Alcom-FT Technical Report Series
ALCOMFT-TR-03-20

Modeling ADSL traffic on an IP backbone link

Nadia Ben Azzouna* Christine Fricker‡ Fabrice Guillemin§

Abstract

Measurements from an Internet backbone link carrying TCP traffic towards different ADSL areas are analyzed in this paper. For traffic analysis, we adopt a flow based approach and the popular mice/elephants dichotomy. The originality of the experimental data reported in this paper, when compared with previous measurements from very high speed backbone links, is in that commercial traffic comprises a significant part due to peer-to-peer applications. This kind of traffic exhibits some remarkable properties in terms of mice and elephants, which are described in this paper. It turns out that by adopting a suitable level of aggregation, the bit rate of mice can be described by means of a Gaussian process. The bit rate of elephants is smoother than that of mice and can also be well approximated by a Gaussian process.

1 Introduction

Characterization of Internet traffic has become over the past few years one of the major challenging issues in telecommunications networks. As a matter of fact, understanding the composition and the dynamics of Internet traffic is essential for network operators in order to offer quality of service and to supervise their networks. Since the celebrated paper by Leland *et al* [11] on the self-similar nature of Ethernet traffic in local area networks, a huge amount of work has been devoted to the characterization of Internet traffic. In particular, different hypotheses and assumptions have been explored to explain the reasons why and how Internet traffic should be self-similar (see for instance [7, 9]).

A common approach to describing traffic in a backbone network consists of observing the bit rate process evaluated over fixed length intervals, say a few hundreds of milliseconds. Long range dependence as well as self-similarity are two basic properties of the bit rate process, which have been observed through measurements in many different situations. Different characterizations of the fractal nature of traffic have been proposed in the literature (see for instance Norros [13] on the monofractal characterization of traffic and Levy-Véhel [16], Abry *et al* [2] on the multifractal properties of traffic). An exhaustive account to fractal characterization of Internet traffic can be found in the book by Park and Willinger [14].

Even though long range dependence and self similarity properties are very intriguing from a theoretical point of view, their significance in network design has recently been questioned in the paper by Cao and Ramanan [5], where it is shown that the overflow probability in a buffer fed with the superposition of a large number of flows satisfying some reasonable regularity assumptions can be well approximated by that obtained when the input process is Poisson. While the above result may not be directly applicable in an access network with limited transmission capacities, the assumption of a large number of flows is reasonable on a high speed backbone link and are in favor of using a simple $M/G/1$ queue for buffer dimensioning in a backbone network composed of gigarouters.

While self-similar models introduced so far in the literature aims at describing the global traffic on a link, it is now usual to distinguish short transfers (referred to as mice) and long transfers (referred to as elephants) [15]. This dichotomy was not totally clear up to a recent past (see for instance network measurements from the MCI backbone network [6]). Yet, the distinction between mice and elephants become more and more evident with the emergence of peer-to-peer (p2p) applications, which give rise to a large amount of traffic on a small number of TCP connections, as it will be shown in the following.

In this paper, we analyze TCP traffic on an Internet backbone link collecting data in direction to several ADSL areas. The primary goal of this paper is to draw attention to several salient features of ADSL traffic. In particular, we consider commercial traffic, which comprises a significant part of p2p traffic, giving rise to very large elephants.

The above observation leads us to analyze ADSL traffic by adopting a flow based approach and more precisely the mice/elephants dichotomy. The intuitive definition of a

*France Telecom R&D, 2 Avenue Pierre Marzin, 22300 Lannion, France. Email: Nadia.Benazzouna@rd.francetelecom.com

†Partially supported by the contract CRE 1 03 D1022 with France Telecom and the Future and Emerging Technologies programme of the EU under contract number IST-1999-14186 (ALCOM-FT)

‡INRIA, Domaine de Voluceau, 78150 Rocquencourt, France. Email: Christine.Fricker@inria.fr

§France Telecom R&D, 2 Avenue Pierre Marzin, 22300 Lannion, France. Email: Fabrice.Guillemin@francetelecom.com

mouse is that such a flow comprises a small number of packets so that it does not leave or leaves slightly the slow start regime. Thus, a mouse is not very sensitive to the bandwidth sharing imposed by TCP. On the contrary, elephants are sufficiently large so that one can expect that they share the bandwidth of a bottleneck according to the flow control mechanism of TCP. As a consequence, mice and elephants have a totally different behavior from a modeling point of view.

The organization of this paper is as follows: Basic definitions are given in Section 2. Mouse traffic is analyzed in Section 3 in the case of non p2p mice and in Section 4 in the case of p2p mice. Elephant traffic is described in Section 5. Finally, some concluding remarks are presented in Section 6. Theoretical results concerning convergence to Gaussian processes as well as the computation of autocorrelation functions and spectral densities are deferred to the Appendix.

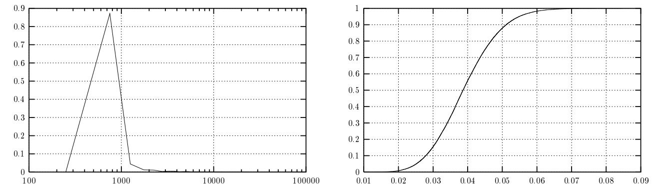
2 Traffic analysis

2.1 Definitions and notation

Throughout this paper, we consider a 1 Gbps link between the France Telecom IP backbone network and several ADSL areas. Traffic originated or in direction to these different ADSL areas is multiplexed on this single link. In the following, we observe TCP traffic from the IP backbone network towards the ADSL areas (downstream traffic). It is worth noting that traffic local to an ADSL area cannot be observed in the collected data.

To analyze traffic characteristics, we adopt the mice and elephants dichotomy. Before proceeding further, we have to emphasize the fact that there is no commonly adopted definition for a mouse. A mouse is intuitively a data transfer, which does not leave or leaves slightly the slow start period. In fact, a mouse is a short time transfer, which has no time to adapt to network conditions according to the fairness criterion imposed by TCP.

As a convention, we adopt in this paper the following definition: a mouse is a data transfer comprising a number of packets less than or equal to 20; a flow is terminated when no packets of the flow have been observed for a time period of 5 seconds. Other definitions for mice are possible; for instance in the paper by Zhang *et al* [17], a small data transfer contains at most 10^4 bytes. If the MTU is equal to 1500 bytes, 10^4 bytes roughly correspond to 8 packets. The value of 20 packets is chosen because if we assume that the maximum congestion window size is 8Kbytes and if there is an ACK for each packet received by the destination, then about 15 packets are necessary to hit the maximum congestion window size at the end of the slow start phase.



(a) Flow size in bytes

(b) PDF of X_t/Z_t

Figure 1: Flows on a backbone link.

The timer of 5 seconds may appear at first glance very sharp. However, since we intend to describe the bit rate of mice, we have to consider the time period when the mouse is active (i.e., when some packets of the mouse are transmitted). Long mice are mostly due to FIN segments, which arrive quite a long time after the last data segment or single SYN segments, which eventually do not initiate any data transfer. This introduces some bias in the evaluation of the duration of mice. To avoid this phenomenon, we use the 5 second timer to remove segments, which are too far away from data segments. The counterpart of this method is that single packet mice artificially appear.

The flow size distribution is displayed in Figure 1(a). It turns out that the majority of flows comprise less than 1000 bytes and actually correspond, as shown in the following, to mice. Even though these flows are the most numerous, they contribute a very small proportion of the total amount of traffic, as shown in Figure 1(b) representing the distribution of X_t/Z_t, where $\{X_t\}$ is the amount of traffic due to mice and $\{Z_t\}$ is the global bit rate process. Mice actually contribute about 6% of global traffic but represent more than 97% of the total number of flows.

Finally, more than 49% of traffic is due to p2p applications (Kazaa, Morpheus, Edonkey, Gnutella, etc.), as shown in Table 1. The significant proportion of p2p traffic gives rise to remarkable phenomena, which are described in the next sections.

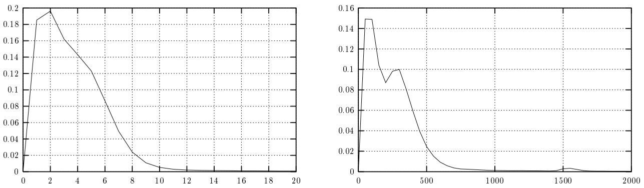
2.2 First observations

Figure 2 displays the distribution of the number of packets and bytes comprised in a mouse. It turns out that most of mice comprise less than 1000 bytes and as stated in the previous section, most of flows are indeed mice (see Figure 1(a)).

