performance in terms of efficiency compared to other sampling techniques. Unfortunately, they didn't use BSS results for real applications such as traffic prediction. In fact, we believe that this sampling technique generates new parameters which are not obvious to identify such as threshold, number of samples…etc. Besides, the resulting samples will be spaced of various time interval. Thus, the application of prediction models on such sampled process is not straightforward.

On the other hand, Traffic prediction has been more investigated since the acceptance of the self-similar and the long-range dependence nature of networks traffic [1, 15, 18, 19]. While this peculiar characteristic causes dramatic effects on network performance in terms of loss and delay, several studies have shown that the self-similarity can be exploited to characterize or to predict the traffic in order to control network resources [2, 9, 11, 12, 17]. A lot of prediction models has been used like *the AutoRegressive Integrated Moving Average* family models (AR, MA, ARMA, ARIMA), *the Linear Minimum Mean Square Error model* (LMMSE) and the Neurofuzzy models like the α_SNF [9, 11, 12, 17].

Traffic sampling and prediction may be useful for developing new techniques able to improve resources utilization in a more flexible and dynamic way. This could be interesting in the context of *wavelength division multiplexing* (WDM). Accordingly, a crucial issue for new generation optical backbone networks is the achievement of traffic engineering strategies that support different types of traffic with dynamic demands. Traffic engineering should select routes for new-established connexions and may redirect traffic flows toward less congested paths, taking into account the different traffic loads and the network state (Fig. 1). Instead of today's static service delivery with manual provisioning and long-term bandwidth allocation, service providers may use intelligent capacity planning to optimize bandwidth provisioning [6].

End-to-end measurements performed on a network infrastructure is a widely used technique for resource management protocols design as well as for traffic behavior analysis. Those measurement results can be used for better resource control [4, 8, 16, 17, 23].

In this paper, we focus on traffic sampling and prediction at IP level to make decisions for resource provisioning at the optical level. In other words, the target is to propose a mechanism of traffic engineering able to dynamically cope with the traffic variation and assign the exact amount of bandwidth (lightpaths) required to handle it.

To achieve this purpose, we derive a new structure-preserving method of sampling. We prove its efficiency for preserving traffic characteristics. We then propose a prediction-based method to detect upcoming high traffic loads using the sampled data. We also focus on finding adequate prediction model by comparing the α_SNF model, the ARIMA model and the LMMSE model.

We note that experiments were done with real traces consisting of the aggregated traffic of a single link. However, the proposed scheme can be applied to the load placed on one lambda channel in the case of an IP/WDM network.

The remaining of the paper is organized as follows. Section 2 introduces the self-similarity in the network traffic.

It describes sampling techniques and the prediction models considered in this work. Section 3 presents the structure preserving sampling method as well as the mathematical proofs demonstrating the self-similarity of the sampled traffic. Our proposal is validated in Sect. 4 by a comparison with other sampling techniques and by prediction experiments. Section 5 concludes with a summary of the obtained results. It also provides a look to the possible applications of the derived technique.

2 Background

In this section, we introduce self-similar processes, the commonly used sampling techniques as well as the prediction models used in this work. We also discuss the criteria that we must take into account to choose the suitable sampling scheme.

2.1 Self-similarity

Last decade's studies on network traffic argue convincingly that LAN and WAN traffic is much better modeled using self-similar processes [15, 19]. The strength of self-similar models is that they are able to incorporate *Long-Range Dependence* (LRD), which informally means significant correlations across arbitrarily large time scales. The self-similar traffic can have serious adverse impact on network performance [18]. While throughput declines gradually as self-similarity increases, queuing delay increases more drastically. In addition, when the traffic is highly self-similar, the queuing delay grows nearly proportionally to the buffer capacity present in the system. However, the non-trivial correlation structure present in LRD traffic at large time scales can be judiciously exploited for accurate traffic prediction.

We present now the most important formulas to our purpose in order to characterize self-similar processes. Given a zero-mean, stationary time series $X = (X_t; t = 1, 2, 3, \ldots)$, we define the m-aggregated series $X^{(m)} = (X_k^{(m)}; t = 1, 2, 3, \ldots)$ by summing the original series X over non-overlapping blocks of size m.

$$X^m(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i). \tag{1}$$

Then we say that X is H-self-similar if for all positive m, X^m has the same distribution as X rescaled by m^H. That is:

$$X_t \cong m^{-H} \sum_{i=(t-1)m+1}^{tm} X_i \quad \text{for all } m \in N. \tag{2}$$

If X is H-self-similar, it has the same autocorrelation function $r(k) = E[(X_t - \mu)(X_{t+k} - \mu)]/\sigma^2$ as the series $X^{(m)}$ for all m.

As result, self-similar processes can show *long-range dependence*. A process with long-range dependence has an autocorrelation function $r(k) \sim k^{-\beta}$ as $k \to \infty$, where $0 < \beta < 1$. Thus the autocorrelation function of such process follows a power law. Since $\beta < 1$, the sum of the autocorrelation values of such series approaches infinity.

This mathematical characteristic has a number of implications. First, the variance of the mean of n samples from such series does not decrease proportionally to $n^{-\beta}$. Second, the power spectrum of such a series is hyperbolic, rising to infinity at frequency zero-reflecting the "infinite" influence of long-range dependence in data.

One of the attractive features of using self-similar models for network traffic, is that the degree of self-similarity of such series is expressed using only a single parameter. This parameter expresses the speed of decay of the autocorrelation function, and it is called *Hurst* parameter $H = 1 - \frac{\beta}{2}$. Thus, for self-similar series with long-range dependence, $\frac{1}{2} < H < 1$. As $H \to 1$, the degree of both self-similarity and long-range dependence increases.

2.2 Sampling techniques

Three categories of sampling techniques have been commonly used in measuring Internet traffic [5, 10]: systematic sampling, stratified random sampling, and simple random sampling (Fig. 2).

- In systematic sampling, every Cth element of the parent process is deterministically selected for sampling, starting from sampling point.
- In stratified random sampling, the time axis is divided into intervals of length C, and one sample is randomly selected each interval.
- In sample random sampling, N packets are randomly selected from the entire population.

