

# An Empirical Study on Time-Correlation of GSM Telephone Traffic

Stefano Bregni, *Senior Member, IEEE*, Roberto Cioffi, and Maurizio Decina, *Fellow, IEEE*

**Abstract**—In this paper, we investigate possible time-correlation of answered call arrivals in sets of real GSM telephone traffic data. Instead of attempting to model the empirical distribution of the call interarrival time, as done in several previous works in literature, we emphasize results obtained by the Modified Allan Variance (MAVAR), a widely used time-domain quantity with excellent capability of discriminating power-law noise types. The call arrival rate exhibits a diurnal trend, with peak hours in the morning and late afternoon. Besides this diurnal rate change, the number of call arrivals in a second is found perfectly uncorrelated to the number of arrivals in other seconds and Poisson distributed, with good consistency by  $\chi^2$ -test evaluation. Uniform and accurate whiteness of call arrivals per second is verified in all hours, regardless the time of the day. In all series analyzed, the empirical statistics of both originated and terminated call arrivals proves excellent consistency with the ideal Poisson model with diurnal variable rate  $\lambda(t)$ . This study may be valuable to researchers concerned about realistic traffic modelling in planning and performance evaluation of cellular networks.

**Index Terms**—GSM, modified Allan variance, Poisson random process, traffic measurement (communication), traffic model.

## I. INTRODUCTION

CLASSIC theory of telephone traffic (a.k.a. teletraffic) was developed when networks were exclusively wired, based on analog Frequency Division Multiplexing (FDM), and the sole offered service was the so-called Plain Old Telephone Service (POTS). Telephone traffic theory and algorithms for dimensioning resources in circuit-switched networks have been developed since the Sixties, without significant changes thereon [1]–[4].

Wireless mobile telephony introduced a new scenario, in which some previous assumptions of traffic statistics may not hold anymore, given the peculiar behaviour of mobile users. Then, the impressive growth of mobile user population yielded the need of characterizing specifically mobile telephone traffic. Algorithms for dimensioning network resources do rely on faithful statistical characterization of traffic. Hence, there is a

Manuscript received January 23, 2007; revised April 3, 2007 and July 7, 2007; accepted August 19, 2007. The associate editor coordinating the review of this letter and approving it for publication was Y.-B. Lin. This paper is largely based, but not exclusively, on “An Empirical Study on Statistical Properties of GSM Telephone Call Arrivals,” by S. Bregni, R. Cioffi, and M. Decina, which appeared in the *Proc. of IEEE GLOBECOM 2006*, San Francisco, CA, USA, Nov. 2006. Work partially funded by the Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR), Italy, under PRIN project MIMOSA. Politecnico di Milano, Dept. of Electronics and Information.

The authors are with Politecnico di Milano, Dept. of Electronics and Information, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy (e-mail: {bregni, decina}@elet.polimi.it).

Digital Object Identifier 10.1109/TWC.2008.070092.

need for telecommunications operators and traffic engineers, for accurate statistical models of mobile telephone traffic.

In classic telephone traffic theory, developed for wired networks, call arrivals to a local exchange are usually modelled as a Poisson process, at least over short observation intervals to assume stationary arrival rate, since the user population served by the exchange is very large and with negligible correlation among users. This assumption of memoryless traffic has been often retained also in presence of mobile users: in literature, incoming calls in cellular networks are mostly modelled as a Poisson process, with both call holding time and interarrival time assumed with negative exponential distribution.

Nevertheless, it has been argued that this Poisson assumption might not be valid in wireless cellular networks for a number of reasons. First, cells partition the user population in small sets, each served by a small number of channels and with possible correlation between users. Moreover, congestion and repeated call attempts, more likely with radio access impairments, are major causes of peaks and bursts in offered traffic and of levelling off the carried traffic. Finally, user mobility during calls (handover) adds further complexity to the problem. Therefore, not surprisingly, traffic characterization in wireless cellular networks has been attracting much attention in literature since few years.

In most cases, researchers focused on characterizing the channel holding time or the call holding time, sometimes based on empirical data. In many studies, the channel holding time has been modelled by negative exponential distribution. Nevertheless, several other works contradicted this simple assumption. In papers [5], [6], for example, the probability distribution that better fits empirical data, by the Kolmogorov-Smirnov test, was found to be a sum of lognormal distributions.

On the other hand, the channel holding time is affected by user mobility, characterized by the cell residence time. With exponentially-distributed call holding time, the merged traffic from new and handoff calls is Poissonian if and only if the cell residence time is exponentially distributed too [7]. For cell residence time having general distribution, the channel holding time distribution was derived analytically in [8], [9]. The channel holding time distribution was also studied in [10], when the cell residence time has Erlang or Hyper-Erlang distribution.

As for the correlation between call arrival times, the distributions of the channel idle time (time between the end of an answered call and the beginning of the next one on the same channel) and of the call interarrival time in a Public Access Mobile Radio (PAMR) cellular system were investigated in

[11]. In that work, the former distribution was approximated by the Erlang- $j,k$  function and the latter resulted different from Poissonian negative exponential.