From Figure 2, we also observe that a large number of mice are composed only of one or two packets. Single packet mice are Reset segments, SYN segments, which are not really associated to a mouse because of transac-

Applications		percentage
non p2p	http	14.6
	ftp	2.1
	nntp	1.9
	others	31.8
	total non p2p traffic	50.4
p2p	Edonkey	37.5
	Kazaa&Morpheus	7.8
	Napster	3.8
	Gnutella	0.5
	Total p2p traffic	49.6

Table 1: Composition of ADSL traffic per application.



(a) Number of packets

(b) Number of bytes (in logarithmic scale)

Figure 2: Characteristics of a mouse.

tion interruption or very long response times by servers, or FIN segments, which arrive far away from the last data segments and which appear as single packet mice because of the 5 second timer used to terminate a mouse. Moreover, a large number of single packet mice are generated by p2p protocols.

Two packet mice are composed of SYN and FIN segments only. This is due to the fact that a large number of TCP connections (associated with HTTP transactions for instance) are opened and immediately closed or not used at all; this may be caused by too long response times by servers, which lead users to interrupt their transactions, or by the fact that certain implementations of HTTP systematically opens several TCP connections in parallel. Actually, only a small number of mice carry useful information (data segments). This phenomenon has to be taken into account when characterizing the mouse arrival process, as shown in the following.

When analyzing more carefully the generation process of mice, it turns out that mice generated by p2p protocols exhibit a behavior, which is quite different from that of other mice (regular mice related to usual applications using TCP, such HTTP, ftp, etc.). This is why we analyze the two types of mice separately. Note that since mice are not sensitive to TCP fairness, global mice traffic is the superposition of p2p and non p2p mice traffic; these two types of traffic do not really interact one with each other and can be analyzed separately.

3 Non p2p mouse traffic

In this section, we analyze mice, which are not generated by p2p protocols, i.e., with port numbers different from 1214 (Kazaa), 4662 (Edonkey), 6346 (Gnutella) and other p2p protocol port numbers. The objective of this section is to describe the bit rate process of those mice and to propose a probabilistic model approximating this process.

3.1 Observation of the bit rate process

Let X_n^m denote the number of bits due to non p2p mice in the time interval $(n\Delta, (n+1)\Delta]$ divided by $\Delta = 100$ ms. The process $\{X_n^m\}$ representing the "instantaneous" bit rate offered by non p2p mice is highly varying as displayed in Figure 3(a); $\{X_n^m\}$ has been observed over a time period of 4900 seconds between 1:27 pm to 2:51 pm; only a time interval of 700 seconds is displayed in Figure 3(a). The empirical distribution of X_n^m is displayed in Figure 3(b). As it will turn out in the following, the process $\{X_n^m\}$ can indeed be approximated by a Gaussian process.

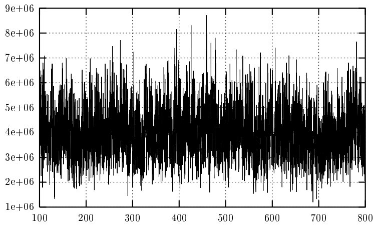
Explaining the form of the sample path of the process $\{X_n^m\}$ requires an in-depth description of non p2p mouse traffic. This is done in the next sections.

3.2 Mice arrival process

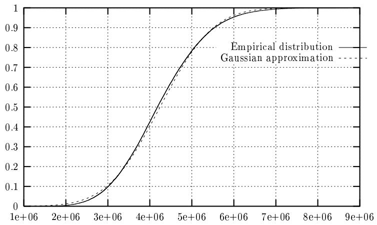
In a first step, we have observed arrivals of individual mice. The mouse inter-arrival time is remarkably exponential and the mouse duration is Weilbull as shown in Figure 4, where the complementary distribution function (cdf) of the inter-arrival time and that of the duration of mice are displayed. The inter-arrival time can be approximated by an exponential random variable with mean 0.003578. The cdf of the mouse duration $\mathcal{S}$ can be well approximated by a Weibullian distribution with zero location parameter, scale parameter $\eta = 1.035$ and skew parameter $\beta = 0.673034$; the mean mouse duration is equal to 1.39 s (which is very close to the theoretical value $\eta\Gamma(1+1/\beta) = 1.36$, $\Gamma(x)$ denoting the Euler Gamma function). This means that

$$\mathbb{P}(\mathcal{S} > x) \sim \exp\left(-\left(\frac{x}{\eta}\right)^{\beta}\right). \tag{1}$$

At first glance, one may conclude that mice arrive according to a Poisson process. However, when we compute



(a) Bit rate



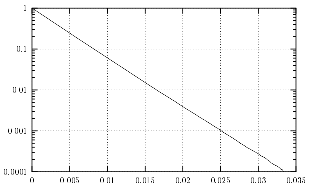
(b) Stationary distribution

Figure 3: Instantaneous bit rate and stationary distribution of the bit rate process $\{X_n^m\}$ estimated over time intervals with length $\Delta = 100$ ms.

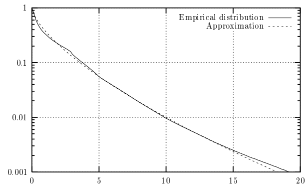
the stationary distribution of the number of active mice at an arbitrary instant, we should obtain a Poisson distribution if the mouse arrival process were Poisson (namely, the stationary distribution of the number of customers in an $M/G/\infty$ queue). In particular, the variance should be equal to the mean value. However, experimental data show that this last property is not verified. We specifically have the mean and the variance equal to 372 and 566.12, respectively. This is sufficient to show that the mouse arrival process is not Poisson.

To overcome this problem, we note that, as mentioned in the previous section, mice are actually not independent. In reality, for a same destination IP address, a certain number of mice arrive near one to each other, forming what we call in the following a macro-mouse. We specifically define a macro-mouse as a set of non p2p mice, which have the same destination address and which arrive within a rather short time interval, say with a length of 1 second; moreover, we impose that a macro-mouse comprises more than one packet.

The inter-arrival time of macro-mice is displayed in Figure 5(b) and the distribution of their duration is displayed in Figure 5(a). Their inter-arrival time is exponential with mean $1/\lambda_m = 0.00562$. The probability distribution of the duration of a macro-mouse can



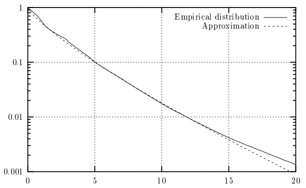
(a) Inter-arrival time



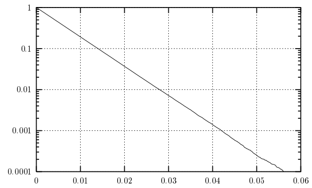
(b) Duration

Figure 4: Complementary distribution functions of the inter-arrival time and the duration of non p2p mice.

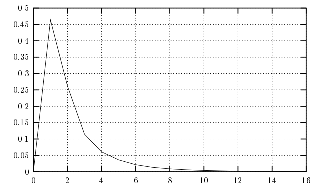
be well approximated by a two parameter Weibullian distribution with scale parameter $\eta_m = 1.78$ and skew parameter $\beta_m = 0.8$; the mean duration of a macro-mouse is $\mathbb{E}[S] = 2.136$ seconds (the theoretical value is $\eta_m \Gamma(1+1/\beta_m) = 2.01$ s). Finally, the distribution of the number of mice in a macro-mouse is displayed in Figure 5(c); the mean value is equal to 2.27.



(a) Duration



(b) Inter-arrival time



(c) Number of mice in a macro-mouse

Figure 5: Characteristics of a mouse group.

When computing the stationary distribution of the number of macro-mice active at a given instant (stationary distribution at an arbitrary instant), we get a Poisson distribution (see Figure 6). Moreover, we have computed the distribution of the number of macro-mice active at the arrival time of a macro-mouse (distribution at arrival instants). It turns out that these two experimental distributions are almost indistinguishable. As a consequence, we have the celebrated ASTA (Arrival See Time Averages) property. In view of the classical ANTIPASTA results [4], it is then reasonable to conjecture that the macro-mouse arrival process is Poisson. Of course, the Poisson process verifies ASTA; this is the well known as PASTA (Poisson Arrival See Time Averages) property. In Figure 6, the theoretical Poisson distribution with mean $\lambda\mathbb{E}[S]$ has been plotted to illustrate the coincidence of the different distributions.

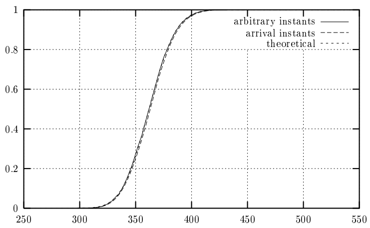


Figure 6: Distribution of the number of active macro-mice at arbitrary and arrival instants.