Once we have selected our sampling scheme, it would seem to be a rather straightforward exercise to take some measurements, calculate some statistics and draw conclusions. There are, however, many things which can go wrong along the way that can be avoided with careful planning and knowing what to watch for. Thus sampling technique must take into account various criteria:

- Preserving Self-similarity: as mentioned in the previous section the Hurst parameter reflects the degree of both self-similarity and long-range dependence. Thus, sampling should preserve the Hurst parameter of the original data.
- Preserving Burstiness: Burstiness behavior in actual traffic creates difficulties for many conventional measures of

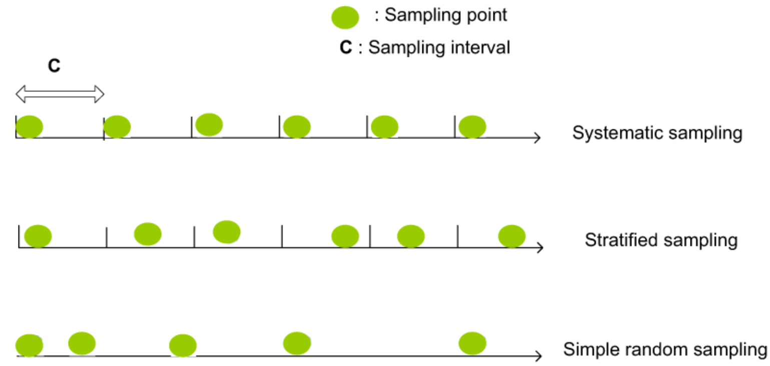


Fig. 2 An illustration of the three sampling techniques [10]

"burstiness". In some previous works the Hurst parameter H was considered as a metric for measuring traffic burstiness and seemed to capture the intuitive notion of burstiness. This consideration was justified by visual assessment and the assumption that the traffic behaves like fractional Gaussian noise [14]. Therefore we use another criterion for "burstiness" assessment. Three commonly used definitions are the ratio of peak bandwidth to mean bandwidth, the coefficient of variation, and the index of dispersion for counts (IDC) [15]. In this paper, we consider the Peak-to-Mean (PM) as a metric to measure traffic burstiness expressed by the following equation:

$$\mathrm{PM} = E\left[\frac{\max_{i=(k-1)m+1}^{i=km}(X_i)}{\sum_{j=(k-1)m+1}^{j=km} X_j/m}\right]. \tag{3}$$

– Precision of the prediction: since we aim to predict traffic for better resource control, sampling technique should help to have an accurate prediction.

In this paper, the proposed sampling technique takes into account all these constraints. It doesn't have to preserve the average or the variance over the data since the final aim is to predict burstiness. We prove that it preserves the self-similarity, the burstiness and the high variability of the traffic. It also allows better prediction performance as compared to the other sampling techniques.

2.3 Prediction models

In what follows, we introduce prediction models considered in this study. The first model is the α_SNF which is a neurofuzzy model. It combines fuzzy logic and neural networks. This model has given better performance and more accurate results than linear predictors [22, 28, 29]. The second model is the Autoregressive Moving Integrated Average which is a linear model. This model was widely used in literature [17, 20]. The third model is the Linear Minimum Square Error which was also used in previous works [9, 11, 12]. We also note that ARIMA and α_SNF are models which need a training phase. The training phase is a phase for the identification of the model parameters. However, LMMSE needs only the N last observations to predict the next value.

The criteria used to evaluate the predictability and to compare the efficiency of used models is the Average Relative Variance:

$$\mathrm{ARV} = \frac{\sum_{i=1}^{n}[x_t - \hat{x}_t]^2}{\sum_{i=1}^{n}[x_t - \mu]^2}, \tag{4}$$

where x_t is the real output, $\hat{x}_t$ is the calculated output, μ is the estimated average over the used data. The advantage of the ARV is that it doest depend on the used scale or the size of data unlike the mean square error (MSE). If the ARV is less than 1, it means that the predictor is doing better than using the average. Thus when the ratio is getting smaller than 1, it is more advantageous to use the predictor.

2.3.1 *The neurofuzzy system (α_SNF)*

The first model is a neurofuzzy model called α_SNF [22, 28, 29]. The fuzzy system is described as a non-linear relation between inputs $x_1, \ldots, x_n$ and an output $Y = f(x_1, \ldots, x_n)$, where n is the number of inputs x_i. This relation is described by a collection of fuzzy rules. Let c be the number of rules in the fuzzy system. We note R_k the kth rule where $1 \le k \le c$. A fuzzy rule R_k is given as the following:

$$R_k: \text{ if } (x_1, \ldots, x_n) \text{ is } A_k \text{ then } y_k \text{ is } b_k, \tag{5}$$

where A_k is called a cluster and y_k is the output of the rule calculated using a real noted b_k.

In fuzzy logic, every point x belongs to a cluster A with a membership degree that has a value between 0 and 1 given by a membership function $\mu_A(x)$. Thus, each rule R^k evaluates the membership of each element $(x_1, \ldots, x_n)$ to each cluster A_k noted $\mu_{A_k}(x_1, \ldots, x_n)$. Then y_k is calculated as:

$$y_k = \mu_{A_k}(x_1, \ldots, x_n).b_k. \tag{6}$$

The rule R_k can be written as:

$$x_1 \text{ is } A_{k1} \text{ and } x_j \text{ is } A_{kj}, \ldots,\ x_n \text{ is } A_{kn} \text{ then } y_k \text{ is } b_k, \tag{7}$$

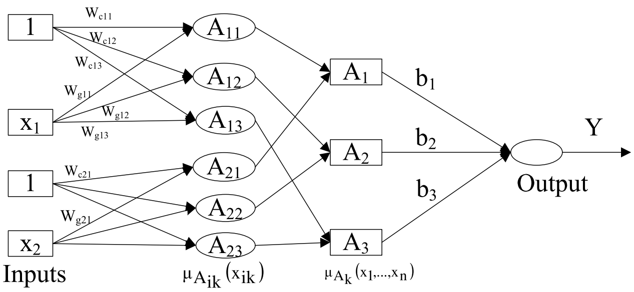


Fig. 3 Example of equivalent neural network α_SNF (2 inputs, 3 rules)

where the cluster A_{ki} is the projection of A_k in the ith dimension. We note $\mu_{A_{ik}}(x_i)$ the membership function of x_i to the cluster A_{ki}. Then $\mu_{A_k}(x_1, \ldots, x_n)$ is given by:

$$\mu_{A_k}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \mu_{A_{ik}}(x_i). \tag{8}$$

We used membership function:

$$\mu_{A_{ik}}(x_i) = \exp\bigl(-|w_{gik}x_i + w_{cik}|^{l_{ik}}\bigr), \tag{9}$$

where w_{gik}, w_{cik} and l_{ik} parameters are used to adjust the general form of the function.