Recently, a further empirical study on real GSM telephone traffic data was reported in [12]. Answered call holding and interarrival times were found to be best modelled by the lognormal-3 function, rather than by the Poissonian negative exponential distribution.

In summary, several studies contradicted the ubiquitous likelihood of the classic Poisson model for telephone traffic in cellular networks and suggested that call arrivals may be significantly time-correlated, due for example to access congestion, user mobility and possible correlation between nearby users. However, we note that the Poisson traffic model is still assumed in almost all works, mainly for the sake of simplicity, when cellular network performance is evaluated. Questions may arise, therefore, on the practical relevance of this simplifying assumption.

In this paper, we analyze a few sets of real GSM telephone traffic data, collected by the Italian mobile telecommunications operator Telecom Italia Mobile (TIM) to billing purposes, which were already studied in [12]. Instead of attempting to model the empirical distributions of the call holding time and of the call interarrival time, as done in [12] and in most previous works on this subject, we investigate possible time-correlation of call arrivals directly in the time domain by means of the Modified Allan Variance (MAVAR), one of the most sensitive tools to this aim and with demonstrated excellent capability of discriminating power-law noise, only recently introduced for network traffic analysis [13], [14].

This paper is organized as follows. In Section II, basic MAVAR properties are briefly recalled for ease of understanding. In Section III, the traffic measurement data and the analysis criteria are described. In Sections IV and V, traffic analysis results are presented and a statistical model is identified. Finally, in Section VI, some conclusions are drawn.

## II. THE MODIFIED ALLAN VARIANCE

The MAVAR was originally conceived for frequency stability characterization of precision oscillators in the time domain [15]-[19]. It was proposed in 1981 by modifying the definition of the two-sample variance (a.k.a. Allan variance), recommended to this purpose by IEEE since 1971 [20], [21]. MAVAR was designed with the goal of discriminating noise types with power-law spectrum of kind  $f^\alpha$  ( $\alpha \in \mathcal{R}$ ,  $\alpha > -5$ ) recognized very commonly in frequency sources. Since then, it has been widely used in clock stability characterization and it has been adopted in telecommunications international standards too [19], [22].

Recently, MAVAR has been also proposed as analysis tool of self-similar and long-range dependent (LRD) traffic [23]. It has been demonstrated to feature superior accuracy in estimation of the Hurst parameter  $H$  and of the exponent  $\alpha$ , coupled with good robustness against nonstationarity in data analyzed [13], [14]. MAVAR has been successfully applied to real Internet traffic analysis, allowing to identify fractional noise in experimental results [13], [14], [24].

This section briefly recalls some basic MAVAR properties most relevant to our aim. For details and demonstration of all

statements, the interested reader is referred to the bibliography cited; in particular, [19], [13], [14] may be suggested as first readings.

### A. Definition

Given an infinite sequence  $\{x_k\}$  of samples of an input signal  $x(t)$ , evenly spaced in time with sampling period  $\tau_0$ , MAVAR is defined as

$$\text{Mod } \sigma_y^2(\tau) = \frac{1}{2n^2\tau_0^2} \left\langle \left[ \frac{1}{n} \sum_{j=1}^n (x_{j+2n} - 2x_{j+n} + x_j) \right]^2 \right\rangle \quad (1)$$

where  $\tau = n\tau_0$  is the observation interval and the operator  $\langle \bullet \rangle$  denotes infinite-time averaging.

In practice, given a finite set of  $N$  samples  $x_k$  spaced by  $\tau_0$  over a measurement interval  $T = (N-1)\tau_0$ , a MAVAR estimate can be computed using the ITU T standard estimator [19], [22]

$$\text{Mod } \sigma_y^2(n\tau_0) = \frac{\sum_{j=1}^{N-3n+1} \left[ \sum_{i=j}^{n+j-1} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2}{2n^4\tau_0^2(N-3n+1)} \quad (2)$$

with  $n = 1, 2, \dots, \lfloor N/3 \rfloor$ . A recursive algorithm for fast computation exists [19], which cuts down the complexity of evaluating MAVAR for all  $\lfloor N/3 \rfloor$  values of  $n$  to  $O(N^2)$  instead of  $O(N^3)$ .

The point estimate (2), computed by averaging  $N-3n+1$  terms, is a random variable itself. Exact computation of confidence intervals is not immediate and, annoyingly enough, depends on the spectrum of the underlying noise [25]-[29]. However, in general, confidence intervals are not negligible only for longest  $\tau$ , where few terms are averaged. In our analysis, we avoided to consider last MAVAR values computed for largest  $n$  (the right end of curves).