Thus, in spite of the fact that the individual mouse arrival process is not Poisson, grouping mice in an adequate manner yields a Poisson process. We then proceed wit the description of the bit rate process on the basis of macro-mice.

3.3 Bit rate process of mice

When we consider the bit rate created by macro-mice, we can adopt a fluid flow approach. More precisely, by neglecting discrete packet arrivals, we assume that the bit rate of a macro-mouse is constant and equal to the total number of bits divided by the duration of the macro-mouse. We then get the fluid approximation of the bit rate of the macro-mouse. The key point is in that since the mean arrival rate $\lambda_m \approx 178$ of macro-mice is large, the fluid bit rate $\{\Lambda_t^m\}$ of macro-mice, defined by

$$\Lambda_t^m = \sum_{j\in\mathcal{A}_t} Y_j$$

where $\mathcal{A}_t$ is the set of macro-mice active at time t and Y_j is the fluid bit rate of the jth macro-mouse, can be approximated in distribution by a Gaussian process, which auto-correlation function is perfectly known (see Appendix). The fluid bit rate over the nth time interval with length Δ is defined by

$$\tilde{\Lambda}_n^m = \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \Lambda_s^m ds. \tag{2}$$

Once we have computed the fluid bit rate process $\{\tilde{\Lambda}_n^m\}$, we can reasonably assume that discrete packet arrivals give rise to a white noise since the number of packets is very large. Thus, the actual bit rate process $\{\bar{X}_n^m\}$ of macro-mice should be equal to the fluid bit rate process $\{\tilde{\Lambda}_n^m\}$ perturbed by a white noise. To validate this approach, we use a Kalman filter to eliminate the white noise altering the actual bit rate process.

For this purpose, we compute the mean and the variance of the difference of the actual bit rate process $\{\bar{X}_n^m\}$ and the approximating process $\{\tilde{\Lambda}_n^m\}$. Experimental data give the mean and the standard deviation equal to $d_m = 372$ bit/s (negligible when compared with the mean bit rate) and $\sigma_m = 840$ Kbit/s, respectively. On the basis of these experimental mean and standard deviation, the actual bit rate has been filtered by a Kalman filter, to give the process $\{\tilde{\bar{X}}_n^m\}$. The comparison between $\{\tilde{\bar{X}}_n\}$ and the process $\{\tilde{\Lambda}_n^n\}$ is illustrated in Figure 7. It appears that the filtered bit rate and the fluid bit rate are very close one to each other and we may reasonably approximate the actual bit rate process $\{\bar{X}_n^m\}$ as

$$\bar{X}_n^m = \tilde{\Lambda}_n^m + \sigma_m \varepsilon_n$$

where $\{\varepsilon_n\}$ is a standard white noise.

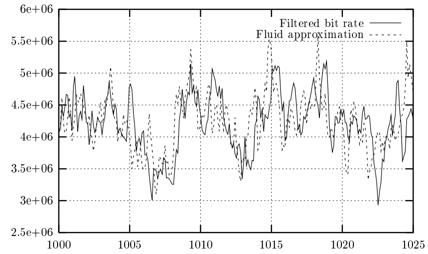


Figure 7: Filtered bit rate process $\{\tilde{\bar{X}}_n^m\}$ and fluid approximation.

Let us now consider the fluid bit rate process. Since the length Δ of the integration interval in equation (2) is small, one may expect that $\tilde{\Lambda}_n^m \sim \Lambda_{n\Delta}^m$. It follows that the auto-correlation function $c_{\tilde{\Lambda}^m}(\ell)$ of the process $\{\tilde{\Lambda}_n^m\}$, defined by

$$c_{\tilde{\Lambda}^m}(\ell) = \frac{\mathrm{cov}[\tilde{\Lambda}_n^m \tilde{\Lambda}_{(n+\ell)}^m]}{\mathrm{var}[\tilde{\Lambda}_n^m]},$$

should be close to $c_{\Lambda^m}(\ell\Delta)$, where $c_{\Lambda^m}(h)$ is the auto-correlation function of the process $\{\Lambda_t^m\}$, given by (see Appendix)

$$c_{\Lambda^m}(h) = \mathbb{E}[Y_m^2(S_m - h)^+]/\mathbb{E}[Y_m^2 S_m],$$

where Y_m and S_m denote the fluid bit rate and the duration of a macro-mouse, respectively.

Of course, the bit rate of a macro-mouse depends upon its duration. The conditional density of bit rate Y_m for different values of the duration S_m is given by Figure 8. It turns out that for a fixed value of S_m, the distribution of Y_m can be well approximated by a gamma distribution. Moreover, it is experimentally observed that $\mathbb{E}[Y_m^2 \mid S_m]$ is almost a constant. For the different values of S_m used in Figure 8, $\mathbb{E}[Y_m^2 \mid S_m] \in [3.10^8, 3.10^9]$. In the following, to simplify the computations, we assume that $\mathbb{E}[Y_m^2 \mid S_m]$ is a constant equal to $\kappa_m = 1.5.10^9$. As a consequence, we have

$$c_{\Lambda^m}(h) \sim \frac{\mathbb{E}[(S_m - h)^+]}{\mathbb{E}[S_m]} = 1 - P(1/\beta_m, (h/\eta_m)^{\beta_m}), \tag{3}$$

where we have taken into account the fact that the distribution of S_m has the form given by equation (1) and $P(a,x)$ is the incomplete Gamma function [1] defined by

$$P(a,x) = \frac{1}{\Gamma(a)} \int_0^x e^{-u} u^{a-1} du.$$

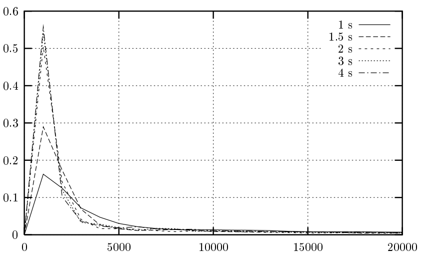
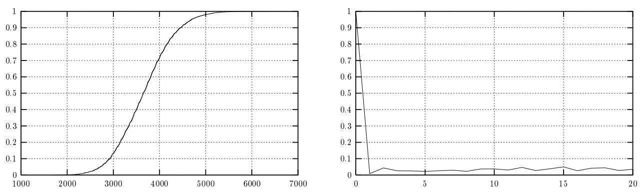


Figure 8: Conditional distribution of the bit rate process Y_m of a mouse (in bit/s) for different values of the macro-mouse duration S_m.

Finally, it remains a fraction of single packet mice, which are not included in macro-mice. The bit rate created by these mice is very small (a few tens of Kbps). The stationary distribution of this residual bit rate process $\{\hat{X}_n^m\}$ is given in Figure 9(a) and can be very well approximated by a Gaussian distribution with mean $\hat{d}_m = 3,668$ bit/s and standard deviation $\hat{\sigma}_m = 606.8$ bit/s. It can be checked by considering linear combinations of $(\hat{X}^m_{n+\ell_1}, \ldots, \hat{X}^m_{n+\ell_k})$ for arbitrary $\ell_1, \ldots, \ell_k$ that the process $\{\hat{X}_n^m\}$ is indeed Gaussian, with an autocorrelation function displayed in Figure 9(b). It turns out that there are almost no correlations and that the process $\{\hat{X}_n^m\}$ can be represented as $\hat{X}_n^m = \hat{d}_m + \hat{\sigma}_m \varepsilon_n$, where $\{\varepsilon_n\}$ is a standard white noise.

By taking into account the above results and assuming independence between the different white noise processes, we come up with the conclusion that the non p2p mouse bit rate can be represented as



(a) Stationary distribution

(b) Autocorrelation function

Figure 9: Characteristics of the residual bit rate $\{\hat{X}_n^m\}$.

$$X_n^m = \tilde{\Lambda}_n^m + \varepsilon_n \sqrt{\sigma_m^2 + \hat{\sigma}_m^2} + d_m + \hat{d}_m, \tag{4}$$

where $\{\varepsilon_n\}$ is a standard Brownian motion and $\{\tilde{\Lambda}_n^m\}$ is a Gaussian process with mean $\mathbb{E}[\tilde{\Lambda}_n^m] = 4.411$ Mbit/s, standard deviation $\sigma_{\tilde{\Lambda}^m} = 591.34$ Kbit/s, and autocorrelation function given by equation (3).