The output of the system Y is given by:

$$Y = \frac{\sum_{k=1}^{c} y_k}{\sum_{k=1}^{c} \mu_{A_k}(x_1, \ldots, x_n)}. \tag{10}$$

The parameters w_{gik}, w_{cik} and l_{ik} are initialized using a method called "*semi-α-cut density*" [22]. The fuzzy model is then incorporated into an equivalent neural network. Figure 3 shows the α_SNF using 3 rules and 2 inputs. Each node A_{ik} calculates the membership function $\mu_{A_{ik}}(x_i)$ using (9) and each node A_k calculates y_k using (6). The output node calculates Y using (10).

The α_SNF model is trained using the back-propagation algorithm [22]. Training the neural network aims at changing the parameters w_{gik}, w_{cik} and l_{ik} in order to reduce the error between the calculated output and the real output. The input variables x_i can be either a lag $y(t-i)$ or another variable measured at time $t-1$. For selecting the number of rules, we found via experiments that using more than 3 rules does not improve the prediction performance. We found that using more than 3 rules does not improve the prediction performance.

2.3.2 *The autoregressive integrated moving average model (ARIMA)*

The most well-known linear forecasting models are the *Autoregressive* (AR), *Moving Average* (MA) and the *the AutoRegressive Moving Average* (ARMA). A time series $y(t)$ is an ARMA(p,q) process if it is stationary and if for every t:

$$\begin{aligned} y(t) &= \phi_1 y(t-1) + \cdots + \phi_p y(t-p) \\ &\quad + \epsilon(t) + \theta_1 \epsilon(t-1) + \cdots + \theta_q \epsilon(t-q), \end{aligned} \tag{11}$$

where ϕ_i and θ_j are the parameters of the model, and $\epsilon(t)$ are error terms. The error terms $\epsilon(t)$are assumed independant, identically distributed sampled from a normal distribution with zero mean and finite variance σ^2.

The equation can also be written in a more concise form as:

$$y(t) = \sum_{i=1}^{p} \phi_i L^i y(t) + (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon(t), \tag{12}$$

where L is the backward shift operator defined as follows: $L^i y(t) = y(t-i)$. We notice that AR and MA are special cases when $q = 0$ or $p = 0$.

The ARMA model fitting procedure assumes the data to be stationary. If the time series exhibits variations that violate the stationary assumption, then there are specific approaches to make the time series stationary. The most common one is what is often called the "differencing operation". It is defined by $(1-L)y(t) = y(t) - y(t-1)$. It can be shown that a polynomial trend of degree k is reduced to a constant by differencing k times, that is, by applying the operator $(1-L)^k y(t)$. We could therefore proceed by differencing repeatedly until the resulting series can plausibly be modeled as a realization of a stationary process. An ARIMA(p,d,q) model is an ARMA(p,q) model that has been differenced d times. Thus, the ARIMA(p,d,q) can be given by:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1-L)^d y(t) = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\epsilon(t). \tag{13}$$

For training the ARMA model, *Powell's function minimization routine* is used to choose the coefficients to minimize the sum of squared prediction errors [26].

In order to estimate parameter p and q, there is no automatic technique. In fact we need to examine sample autocorrelation function (ACF) and partial auto-correlation function (PACF) to get an idea of potential p and q values. Consequently, we specified parameters as used usually in literature [20] thus $p = 4$, $d = 1$ and $q = 4$.

2.3.3 *Linear minimum square error (LMMSE)*

LMMSE Predictor consists of predicting the aggregate series sample, $y(t+1)$, in the next interval as a weighted sum of the past n average samples:

$$y(t+1) = [a_1 \quad a_2 \quad \cdots \quad a_n] \begin{bmatrix} y(t) \\ \cdots \\ y(t-n+1) \\ y(t-n) \end{bmatrix}, \tag{14}$$

where $a_1\, a_2 \cdots a_n$ are the LMMSE coefficients. Those coefficients can be expressed as:

$$\begin{aligned} &[a_1 \quad a_2 \quad \cdots \quad a_n] \\ &\quad = [R(1) \quad \cdots \quad R(n-1) \quad R(n)] \\ &\quad\quad \times \begin{bmatrix} R(0) & R(1) & \cdots & R(n-1) \\ R(1) & R(0) & \cdots & R(n-2) \\ \cdots & & & \\ R(n-1) & R(n-2) & \cdots & R(0) \end{bmatrix}^{-1}, \end{aligned} \tag{15}$$

where $R(i)$ is the covariance function of the time series, and can be estimated as:

$$R(i) = \frac{1}{n} \sum_{t=i+1}^{n} y(t)y(t-i), \quad 0 \le i \le n-1, \tag{16}$$

where n is the number of aggregate series samples kept and it is a tunable parameter. We consider $n = 10$ for the simulation.

3 A structure preserving method of traffic sampling

3.1 Maximum-based sampling

Traffic measurement consists of collecting data from the network. To achieve this task, two main approaches are used.

The first one consists of examining every single packet traversing a given router (measurement spot). It is obvious that this approach is infeasible since the required processing resource and the storage cost grow exponentially as the throughput increases.

The second approach employs sampling at the packet level to control the consumption of resources in measurement networks infrastructure. Many routers that are equipped with traffic measurement tools provide statistics from a sampled streams because of the limitations on the memory size and the processing resource of the collected data. Hence, the main advantage of this traffic measurement approach is the reduction of storage and processing at the collector equipment. In this work, we use the throughput of the traffic (Mbps) which can be obtained easily using any traffic measurement tool. The throughput is defined as the incoming input rate in Megabit per second.

Several studies in the literature have analysed the network traffic [17, 24]. The observations from the collected time series lead to three main findings:

1. There are strong periodicities in the data.
2. The time series exhibit evident self-similarity and long-range dependence, i.e. non-stationary behavior.
3. The traffic have a heavy-tailed distribution which leads to is that it exhibits extreme variability. Practically speaking, a heavy-tailed distribution gives rise to very large values with nonnegligible probability so that sampling from such distribution results in the bulk of values being "small" but a few samples having "very" large values. Not surprisingly, heavy-tailedness impacts sampling by slowing down the convergence rate of the sample mean to the population mean, dilating it as the tail index approaches 1.

Such findings can be exploited in the forecasting process. In fact, periodicity and long-range dependence imply that the time series behavior from one interval to the next can be predicted.

Consequently our objective is to consider the traffic parameters during cycle n (the peak rate in our study), we aim to design a method to predict the required bandwidth by the next cycles $(n, n+1, \ldots)$. Hence, the resource provisioning can be done efficiently and intelligently, especially in the context of IP over WDM.

For instance, intelligent and efficient provisioning at the optical layer requires adding lightpaths in a specific areas of the network according to Internet layer dynamics. Therefore, we propose to perform traffic forecast at peak level of traffic instead of the original traffic. In other words, we aim to predict successive traffic peaks that represent potential link congestion causes. By accurately predicting upcoming high traffic loads, the network controller can provision optical resources accordingly.

In order to achieve this aim, we propose a periodic sampling method (so-called *maximum-based sampling*) that picks one measurement during a sampling interval of size τ. This measurement represents the maximum value calculated over the sampling interval i.e. the peak rate. The online traffic monitoring approach consists of the subsequent tasks (Fig. 4):

1. Set the sampling interval size equal to τ.
2. Determine corresponding maximums (M_i) in interval $(I_i)_{i=1,\ldots,N}$.
3. Apply the prediction method on the newly constructed process $(M_i)_{i=1,\ldots,N}$.

Elbiaze et al. [7] have observed that the peak rates calculated over non-overlapping intervals varies from one interval to the next keeping the same behavior as the original traffic, i.e., incorporating correlation. On the other hand, the

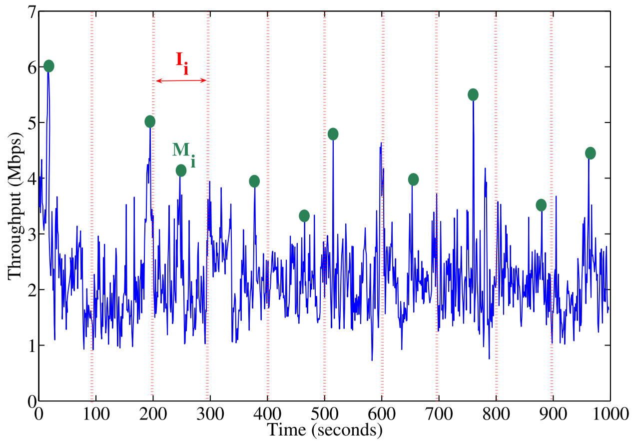


Fig. 4 An illustration of the max-based sampling technique

family of peak rates behaves in the same way in different lengths, i.e. in many time scales. Intuitively, that means the presence of self-similarity property in the sequence of peak rates. Based on this constatation, we try to prove mathematically the self-similarity of such sampled data.

3.2 Mathematical proofs for the self-similarity preservation

The most useful definition of self-similar for our purpose is the following: A process $X(t)$ $-\infty < t < \infty$ is self-similar with parameter H if $X(at)$ and $a^H X(t)$ have identical finite dimensional distributions for all a. This is equivalent to say that for any positive a and for any finite integer m

$$P(X(at_1) \le x_1, \ldots, X(at_m) \le x_m) \\ = P(a^H X(t_1) \le x_1, \ldots, a^H X(t_m) \le x_m). \tag{17}$$

Theorem 1 *Let X(t) be a continuous self-similar (S.S) with parameter H and let P be a partition of* $[0, \infty)$ *into disjoint half-open intervals* $[t_0, t_1), [t_1, t_2), \ldots, [t_n, t_{n+1})$ *with* $t_0 = 0$ *such that*

$$\bigcup_{i=0}^{\infty} [t_i, t_{i+1}) = [0, \infty).$$

Define

$$Y(t_i) = \max_{t \in [t_i, t_{i+1})} X_t.$$

Then $\{Y(t_i)\}_{i=1}^{\infty}$ *is self-similar with the same parameter* H.

Proof of Theorem 1 We will first consider a sequences of closed disjoint intervals $I_{i,n} = [t_i, t_{i+1} - 1/n]$ and define

$$Y^{\star}_{t,n} = \max_{t \in I_{i,n}} X_t$$

and then for each fixed n we will show that the $\{Y^{\star}_{t,n}\}_{i=1}^{\infty}$ is S.S with parameter H.

By the continuity of the stochastic process X_t it follows that if $\{Y^{\star}_{i,n}\}_{i=1}^{\infty}$ is S.S with parameter H then $\{Y(t_i)\}_{i=1}^{\infty}$ is also.

Fix n and consider $\{Y^{\star}_{i,n}\}_{i=1}^{\infty}$ (for convenience we drop the subscript n and write simply $Y^{\star}_i$).

We must show first that

$$P(Y^{\star}(at_i) \le x) = P(a^H Y^{\star}(t_i) \le x). \tag{18}$$

We must then show that $Y^{\star}(at_i)$ and $a^H Y^{\star}(t_i)$ have the same finite dimensional distributions. This is equivalent to showing for disjoint intervals $[t_{i_1}, t_{i_1+1} - 1/n) \cdots [t_{i_m}, t_{i_m+1} - 1/n)$

$$P(Y^{\star}(at_{i_1}) \le x_1, \ldots, Y^{\star}(at_{i_m}) \le x_m) \\ = P(a^H Y^{\star}(t_{i_1}) \le x_1, \ldots, a^H Y^{\star}(t_{i_m}) \le x_m). \tag{19}$$

□

Proof of (18) We know that $\forall t \in I_{i,n}$

$$P(X(at) \le x) = P(a^H X(t) \le x) \tag{20}$$

it follows

$$P\left(\max_{t \in I_{i,n}} X(at) \le x\right) \le P\left(\max_{t \in I_{i,n}} (a^H X(t)) \le x\right) \quad \forall t \in I_{i,n}.$$

Which is equivalent to

$$P(Y^{\star}(at_i) \le x) \le P(a^H Y^{\star}(t_i) \le x). \tag{21}$$

It also follows from (20) that

$$P\left(\max_{t \in I_{i,n}} a^H X(t) \le x\right) \le P\left(\max_{t \in I_{i,n}} X(at) \le x\right) \quad \forall t \in I_{i,n}.$$

Which is equivalent to

$$P(a^H Y^{\star}(t_i) \le x) \le P(Y^{\star}(at_i) \le x). \tag{22}$$

The inequality (21) and the reverse (22) are true and thus the equality 18 is true. □

Proof of (19) We know by property (17) which defines the self-similarity of X that for each $t_1 \in I_{1,n}, \ldots, t_m \in I_{m,n}$

$$P(X(at_1) \le x_1, \ldots, X(at_m) \le x_m) \\ = P(a^H X(t_1) \le x_1, \ldots, a^H X(t_m) \le x_m). \tag{23}$$

By repeating the process we used to show that (20) ⇒ (18) in a nested fashion, we can show that (3.2) ⇒ (19). Thus (19) is true. □