To check the above representation for the bit rate process, we compute the spectral density ψ_{X^m} of the chronological series $\{X_n^m\}$, for $n = 0, 1, 2, \ldots$ defined by $\mathrm{cov}[X_n^m, X_{n+\ell}^m] = \int_{-\pi}^{\pi} e^{i\ell x} \psi_{X^m}(x) dx$. The quantity $\mathrm{cov}[X_n^m X_{n+\ell}^m]$ is computed by averaging $X_n^m X_{n+\ell}^m$ for $n = 1, 2, \ldots, M$ where M is the total number of samples. From equation (3), we have

$$2\pi \psi_{X^m}(x) \sim \sigma_m^2 + \hat{\sigma}_m^2 + 2\pi \kappa_m \psi_{L^m}(x),$$

where ψ_{L^m} is the spectral density associated with the process $\{L^m_{n\Delta}\}$, L_t^m denoting the number of macro-mice active at time t. It then follows that for $x \in [-\pi, \pi]$, we should have

$$\psi_{X^m}(x/\Delta) \sim (\sigma_m^2 + \hat{\sigma}_m^2)/(2\pi) + \kappa_m \Delta \psi_m(x),$$

where ψ_m is the spectral density of the process $\{L_t^m\}$ and is given by Proposition 2 in the Appendix. Note that the motivation for using spectral densities is in that such functions are more robust to possible non stationary phenomena and characterize chronological series.

The comparison between $\psi_{X^m}(./\Delta)$ obtained by filtering the white noise and $\kappa_m \Delta \psi_m$ is illustrated in Figure 10. It is clearly appears that the two spectral densities are very close one to each other. This validates representation (4) for the bit rate of non p2p mice.

4 P2p mouse traffic

To describe the bit rate process of p2p mice, we proceed as in the previous section by introducing macro-mice via the aggregation of p2p mice according to some criterion and over an adequate time interval.

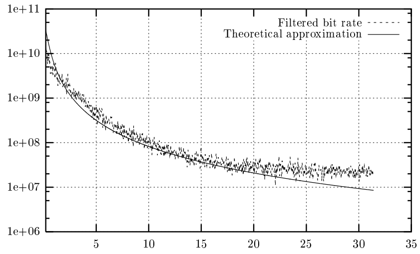


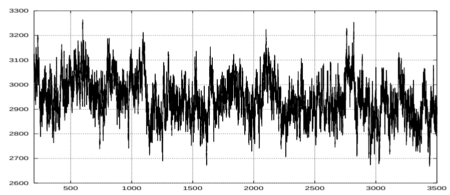
Figure 10: Comparison of spectral densities for non p2p mice.

4.1 Introduction of p2p macro-mice

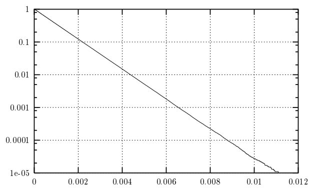
We analyze in this section the traffic offered by p2p applications, i.e., mice with a port number corresponding to a p2p protocol. Figure 11(a) represents the number of active p2p mice. The empirical complementary probability distributions of the inter-arrival time as well as the duration of p2p mice are displayed in Figures 11(b) and 11(c), respectively. While the p2p mouse inter-arrival time is remarkably exponential, it clearly appears that the process counting the number of active p2p mice is composed of "bursts" as shown in Figure 11(a). This indicates that the p2p mouse arrival process is not Poissonian. To complete the characterization of p2p mice, we note that the cdf of their duration can be approximated by a two-parameter Weibullian distribution with scale parameter $\eta = 2.8$ and skew parameter $\beta = 1.03$. Finally, the size of p2p mice is rather small (in general less than 8 packets).

As in the previous section, we are led to group mice comprising more than one packet according to some criterion. At a first glance, we may group p2p mice according to their source address. Intuitively, this criterion corresponds to the fact that a member of a p2p network seeking a content sends requests to different nodes. But this level of aggregation is not sufficient because the process counting the aggregated p2p mice on the basis their source address remains quite irregular. In fact, the search for a content and the transmission of requests give rise to response messages by the users connected to the p2p network. Hence, a second level of aggregation consists of grouping the aggregated p2p mice on the basis of their destination address.

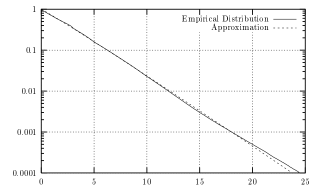
This second level of aggregation gives rise to macro p2p mice, which are composed of p2p mice with the same IP source address and/or the same destination address and arriving in a time interval of $\delta = 1$ second. Note that δ is a critical parameter since p2p mice are aggregated over time intervals of δ seconds, which should correspond to the time needed to send requests and get answers by the different members of a p2p network.



(a) Number of active p2p mice



(b) Inter-arrival time



(c) Duration

Figure 11: Characteristics of p2p mice in terms of the cdf of the duration and the inter-arrival time, and the number of active p2p mice over time.

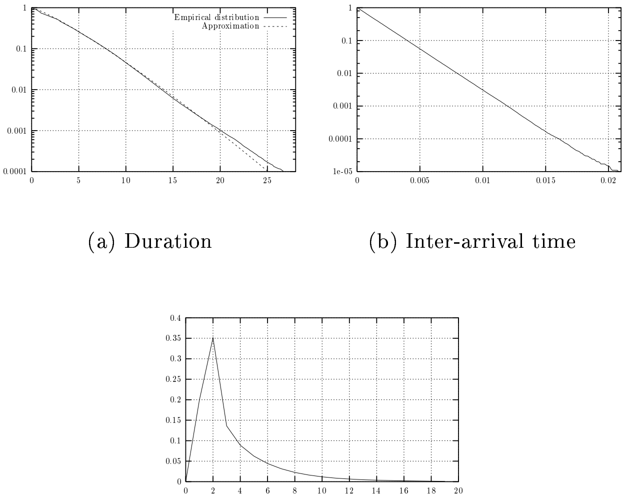
The cdf of the inter-arrival times and the duration of p2p macro-mice are given in Figure 12(b) and 12(a). It turns out that the inter-arrival time is remarkably exponential with mean $1/\lambda_\mu = 0.00174$.

Moreover, the macro-mice duration can be approximated by a two parameter Weibullian distribution with skew parameter $\beta_\mu = 1.1927$ and scale parameter $\eta_\mu = 3.898$; the mean value of the duration is $\mathbb{E}[S_\mu] = 3.55$ s (which is close to the theoretical value $\eta_\mu \Gamma(1 + 1/\beta_\mu) = 3.67$ s). Finally, the distribution of the number of p2p mice in a macro-mouse is displayed in Figure 12(c); the mean number of p2p mice in a macro-mouse is 3.43.

To check that the process counting p2p macro-mice is Poisson, we compute the distribution of the number of active p2p macro-mice at an arbitrary instant and the distribution of the number of active macro-mice at the arrival instant of a macro-mouse. These two distributions are given in Figure 13 and are very close one to each other. It follows that we may reasonably assume that the p2p macro-mouse arrival process is Poissonian.

4.2 Bit rate of p2p macro-mice

From a theoretical point of view, p2p macro mice can be described as Poisson clouds. But, for characterizing their offered bit rate, we can consider as in the previous section the fluid bit rate $\{\Lambda_t^\mu\}$ of p2p macro mice. We then assume that the exact bit rate can then be roughly approximated by the fluid bit rate perturbed by a white



(a) Duration (b) Inter-arrival time

(c) Number of mice

Figure 12: Characteristics of p2p macro-mice (cdf of duration and the inter-arrival time, and number of p2p mice in a macro-mouse).

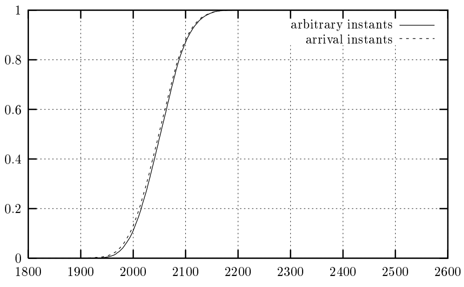


Figure 13: Distribution of the number of active p2p macro-mice at arbitrary and arrival instants.

noise; this latter white noise is due to discrete packet arrivals.

To check the above assumption, we compute the mean and the standard deviation of the difference between the bit rate $\{\bar{X}_n^\mu\}$ of p2p macro-mice and the fluid bit rate $\{\tilde{\Lambda}_n^\mu\}$ over the nth time interval with length Δ defined as in equation (2). These two quantities are equal to $d_\mu = 76.471$ Kbit/s and $\sigma_\mu = 165.86$ Kbit/s, respectively. We then use a Kalman filter to obtain the filtered bit rate $\{\tilde{\bar{X}}_n^\mu\}$. Ideally, the two processes $\{\tilde{\bar{X}}_n\}$ and $\{\tilde{\Lambda}_n + d_\mu\}$ should be very close one to each other. This is illustrated in Figure 14. Since the two processes are reasonably close one to each other, we can then approximate the process $\{\bar{X}_n^\mu\}$ as

$$\bar{X}_n^\mu = \tilde{\Lambda}_n^\mu + \sigma_\mu \varepsilon_n + d_\mu,$$

where $\{\varepsilon_n\}$ is a standard white noise.

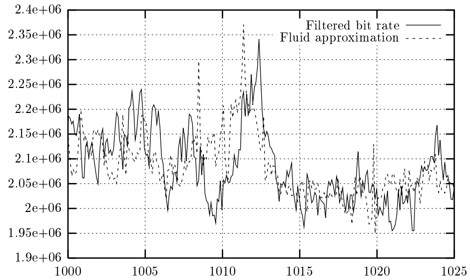


Figure 14: Comparison of the filtered bit rate process of p2p macro-mice and the associated fluid bit rate.

Finally, as in the previous section, it remains a residual bit rate due to single packet p2p mice. We admit without further details that this residual bit rate is a white noise with mean $\hat{d}_\mu \approx 4$ Kbit/s and standard deviation $\hat{\sigma}_\mu = 432.35$ bit/s.

As a consequence, in view of the above results, we have the following representation for the bit rate of p2p mice

$$X_n^\mu = \tilde{\Lambda}_n^\mu + \sqrt{\sigma_\mu^2 + \hat{\sigma}_\mu^2}\varepsilon_n + d_\mu + \hat{d}_\mu, \quad (5)$$

where $\{\varepsilon_n\}$ is a standard white noise. The process $\{\tilde{\Lambda}_n^\mu\}$ is Gaussian process with mean $\mathbb{E}[\tilde{\Lambda}_n^\mu] = 1.992$ Mbit/s and standard deviation $\sigma_{\tilde{\Lambda}^\mu} = 70.075$ Kbit/s. ($\{\tilde{\Lambda}_n^\mu\}$ is Gaussian because $\lambda_\mu \approx 574$ is large - see Appendix.)

To check the above representation, we compute the spectral density of the process $\{X_n^\mu\}$ and that of the process on the right hand side of equation (5). As in the previous section, as far the auto-correlation function is concerned, we have $\tilde{\Lambda}_n^\mu \sim \Lambda_{n\Delta}^\mu$ and then the autocorrelation function of the process $\{\tilde{\Lambda}_n^\mu\}$ can be approximated by $c_{\Lambda^\mu}(n\Delta)$, where $c_{\Lambda^\mu}(h)$ is the autocorrelation function of the process $\{\Lambda_t^\mu\}$, given by

$$c_{\Lambda^\mu}(h) = \frac{\mathbb{E}[Y_\mu^2(S_\mu - h)^+]}{\mathbb{E}[Y_\mu^2 S]},$$

Y_μ and S_μ denoting the bit rate and the duration of a p2p macro-mouse, respectively.

To compute the above auto-correlation function, we need estimate $\mathbb{E}[Y_\mu^2 \mid S_\mu]$. The function $s \to \mathbb{E}[Y_\mu^2 \mid S_\mu = s]$ is illustrated in Figure 15. To simplify, we assume in the following that $\mathbb{E}[Y_\mu^2 \mid S_\mu]$ is a constant (equal to $\kappa_\mu = 1.0e7$). Hence, for computing the auto-correlation function of the process $\{\Lambda_t^\mu\}$, we assume that $\mathbb{E}[Y^2 \mid S = s]$ is a constant (equal to κ) and we get

$$c_{\Lambda^\mu}(h) \sim \frac{\mathbb{E}[(S_\mu - h)^+]}{\mathbb{E}[S_\mu]} = 1 - P(1/\beta_\mu, (t/\eta_\mu)^{\beta_\mu});$$

the autocorrelation function of the process $\{\tilde{\Lambda}_n^\mu\}$ is approximated as $c_{\tilde{\Lambda}^\mu}(\ell) \sim c_{\Lambda^\mu}(\ell\Delta)$.

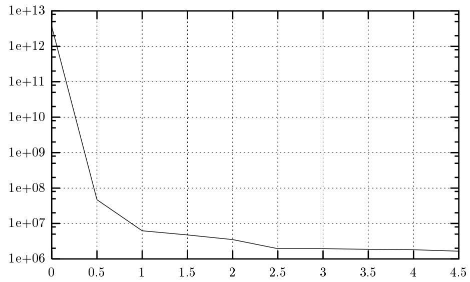


Figure 15: Graph of the function $s \to \mathbb{E}[Y^2 \mid S]$.

To verify the above approximations and representation (5) for the bit rate process of p2p mice, we compute the experimental spectral density of the process $\{X_n^\mu\}$ and the theoretical density given by

$$\frac{\sigma_\mu^2 + \hat{\sigma}_\mu^2}{2\pi} + \kappa_\mu \Delta\psi_\mu(x),$$

where $\psi_\mu(x)$ is the spectral density of the process $\{L_t^\mu\}$ describing the number of active macro-mice over time. Unfortunately, the exact spectral density $\psi_\mu(x)$ (given by equation (16) in the Appendix) is very difficult to compute numerically. Only the approximating function given by equation (18) is displayed in Figure 16. It turns out that the approximating function and the empirical spectral density are reasonably close one to each other and we may then conclude that representation (5) is valid.

5 Characteristics of elephants

In this section, we investigate the bit rate created by elephants. We proceed as in the previous sections via the

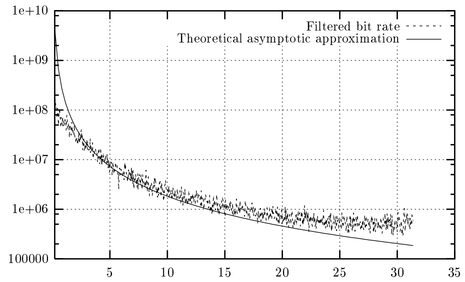


Figure 16: Spectral density of the filtered bit rate process and approximation by the function $\kappa_\mu \Delta\psi_\mu(x)$.

identification of the spectral measure associated with the bit rate process.

5.1 First observations

Figure 17 displays the global bit rate due to elephants on the link. The first observation we can make is that the bit rate of elephants is larger and much smoother than the bit rate of mice (see for instance Figure 3(a)). In fact, the bit rate of elephants is oscillating around a mean value.

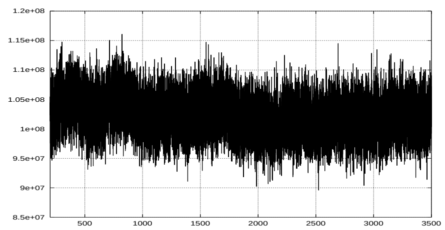


Figure 17: Bit rate created by elephants.

While it is usually assumed that elephants share the transmission capacity of a bottleneck link according to some fairness criterion (max-min or proportional fairness), we first have to draw attention to the fact that certain elephants are by nature with a very small bit rate. This is typically the case of elephants composed of ACK segments generated by a terminal retrieving data. In the case considered in this paper, since we observe downstream traffic, ACK segments correspond to data retrieved from a host connected to one of the ADSL areas connected to the IP backbone via the observed link.

To distinguish those elephants mainly composed of ACK segments, we have fixed a threshold for the mean value of the length of packets comprised in an elephant. If the mean packet length is less than 80 bytes, this certainly means that the elephant is essentially composed of ACK segments and its bit rate is small. Those elephants, referred to as ACK elephants, represent a small fraction of the global bit rate of elephants (about 1 Mbit/s against 100 Mbit/s for the global elephant bit rate).

Figure 18 represents the spectral density of the bit rate created by ACK elephants. As a first approximation, we note that the spectral density rapidly becomes a constant and thus it may reasonably be assumed that the bit rate created by ACK elephants is a white noise. This assumption is all the more motivated by the fact that the contribution of ACK elephants to the global elephant bit rate is negligible. Nevertheless, around the origin, the spectral density is not constant and this phenomenon could be due to several reasons (non stationarity, dependencies in the bit rate created by ACK elephants, or maybe long range dependence).

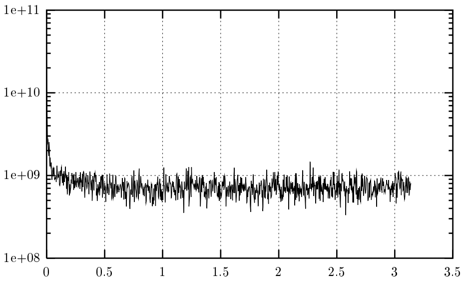


Figure 18: Spectral density of the bit rate created by elephants composed of ACK segments.