Proof of Self-similarity By continuity of the stochastic process since $\lim_{n\to\infty}(t_{i+1} - 1/n) = t_{i+1}$ and using (19)

$$\begin{aligned} &P(Y^\star(at_{i_1}) \le x_1, \ldots, Y^\star(at_{i_m}) \le x_m) \\ &\quad = P(a^H Y^\star(t_{i_1}) \le x_1, \ldots, a^H Y^\star(t_{i_m}) \le x_m) \\ &\quad \Rightarrow P(Y(at_1) \le x_1, \ldots, Y(at_m) \le x_m) \\ &\quad = P(a^H Y(t_1) \le x_1, \ldots, a^H Y(t_m) \le x_m) \end{aligned} \tag{24}$$

and this proves the self-similarity of Y. □

4 Experimental results

The first part of the simulation experiments was conducted to investigate a comparison between the max-based sampling, the systematic sampling, the stratified sampling and the aggregated data. It aims to show which sampling technique preserves the traffic characteristics using some statistical parameters. Thus we evaluate the sampled mean (the average of sampled data), the sampled variance (the variance of the sampled data), the self-similarity and the burstiness of the sampled data. The second part investigates a comparison of the prediction performance of the introduced models α_SNF, ARIMA and LMMSE using the sampled data (with the max-based, the systematic and the stratified techniques for various time scales). This comparison aims to show which sampling technique allows more accurate prediction.

4.1 Traces and preprocessing

The used trace is the Auckland-VIII data set. The trace contains a two weeks GPS-synchronized IP header trace captured with an Endace DAG3.5E tap Ethernet network measurement card in December 2003 by NLANR.[1] In order to reduce processing time, we used 30 minutes of the data in experiments. For models which require training phase, we divide data into two sets of 15 minutes. The first set is used to estimate the model parameters (the training phase); the second set is used for the evaluation of the performance of the selected model. We extracted the throughput of the data every 10 ms. We performed max-based sampling on the throughput using various interval size $\tau = 100$ ms to 1000 ms. We also aggregated the traffic using (2) to obtain throughput for the various granularities in order to compare aggregated traffic to the sampled one.

[1]National Laboratory for Applied Network Research, http://pma.nlanr.net/Special/auck8.html.

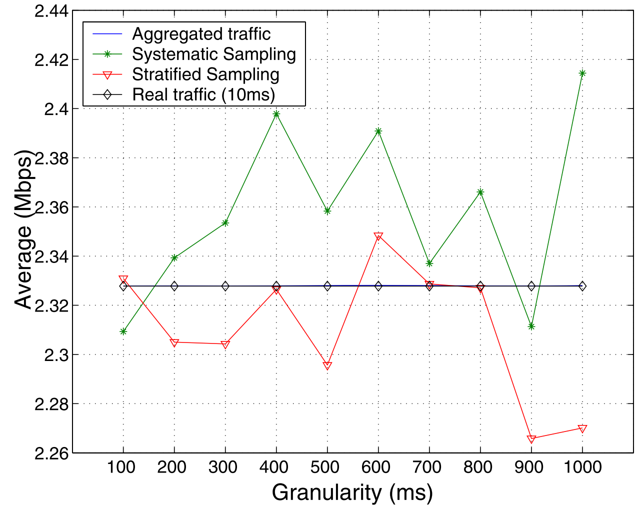


Fig. 5 Average of sampled data

4.2 Comparing the max-based technique with the existing sampling techniques

As mentioned before, the sampling technique has to show the self-similarity and the burstiness of the traffic. The aim is to find the suitable sampling technique able to improve the prediction performance. Unlike previous work [10], the preservation of the mean or the variance of the data isn't really necessary since the sampling here aims to prepare data for the prediction. In fact, the resource provisioning for the next interval time doesn't need the prediction of the average peak rate but it needs the maximum peak rate.

However, we performed some comparisons between sampling techniques based on some statistical parameters. Several experiments are performed using various sampling interval (form 100 ms to 1000 ms).

Figure 5 shows the average of the obtained sampled data compared to real traffic average. The stratified sampling scheme in most cases underestimates the average whereas the systematic sampling overestimates it. Only aggregation of the traffic preserves its average. However, we didn't draw the average of the max-based sampled data because its obviously higher than for the other techniques.

Figure 6 depicts the variance of the sampled traffic compared to that of the real traffic. It shows that aggregating the data clearly reduces the variance of the data.

This result can be easily proved mathematically. Recall that the variance of the sample mean var($\bar{Z}$) of a random variable Z satisfies var($\bar{Z}$) $= \sigma_Z^2/m$ where m is the sample size. From (2) it follows that $var(X^{(m)}) = \sigma^2 m^{2H-2}$; this means that when aggregating, variance of the data is decaying in function of m and H. This could be inadequate for the prediction of peaks and bursts because the prediction model won't detect the variability lost when aggregating. The comparison between the aggregated data and the

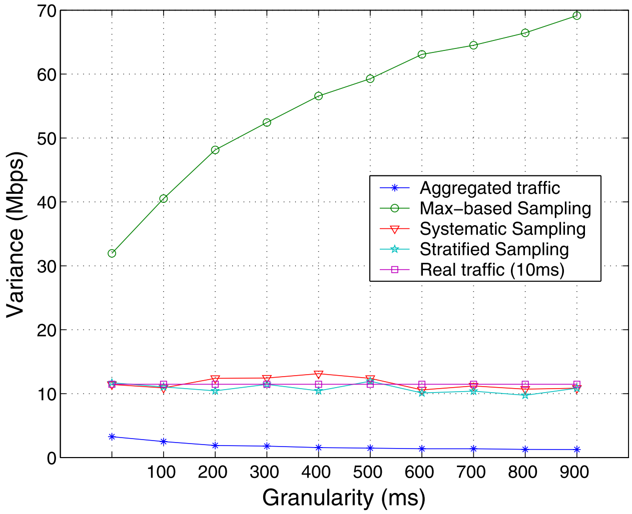


Fig. 6 Variance of sampled data

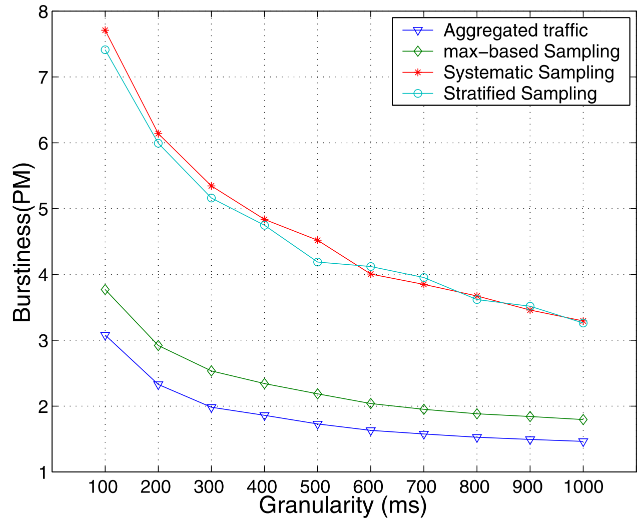


Fig. 8 Measurement of burstiness (peak-to-mean)