In the following, we focus our attention to "regular" elephants, that is, elephants which are not ACK elephants. When observing the dynamics of such an elephant, we note that the transmission of packets is not regular but is interrupted for a few seconds, starts again with a few packets before the transmission of a larger number of packets, and so on. This phenomenon may be due to congestion (lost packets followed by slow start periods), to the regulation of the source by the destination, etc. Hence, we are led to cut elephants into mini-elephants and elephant mice. A mini-elephant is composed of a number of packets larger than 20 packets and a mini-elephant is terminated when no packets of the mini-elephant have been observed for a time period of 20 seconds. Elephant-mice are groups of packets of an elephant, which do not belong to mini-elephants and which are composed of less than 20 packets.

5.2 Bit rate of mini-elephants

Let $\{X_n^{me}\}$ denote the bit rate process created by mini-elephants, where X_n^{me} is the number of bits due to mini-elephant over the interval $(n\Delta, (n+1)\Delta]$ divided by Δ. In a first step, we compute the fluid bit rate associated with mini-elephants, which is equal to the number of bytes contained in a mini-elephant divided by its duration. This gives rise to the continuous time fluid bit rate process $\{\Lambda_t^{me}\}$. Moreover, we compute the fluid bit rate process $\{\tilde{\Lambda}_n^{me}\}$ over the time intervals $(n\Delta, (n+1)\Delta]$ for $n = 0, 1, 2, \ldots$ as in equation (2).

We then calculate the mean and the variance of the difference between the fluid bit rate process $\{\tilde{\Lambda}_n^{me}\}$ and the actual bit rate process $\{X_n^{me}\}$; these two quantities are equal to $d_{me} = 18$ Kbit/s and $\sigma_{me} = 3$ Mbit/s, respectively. On the basis of these two values, we then use a Kalman filter to eliminate the white noise due to discrete packet arrivals and altering the actual bit rate process $\{X_n^{me}\}$; this gives rise to the filtered bit rate process $\{\tilde{X}_n^{me}\}$. The comparison between the filtered bit rate process $\{\tilde{X}_n^{me}\}$ and the fluid bit rate process $\{\tilde{\Lambda}_n^{me}\}$ is illustrated in Figure 19. It turns out that these two processes are reasonably close one to each other (the maximum relative error is about 1%).

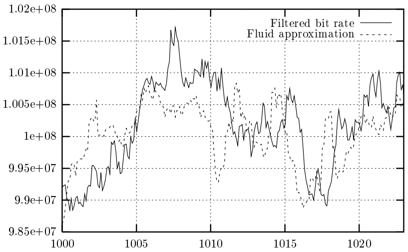


Figure 19: Fluid bit rate vs. filtered bit rate of mini-elephants.

The distribution of the size of mini-elephants (expressed in bits) is illustrated in Figure 20. It turns out that the distribution of the mini-elephant size B can be well approximated by a Pareto distribution as

$$\mathbb{P}\{B > x\} \sim a/(c+x)^b$$

with $a = 6.447e12$, $b = 2.14$ and $c = 1.66e6$; the theoretical mean $a/((b-1)c^{b-1})$ and the experimental mean are equal. This Pareto property is in line with previous measurements from IP backbone links [7].

Moreover, the distributions of the number of active mini-elephants seen at an arbitrary instant and that seen by an arriving mini-elephant are given in Figure 21. It turns that these two distributions are very close one to each other and one may reasonably assumed that the mini-elephant arrival process is Poisson, with intensity $\lambda_{me} = 26.4\ s^{-1}$.

Contrary to mice studied in the previous sections, mini-elephants, which are themselves included in elephants, are sufficiently large to be in congestion avoidance regime and one may then expect that mini-elephants

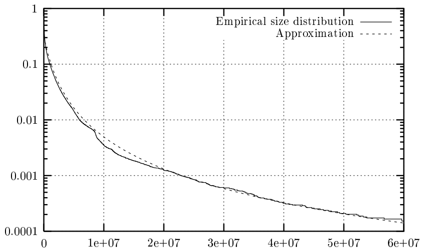


Figure 20: Distribution of the size of mini-elephants (in bits).

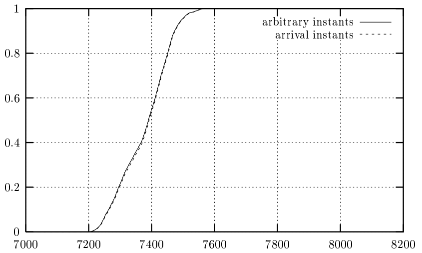


Figure 21: Probability distribution function of the number of active mini-elephants at an arbitrary instant and that seen by an arriving mini-elephant.

share the bandwidth of the network according to the TCP control loop. But, we have to draw attention to the fact that the observed link carry a large number of elephants, which may be bottlenecked somewhere else in the network, in particular in the access network. Therefore, using an $M/G/\infty$ model as in the previous section may not be so far from the reality. To check this assumption, we compute the spectral density of the bit rate process and we compare it with the theoretical one obtained by using an $M/G/\infty$ model.

By using the same arguments as in the previous sections, we have $\tilde{\Lambda}_n^{me} \sim \Lambda_{n\Delta}^{me}$. We then approximate $\Lambda_{n\Delta}^{me}$ as $\sqrt{\kappa_{me}} L_{n\Delta}^{me}$ for some $\kappa_{me} > 0$, where L_t^{me} is the number of mini-elephants active at time t. As a consequence, the spectral density $\psi_{X^{me}}$ of the bit rate process $\{X_n^{me}\}$ should be related to the spectral density $\psi_{L^{me}}$ of the process $\{L_{n\Delta}^{me}\}$ as

$$\psi_{X^{me}}(x) \sim \frac{\sigma_{me}^2}{2\pi} + \kappa_{me}\psi_{L^{me}}(x); \qquad (6)$$

experimental data show that $\kappa_{me} = 5.6e9$.

Furthermore, it turns out that the duration of mini-elephants can be well approximated by a two-parameter Weibullian distribution with the scale parameter $\eta_{me} = 0.466$ and skew parameter $\beta_{me} = 126.88$, as shown in Figure 22. Now, by using the $M/G/\infty$ assumption, we can approximate $\psi_{L^{me}}$ by ψ_{me} given by equation (15) for the parameters λ_{me}, β_{me} and η_{me}. We then come up with the approximation

$$\psi_{X^{me}}(x/\Delta) \sim \sigma_{me}^2/(2\pi) + \kappa_{me}\Delta\psi_{me}(x). \qquad (7)$$

This approximation is illustrated in Figure 23, where we have considered the filtered bit rate; this amounts to removing the term $\sigma_{me}^2/(2\pi)$ in equation (7). Numerical evidence shows that the spectral densities are in good agreement and hence, the $M/G/\infty$ model can be used to described the superposition of mini-elephants. This is certainly due to the fact that the observed link is actually not overloaded or that mini-elephants are constrained somewhere else in the network.

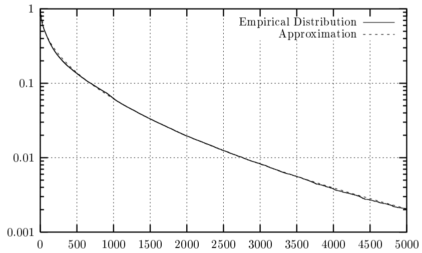


Figure 22: Cdf of the duration of mini-elephants.

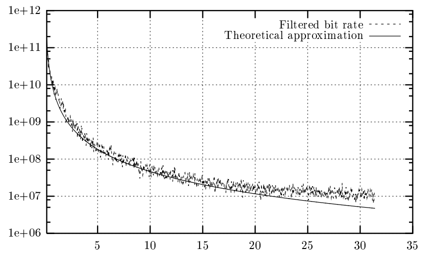


Figure 23: Comparison between the spectral density of the filtered bit rate process $\{X_n\}$ and $\kappa_{me}\Delta\mu_L^{me}$.

5.3 Bit rate of elephant mice

In this section, we deal with elephant mice. Such a mouse is a group of packets of an elephant, which do belong to a sustained transmission phase of the elephant. To describe the bit rate process $\{X_n^{em}\}$, we proceed as for other mice. The distribution of the duration of elephant mice can be approximated by a two parameter Weibullian distribution with scale parameter $\eta_{em} = 16.94$ and skew parameter $\beta_{em} = 0.76$. Moreover, by computing the distributions of active elephant mice at an arbitrary instant and at arrival instant, we can show that the elephant mouse occurrence process can be assumed to be Poisson.