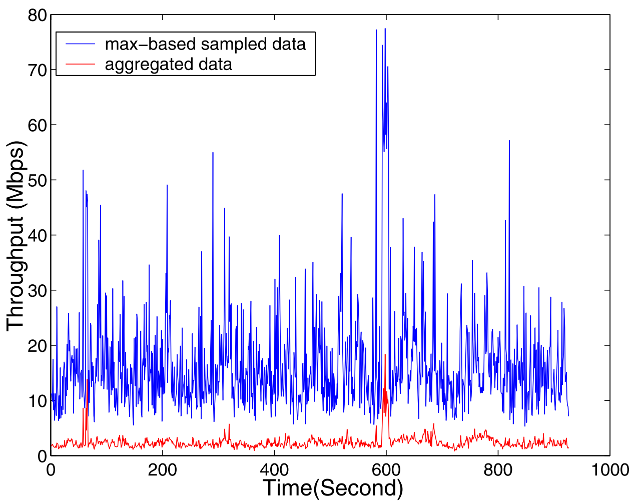


Fig. 7 Comparison of aggregated data and max-based sampled data

max-based sampled data (Fig. 7) shows that the aggregated data has lost its variability which is important for predicting the traffic.

Figure 6 shows also that the variance of the signal is almost preserved for the systematic and the stratified sampling techniques.

Figure 8 depicts the burstiness (PM) versus the granularity for the aggregated traffic, the sampled traffic with the three sampling techniques. It shows also that burstiness of systematic and stratified sampled data are very important. Whereas aggregated traffic and max-based sampled data are less bursty. Intuitively, low burstiness makes the data more "predictible" i.e. we will have better prediction performance in terms of error. Consequently, according to Fig. 8, aggregated data and max-based sampled data will improve the quality of the prediction. This will be proved in prediction experiments in the next paragraph.

On the other hand, the obtained Hurst parameter (Fig. 9) differs for various sampling interval size and for the different sampling techniques. We used the wavelet-based method to estimate the Hurst parameter [1] with 95% confidence interval.

It is clear that for the systematic and the stratified sampling the Hurst parameter is decreased. It decreases under 0.5 (granularity 300 ms) which means that the data has lost its self-similarity. The max-based sampling almost preserves the same Hurst parameter for all granularities. However, traffic aggregating remarkably increases the Hurst parameter i.e. it increases the self-similarity and the long-range dependence of the data. This could mean that aggregated data and max-based sampled data will be more predictible.

We also notice that the obtained results show that Hurst parameter (self-similarity) doesn't always reflect burstiness. In fact our burstiness metric shows that systematic and stratified sampling techniques has led to a burstier traffic than the aggregated or the max-based sampled data (Fig. 8). Whereas the estimation of Hurst parameter shows that the data obtained by systematic and stratified sampling are less self-similar and long-range dependant than the data resulting from the aggregation or the max-based sampling techniques (Fig. 9). It means that high burstiness doesn't reflect self-similarity. This phenomena justifies the use of the PM as a metric of burstiness instead of Hurst parameter (Sect. 2.2). This observation is important for future analysis of the traffic in the way that we must take into account that burstiness doesn't involve self-similarity and vice-versa.

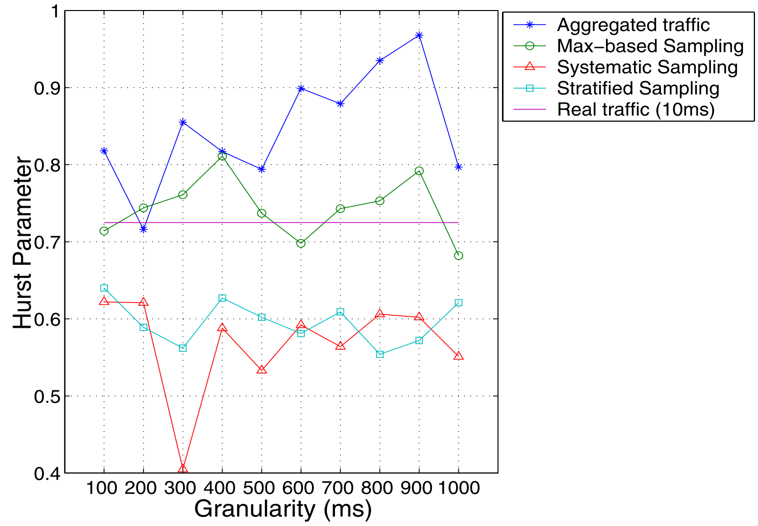


Fig. 9 Hurst parameter for various granularities

4.3 Exploiting the sampling techniques for traffic prediction

In this paragraph, we undertake several prediction experiments using the sampled data obtained from the Auckland data set. We use the three sampling techniques and the aggregation for treating data before performing the predictions. The aim is to compare the prediction performance obtained when using each sampling technique.

Figure 10 depicts obtained prediction error (ARV) using the LMMSE predictor with respect to the granularity. Each curve represents the obtained ARV using a sampling technique for several granularities. The curves shows that the prediction errors using systematic or the stratified sampling are important for all granularities. The prediction using aggregated data improves the performance compared to the stratified and systematic sampling. For the max-based sampling technique, the results are improved. The figure shows that the achieved performance is steady. The prediction error is often less than 1 for all granularities.

The results confirm our observations in the previous paragraph. In fact, stratified and systematic sampling have provided a burstier data which has lost its self-similarity. Consequently the data has became less "predictible" (important prediction error). On the contrary, The aggregation and the max-based sampling has provided a less bursty data which has conserved the self-similarity and the long-range dependence. Consequently, the resulting data has became more "predictable". We also note that even though aggregated traffic provides good prediction performance in terms of error, data has lost its variance when aggregating. This means that predicted data won't show the variance of the data. However the max-based sampling provides better prediction performance while data preserves its high variance.

Figure 10 shows a comparison between the performance found using the prediction models. We undertook experiments for several granularities. We used sampled data using the max-based technique. In experiments, the number of input is 10 delays for the LMMSE model, 4 inputs for the ARIMA model. The number of inputs for the α_SNF model is determined before each prediction using the concept of mutual information (Sect. 2.3.1). The number of inputs for the α_SNF model is specified in the figure for each prediction and for each granularity. Prediction models having a training phase (α_SNF and ARIMA) give better performance ($ARV \prec 1$) for all granularities (Fig. 11). The LMMSE model gives higher prediction error especially for small granularities. We can note a small improvement of prediction performance for the ARIMA$(4, 1, 4)$ model compared the α_SNF. However, the α_SNF model uses less input variables for all cases (less than 4 inputs). It gives almost the same prediction performance of the ARIMA model.

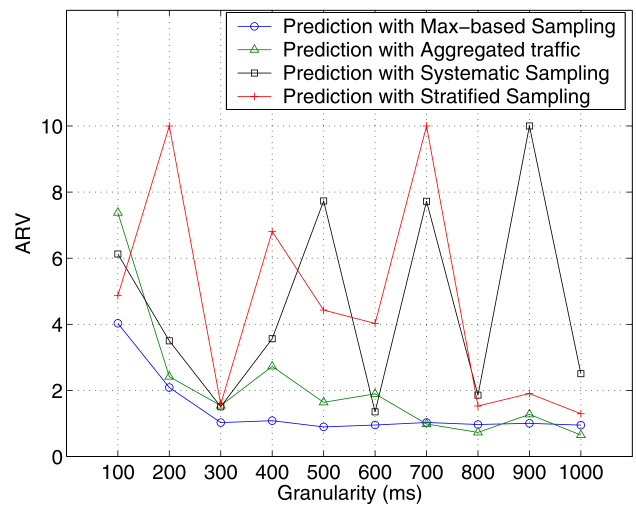


Fig. 10 LMMSE-Prediction using data sampled with various techniques

Table 1 Summary of the obtained results

	Average	Variance	Self-similarity	Burstiness	Prediction performance
Max-based Sampling	Treats peaks	High	Preserved	Preserved	Good
Systematic Sampling	Lost	Preserved	Lost	Preserved	Unreliable
Stratified Sampling	Lost	Preserved	Lost	Preserved	Unreliable
Aggregating	Preserved	Lost	Overestimated	Lost	Bad for bursts

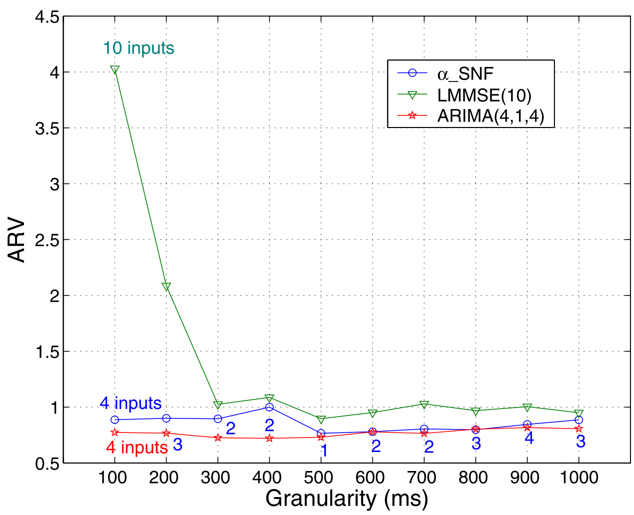


Fig. 11 Comparison of the prediction models

We can also notice that for small granularities, the prediction error is important (Figs. 10, 11). This can be explained by the high burstiness of the traffic for these granularities (Fig. 8). This validates our conjecture that bursty traffic makes data less "predictible". Consequently, using high granularities reduces burstiness. Hence, it improves the prediction performance.

Table 1 provides a summary of the obtained results. It compares the sampling techniques with respect to the preservation of the statistical parameters, the self-similarity, the burstiness and the effect on the prediction performance.

5 Conclusions and future work

In this paper, we have investigated several important issues in employing sampling techniques for measuring Internet traffic. We have proposed a structure-preserving sampling method to be used together with the LMMSE, ARIMA or α_SNF prediction-based models. The whole technique should be used to improve the resource provisioning mechanism (Fig. 1).

We showed here that in case of self-similar traffic, knowledge of fundamental characteristics of the traffic can provide new insight into data treatment toward traffic prediction. Therefore, the sampling technique applied to the data has an important impact on the performance of the prediction. In this context, we showed the efficiency of the max-based sampling as structure preserving method of Internet traffic compared to common sampling and aggregating techniques used in literature. We carried out experiments using real Internet traces. We confirmed the effectiveness of the max-based sampling to preserve traffic characteristics as well as to improve the prediction performance compared to other sampling and aggregating techniques.

We also undertook prediction experiments using the sampled data for various time scales. Thus, we find that the ARIMA and the α_SNF models—which need training phase—give more accurate predictions in terms of error than the LMMSE which gives bad prediction performance for different time scales.

We believe that the proposed structure-preserving sampling method can be judiciously exploited in multiple networking areas such as monitoring, alarms generation, resource provisioning…etc. Thus, our future work focuses on building several simulation scenarios to exploit the performance of the predictions using the structure preserving method in a wide range networking areas:

Intelligent alarm generation We believe that the maximum-based traffic sampling together with ARIMA or α_SNF prediction can help generating alarms in an efficient manner. For instance, a threshold alarm can be generated when a parameter of interest (such as traffic on a link, disk usage) exceeds a certain threshold. Hence, if the system keeps track of the past alarms generated from a threshold excess, the arrival of the upcoming alarm can be predicted and the operator can be prepared to take the necessary decisions accordingly.

Traffic engineering and control It is necessary to make the reservations a significant amount of time ahead of the arrival of the traffic in order to allow for resource assignment delays. This prediction-based method is vital for different network contexts, especially for IP over WDM networks where a major part of optical cross-connects still configured manually. This necessitates the prediction of the future requirements of source-destination traffic flows Hence, we will build a simulation-based model for a mesh optical network to investigate the bandwidth allocation scheme using our structure preserving sampling method. We study the performance of this allocation technique and compare it with other approaches.