By using the same technique as for mice and in particular a Kalman filter with mean $d_{em} = 45$ Kbit/s and variance $\sigma_{em} = 180$ Kbit/s, we can approximate the spectral density of the bit rate process created by elephant mice as

$$\psi_{X^{em}}(x/\Delta) = \sigma_{em}^2/(2\pi) + \kappa_{em}\Delta\psi_{em}(x), \quad (8)$$

with $\kappa_{em} = 6e8$, $\sigma_{em} = 180$ Kbit/s and $\psi_{em}(x)$ given by equation (15) with $\lambda_{em} = 23.35$ s^{-1}, $\eta_{em} = 16.94$ and $\beta = 0.76$. This approximation is illustrated in Figure 24, where the spectral density of the filtered elephant mice bit rate process and the approximation $\kappa_{em}\Delta\psi_{em}(x)$ are displayed. For elephant mice, we have taken a sampling period $\Delta = 500$ ms in order to obtain a sufficient amount of data generated by mice elephant over each sampling interval. The above approximation, even though not very accurate, is reasonable.

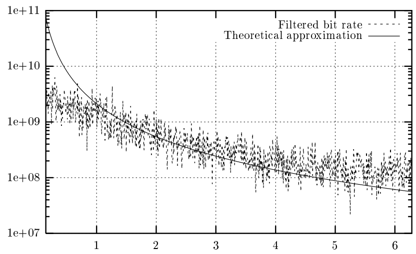


Figure 24: Spectral density of the filtered elephant mice bit rate process and the approximation $\kappa\Delta\mu(x)$.

5.4 Synthesis of the global bit rate

From the previous sections, we come up with the following representation of the bit rate process $\{X_n^e\}$ of elephants:

$$X_n^e = \tilde{\Lambda}_n^{me} + X_n^{ack} + \tilde{\Lambda}_n^{em} + \sigma_e\varepsilon_n + d_e, \quad (9)$$

where

- $\{X_n^{ack}\}$ is the bit rate process of ACK elephants, that we can assume in a first approximation to be of the form $d_{ack} + \sigma_{ack}\varepsilon_n$, where $\{\varepsilon_n\}$ is a standard white noise, $\sigma_{ack} \approx 71.6$ kbit/s and $d_{ack} = 1$ Mbit/s,
- $\{\tilde{\Lambda}_n^{me}\}$ is the fluid bit rate of mini-elephants, which is Gaussian with mean 99.65 Mbit/s and spectral density $\kappa_{me}\Delta\psi_{me}(\Delta x)$,
- $\{\tilde{\Lambda}_n^{em}\}$ is the Gaussian process associated with the fluid bit rate of elephant mice, with mean 963.6 Kbit/s and spectral density $\kappa_{em}\Delta\psi_{em}(\Delta x)$,
- $\{\varepsilon_n\}$ is a standard white noise and

$$\sigma_e = \sqrt{\sigma_{em}^2 + \sigma_{me}^2},$$

- d_e is a constant equal to $d_{em} + d_{me} = 63$ Kbit/s.

More generally, when we consider the global bit rate $\{X_n\}$, we can approximate $\{X_n\}$ as the superposition of several Gaussian processes and a white noise; each of these Gaussian processes describe one component of global traffic: p2p mice, non p2p mice, elephant mice and mini-elephants. We specifically have

$$X_n = \tilde{\Lambda}_n + \sigma\varepsilon_n,$$

where $\{\varepsilon_n\}$ is standard white noise, $\{\tilde{\Lambda}_n\}$ is a Gaussian process with mean about 108 Mbit/s and with spectral density $\Delta\psi(\Delta x)$, with

$$\psi(x) = \kappa_m\psi_m(x) + \kappa_\mu\psi_\mu(x) + \kappa_{me}\psi_{me}(x) + \kappa_{em}\psi_{em}(x) \quad (10)$$

and

$$\sigma = \sqrt{\sigma_m^2 + \hat{\sigma}_m^2 + \sigma_\mu^2 + \hat{\sigma}_\mu^2 + \sigma_{me}^2 + \sigma_{em}^2 + \sigma_{ack}^2} \sim \sigma_{me}.$$

It is worth noting that while mini-elephants offer the prevalent part of traffic, no term in the sum on the r.h.s. of equation (10) is dominant. In particular, $\kappa_m\psi_m$ and $\kappa_{me}\psi_{me}$ are comparable, especially around the origin. Hence, when computing the global spectral density of the global bit rate $\{X_n\}$, adequate filtering should be performed in order to recover the relevant traffic characteristics (for instance the arrival intensities of mice and elephants as well as the mean bit rate experienced by elephants). This point will be addressed in further studies.

6 Conclusion

We have analyzed in this paper TCP traffic delivered by an IP backbone network to several ADSL areas. One salient feature of the observed traffic traces is in that a significant part of global traffic is due to p2p applications, which exhibit some remarkable properties, in particular with regard to the number of mice, which are permanently exchanged between the different members of a peer-to-peer network.

It is possible to decompose traffic into several components on the basis of the mice/elephants dichotomy. By analyzing each component separately and by adopting an adequate level of aggregation, it is possible to describe each component by means of a Gaussian process perturbed by a white noise. In the description of each component, it is quite remarkable that the duration of flows can be well approximated by a two parameter Weibullian distribution, even in the case of mini-elephants in spite of the fact that their volume of information follows a Pareto distribution.

A Proofs of technical results

A.1 Calculation of autocorrelation functions

Let t_n be the nth point of a Poisson process with intensity λ. Consider a sequence $\{(Y_n, S_n)\}$ of i.i.d couples of random variables taking values in $\mathbb{R}_+ \times \mathbb{R}_+$; note that Y_n and S_n may be correlated. Let $\{\Lambda_t\}$ be the process defined by

$$\Lambda_t = \sum_{n=0}^{\infty} Y_n \mathbb{1}_{\{t_n \le t < t_n + S_n\}}.$$

Note that $\Lambda_t = \mathcal{N}(f_t)$ where $\mathcal{N}$ is a Poisson process with intensity $\lambda du \otimes \mathbb{P}_{(Y,S)}(dy, ds)$, with $\mathbb{P}_{(Y,S)}(dy, ds)$ denoting the joint probability density function of the couple (Y_1, S_1), and where f is defined by

$$f_t(u, y, s) = y \mathbb{1}_{A_t}(u, s)$$

with

$$A_t = \{(u, s) : u \le t,\ u + s > t\}.$$

Then, the following lemma is an elementary result on Poisson processes.

Lemma 1 *If $\mathcal{N}$ is a Poisson process with intensity μ, then for all measurable functions f and g*

$$\mathbb{E}\big(e^{-s\mathcal{N}(f)}\big) = \exp\left(-\int (1 - e^{-sf}) d\mu\right)$$

and

$$\operatorname{cov}\big(\mathcal{N}(f), \mathcal{N}(g)\big) = \int fg d\mu.$$

The expression of the Laplace transform of $\mathcal{N}(f)$ given by this lemma is well-known (see [12] for example) and the expression of the covariance is derived by the same method. Using this lemma, we have the following result.

Proposition 1 *The Laplace transform of the random variable Λ_t is given by*

$$\mathbb{E}\left(e^{-s\Lambda_t}\right) = \exp\left(-\lambda\mathbb{E}\big((1 - e^{-sY})(t \wedge S)\big)\right)$$

and in the stationary regime, the Laplace transform of Λ_∞ is given by

$$\mathbb{E}\left(e^{-s\Lambda_\infty}\right) = \exp\left(-\lambda\mathbb{E}\big((1 - e^{-sY})S\big)\right)$$

Moreover, we have

$$\operatorname{cov}\big(\Lambda_t, \Lambda_{t+h}\big) = \lambda\mathbb{E}\big(Y^2\big(t \wedge (S - h)^+\big)\big)$$

Proof. By definition, $\Lambda_t = \mathcal{N}(f_t)$. Thus, by Lemma 1 and using the expression of the intensity μ of $\mathcal{N}$, the Laplace transform $\mathbb{E}[\exp(-s\Lambda_t)]$ is given by

$$\exp\left(-\int (1 - e^{-sy})\mathbb{1}_{\{t-u \in [0,s]\}} \lambda du \otimes \mathbb{P}_{(Y,S)}(dy, ds)\right).$$