References

1. Abry, P., & Veitch, D. (1998). Wavelet analysis of long range dependent traffic. *IEEE Transactions on Information Theory*, *44*(1), 2–15.
2. Barakat, C., Thiran, P., Iannaccone, G., Diot, C., & Owezarski, P. (2003). Modeling Internet backbone traffic at the flow level. *IEEE Transactions on Signal Processing*, **51**(8).
3. Beran, J., Sherman, R., Taqqu, M. S., & Willinger, W. (1995). Long range-dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, *43*(2–4), 1566–1579.
4. Cetinkaya, C., & Knightly, E. (2000). Egress admission control. In *Proceedings IEEE INFOCOM*.
5. Claffy, K. C., Polyzos, G. C., & Braun, H. W. (2003). Application of sampling methodologies to network traffic characterization. In *Proceedings of ACM. SIGCOMM'93*.
6. DiMicelli, T. (2004). Optical control planes: time to shake up the network. In *FibreSystems Europe in association with LIGHTWAVE Europe* (p. 16).
7. Elbiaze, H., Chahed, T., Atmaca, T., & Hébuterne, G. (2003). Shaping self-similar traffic at access of optical network. *Performance Evaluation Journal*, *53*(3–4), 187–208.
8. Floyd, S., Handley, M., Padhye, J., & Widmer, J. (2000). Equation-based congestion control for unicast applications. *SIGCOMM Computer Communication Review*, *30*(4), 43–56.
9. Gao, Y., He, G., & Hou, J. C. (2002). On leveraging traffic predictability in active queue management. In *Proceedings IEEE INFOCOM*, New York.
10. He, G., & Hou, J. C. (1993). An in-depth analytical study of sampling techniques for self similar Internet traffic. In *Proceedings of the 25th IEEE international conference on distributed computing systems*.
11. He, G., & Hou, J. C. (2003). On exploiting long range dependency in measuring cross traffic. In *Proceedings of IEEE INFOCOM 2003*, San Francisco, CA.
12. He, G., Gao, Y., Hou, J. C., & Park, K. (2002). A case for exploiting self-similarity of network traffic in tcp. In *ICNP '02: Proceedings of the 10th IEEE international conference on network protocols* (pp. 34–45), Washington, DC, USA. IEEE Computer Society.
13. Hohn, N., & Veitch, D. (2006). Inverting sampled traffic. *IEEE/ACM Transactions on Networking*, *14*(1), 68–80.
14. Leland, W. E., & Wilson, D. V. (1991). High time-resolution measurement and analysis of LAN traffic: implications for LAN interconnection. In *Proceedings IEEE INFOCOM* (pp. 1360–1366).
15. Leland, W., Taqqu, M. S., Willinger, W., & Wilson, D. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, *2*, 1–15.
16. Mathis, M., Semkeand, J., Mahdavi, J., & Ott, T. (1997). The macroscopic behavior of the TCP congestion avoidance algorithm. *Computer Communications Review*, *27*, 67–82.
17. Papagiannaki, K., Taft, N., Zhang, Z.-L., & Diot, C. (2003). Long-term forecasting of Internet backbone traffic: observations and initial models. In *IEEE Infocom*, San Francisco.
18. Park, K., Kim, G., & Crovella, M. (1997). On the effect of traffic self-similarity on network performance. In *Proceedings SPIE int. conf. performance and control of network systems*.
19. Paxson, V., & Floyd, S. (1995). Wide-area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, *3*(3), 226–244.
20. Qiao, Y., Skicewicz, J., & Dinda, P. (2002). *Multiscale predictability of network traffic* (Technical Report NWU-CS-02-13). Department of Computer Science, Northwestern University.
21. Rabbat, M. G., Coates, M. J., & Nowak, R. D. (2006). Multiple-source Internet tomography. *IEEE Journal on Selected Areas in Communications*, *24*(12), 2221–2234.
22. Rouai, F. A., & Ahmed, M. B. (2001). A new approach for fuzzy neural network weight initialization. In *INNS-IEEE international joint conference on neural network (IJCNN)*, 2.
23. Rubenstein, D., Kurose, J., & Towsley, D. (2000). Detecting shared congestion of flows via end-to-end measurement. In *Proceedings of ACM SIGMETRICS*.
24. Soule, A., Nucci, A., Cruz, R., Leonardi, E., & Taft, N. (2004). How to identify and estimate the largest traffic matrix elements in a dynamic environment. In *SIGMETRICS '04/performance '04: proceedings of the joint international conference on measurement and modeling of computer systems* (pp. 73–84), New York, NY, USA. ACM Press.
25. Tellenbach, B., Brauckhoff, D., & May, M. (2008). Impact of traffic mix and packet sampling on anomaly visibility. In *The third international conference on internet monitoring and protection, ICIMP* (pp. 31–36).
26. Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1986). *Numerical recipes in Fortran*. Cambridge: Cambridge University Press.
27. Zhang, L., & Zhang, T. (2007). Small packet threshold adaptive sampling for network management. In *International conference on mechatronics and automation* (pp. 811–815).
28. Zhani, M. F., Elbiaze, H., & Kamoun, F. (2007). α_SNFAQM: an active queue management mechanism using neurofuzzy prediction. In *Proc. of the 12th IEEE symposium on computers and communications (ISCC)*.
29. Zhani, M. F., Elbiaze, H., & Kamoun, F. (2008). Analysis of prediction performance of training-based models using real network traffic. In *Proc. of international symposium on performance evaluation of computer and telecommunication systems (SPECTS)*.

Halima Elbiaze received the BS degree in Applied Mathematics from the University of Rabat, Morocco, in 1996. She received the MS and Ph.D. degrees in Computer Science from the University of Versailles, France, in 1998 and 2002, respectively. She is currently an Assistant Professor in the Department of Computer Science, University of Quebec in Montreal, Canada. Her research interests are in the area of Quality of Service, performance evaluation and traffic engineering for high speed networks (IP/WDM, TCP/IP, ATM, FR, etc.).

Mohamed Faten Zhani was born in Kairouan, Tunisia on January 25, 1980. He received the Engineering and the M.S. degrees from the National school of computer engineering of Tunis in 2003 and 2005, respectively. He is currently a Research Assistant in the department of Computer Science in the University of Quebec in Montreal and he is currently working toward the Ph.D. degree in Computer Science at the National school of computer engineering of Tunis and the University of Quebec in Montreal. His research interests include network traffic modeling, analysis and prediction, congestion control and active queue management.

Omar Cherkaoui is professor in computer science department of University of Quebec in Montreal, Canada. He received his Ph.D degree in 1990 from University of Montreal. His research interests include management of high speed networks and multimedia application over IP networks. He is the head of "teleinformatique" laboratory where his leading different R&D projects.

Farouk Kamoun was born in Sfax, Tunisia on October 20, 1946. He received the Engineering Degree from Ecole Supérieur d'Electricité, Paris, France in 1970 and the Ph.D. degree in computer science, in 1976, from UCLA, where he participated in the ARPA Network Project and did his research on design considerations for large computer communication networks. He is currently Professor and Director of CRISTAL Research Laboratory at the Ecole Nationale des Sciences de l'Informatique, Tunisia. His research areas of interest deal with protocols for ad-hoc networks, quality of service and network measurements.