Fubini's formula and the elementary fact that for two real numbers a and b, $(a-b)^+ = a - a \wedge b$ gives the expression of $\mathbb{E}\left(e^{-s\Lambda_t}\right)$. It converges to $\exp\big(-\lambda\mathbb{E}\big(S(1 - e^{-sC})\big)\big)$ by Monotone convergence Theorem as t tends to infinity. Then, by Lemma 1,

$$\operatorname{cov}\big(\Lambda_t, \Lambda_{t+h}\big) = \int f_t f_{t+h} d\mu.$$

But

$$f_t f_{t+h} = y^2 1_{A_t \cap A_{t+h}}(u, s)$$

where $A_t \cap A_{t+h} = \{(u, s), u \le t, u + s > t + h\}$. We conclude as for the Laplace transform. This ends the proof. □

Corollary 1 *We have the following convergence results: when $t \to \infty$,*

$$\mathbb{E}\big(\Lambda_t\big) = \lambda\mathbb{E}\big(Y(t \wedge S)\big) \to \mathbb{E}\big(\Lambda_\infty\big) = \lambda\mathbb{E}\big(YS\big), \tag{11}$$

$$\operatorname{var}\big(\Lambda_t\big) = \lambda\mathbb{E}\big(Y^2(t \wedge S)\big) \to \operatorname{var}\big(\Lambda_\infty\big) = \lambda\mathbb{E}\big(Y^2 S\big), \tag{12}$$

and

$$\frac{\operatorname{cov}\big(\Lambda_t, \Lambda_{t+h}\big)}{\operatorname{var}\big(\Lambda_t\big)} = \frac{\mathbb{E}\big(\big(Y^2(t \wedge (S-h)^+)\big)\big)}{\mathbb{E}\big(Y^2(t \wedge S)\big)} \to c_\Lambda(h) \stackrel{def}{=} \frac{\mathbb{E}\big(Y^2(S-h)^+\big)}{\mathbb{E}\big(Y^2 S\big)}. \tag{13}$$

A.2 Heavy traffic results

We study the behavior of the process $\{\Lambda_t\}$ when time is rescaled by a factor n. In other words, the arrival process has a rate $n\lambda$. Define the mean function $m(t)$ as

$$m(t) = \lambda\mathbb{E}\big(Y(t \wedge S)\big).$$

Then, we have the following result, which is due, in a more general case, to Borovkov [3] and Iglehart [10].

Theorem 1 *Under the assumption $\mathbb{E}(Y^2S) < \infty$, the process $\{\hat{\Lambda}_t^n\}$ defined by*

$$\hat{\Lambda}_t^n = \frac{\Lambda_t^n - nm(t)}{\sqrt{n}}$$

converges in distribution to the Gaussian process $\{\hat{\Lambda}_t\}$ characterized by the covariance function

$$\gamma(t, t+h) = \lambda\, \mathbb{E}\big(Y^2(t \wedge (S-h)^+)\big).$$

It is worth noting that the centered Gaussian process $\{\hat{\Lambda}_t\}$ has the same autocorrelation function as the process the original process $\{\Lambda_t\}$. This is this last property, which is used in this paper in order to identify the limiting process $\{\hat{\Lambda}_t\}$.

B Computation of the spectral density ψ

Let ψ be the spectral density of the process $\{L_t\}$ describing the number of customers in an $M/G/\infty$ queue in the stationary regime, where the intensity of the input Poisson process is denoted by λ and where service times are Weibullian with scale parameter η and skew parameter β. From the previous section, we know that the autocovariance function of the process $\{L_t\}$ is given for $h \in \mathbb{R}$ by

$$\operatorname{cov}(L_t, L_{t+h}) = \frac{\lambda\eta}{\beta}\Gamma\left(\frac{1}{\beta}, \left(\frac{|h|}{\eta}\right)^{\beta}\right),$$

where $\Gamma(a,x) = \int_x^\infty u^{a-1}e^{-u}du$. The spectral density ψ is then defined as

$$\int_{-\infty}^{\infty} e^{ihx}\psi(x)dx = \frac{\lambda\eta}{\beta}\Gamma\left(\frac{1}{\beta}, \left(\frac{|h|}{\eta}\right)^{\beta}\right). \qquad (14)$$

Proposition 2 *The spectral density $\psi(x)$ is given by*

$$\psi(x) = \frac{\lambda\eta^2}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\frac{\Gamma(n\beta+1)}{(x\eta)^{n\beta+2}}\cos\left(n\beta\frac{\pi}{2}\right) \qquad (15)$$

for $\beta \in (0,1]$ and

$$\psi(x) = \frac{\lambda\eta^2}{\pi\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}(\eta x)^{2n}\Gamma\left(\frac{2n+2}{\beta}\right) \qquad (16)$$

for $\beta > 1$.

Proof. Taking derivatives in the definition (14) of ψ, we have

$$\int_{-\infty}^{\infty} ixe^{ihx}\psi(x)dx = -\operatorname{sgn}(h)\lambda e^{-(|h|/\eta)^\beta},$$

where $\operatorname{sgn}(h)$ is the sign of h. This entails that by Fourier inverse transform formula

$$x\psi(x) = \frac{\lambda}{\pi}\int_0^\infty \sin(hx)e^{-(h/\eta)^\beta}dh. \qquad (17)$$

It turns out that the above Fourier inverse function is known in the literature to belong to the class of general singular integral of Weierstrass [8]. Modulo a few adaptations, it is easy to derive equations (15) and (16). $\square$

Note that when $\beta < 1$ and $x \to \infty$, it is readily checked from equation (15) that $\psi(x) \sim \lambda/(\pi x^2)$. This result also holds when $\beta > 1$; as shown by the following result.

Proposition 3 *Under the assumption $\beta > 1$, when $x \to \infty$,*

$$\psi(x) \sim \lambda/(\pi x^2). \qquad (18)$$

Proof. Let us consider the integral

$$J(x) = \int_0^\infty e^{ixz}e^{-(\frac{z}{\eta})^\beta}dz.$$

From equation (17), we have, for real x, $x\psi(x) = \lambda\left(J(x) - \overline{J(x)}\right)/2i\pi$. Via an integration by part, we can easily show that $J(x) \sim i/x$ when $\xi \to \infty$ and equation (18) follows. $\square$

References

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions.* National Bureau of Standards, Applied Mathematics Series 55, 1972.

[2] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. The multiscale nature of network traffic. *IEEE Signal Processing Magazine*, 2002.

[3] A. A. Borovkov. Limit laws for queueing processes in multichannel systems. *Sibirsk. Mat. Ž.*, 8:983–1004, 1967.

[4] P. Brémaud and R. Mazumdar. Event and time averages: a review. *Adv. Appl. Prob.*, 24:377–411, 1992.

[5] J. Cao and K. Ramanan. A Poisson limit for buffer overflow probabilities. In *Proc. Infocom 2002*, New York, June 2002.

[6] K. Claffy, G. Miller, and K. Thompson. The nature of the beast: Recent traffic measurement from an Internet backbone. In *Proc. of Inet*, 1998.

[7] M. Crovella and A. Bestravos. Self-similarity in world wide web. Evidence and possible causes. *IEEE/ACM Trans. on Networking*, pages 835–846, December 1997.

[8] A. Erdélyi. *Higher transcendal functions.* Mc Graw Hill, 1953.

[9] A. Feldmann, A.C. Gilbert, W. Willinger, and T. Kurtz. The changing nature of network traffic: Scaling phenomena. In *Computer Communication Review*, volume 28, pages 5–19, 1998.

[10] Donald L. Iglehart. Weak convergence of compound stochastic process. I. *Stochastic Processes Appl.*, 1:11–31; corrigendum, ibid. 1 (1973), 185–186, 1973.

[11] W. Leland, M..Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. *IEEE/ACM Trans. Net.*, pages 1–15, 1994.

[12] J. Neveu. Processus ponctuels. In *École d'Été de Probabilités de Saint-Flour, VI—1976*, pages 249–445. Lecture Notes in Math., Vol. 598. Springer-Verlag, Berlin, 1977.

[13] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE J. Sel. Areas Commun.*, 13(6), August 1995.

[14] K. Park and W. Willinger, editors. *Self similar network traffic and performance evaluation.* Wiley Interscience, 2000.

[15] V. Paxson and S. Floyd. Wide area traffic: The failure of the Poisson assumption. *IEEE/ACM Trans. on Networking*, pages 226–244, 1995.

[16] J. Lévy Véhel and R. Riedi. *Fractals in Engineering*, chapter Fractional Brownian motion and data traffic modeling: The other end of the spectrum. Springer, 1997.

[17] Z.L.Zhang, V. Ribeiro, S. Moon, and C. Diot. Small time scaling behavior of Internet backbone traffic: An empirical study. In *Proc. Infocom 2003*, 2003.