### B. Application to Estimation of Fractional Noise Parameters

As customary in characterization of phase and frequency noise in precision oscillators [18], [19], [20], [26], [30], we deal with random processes  $x(t)$  with one-sided power spectral density (PSD) modelled as

$$S_x(f) = \begin{cases} \sum_{i=1}^P h_{\alpha_i} f^{\alpha_i} & 0 < f \leq f_h \\ 0 & f > f_h \end{cases} \quad (3)$$

where  $P$  is the number of noise types considered in the model,  $\alpha_i$  and  $h_{\alpha_i}$  are parameters ( $\alpha_i, h_{\alpha_i} \in \mathcal{R}$ ) and  $f_h$  is the upper cut-off frequency. Such random processes are commonly referred to as *power-law noise* or *fractional noise*. Note that  $x(t)$  is not necessarily assumed Gaussian in this model.

Power-law noise with  $-4 \leq \alpha_i \leq 0$  has been revealed in practical measurements of various physical phenomena, including phase noise of precision oscillators [18], [19], [20], [26], [30] and Internet traffic [13], [14], [24], [31], whereas  $P$  should be limited to few units for the model being meaningful. Although values  $\alpha_i \leq -1$  yield model pathologies, such as infinite variance and even nonstationarity, this model is

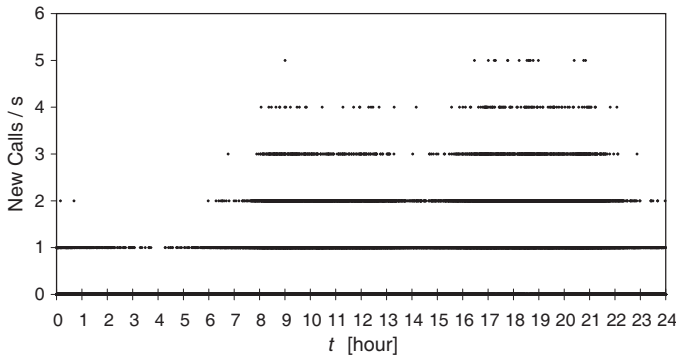


Fig. 1. Originated call arrivals  $\{x_{Ok}\}$  (24 Oct. 2003,  $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

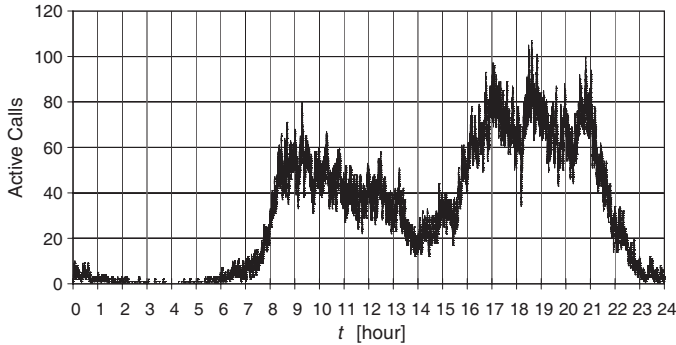


Fig. 2. Originated active calls  $\{n_{Ok}\}$  (24 Oct. 2003,  $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

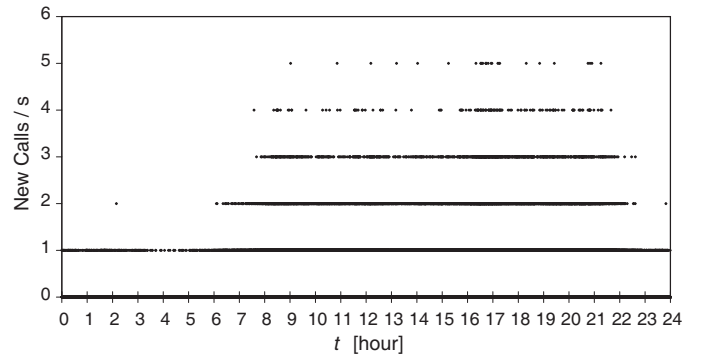


Fig. 3. Originated call arrivals  $\{x_{Ok}\}$  (23 Jan. 2004,  $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

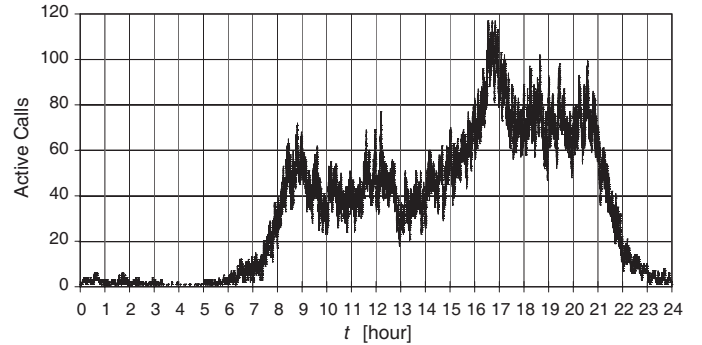


Fig. 4. Originated active calls  $\{n_{Ok}\}$  (23 Jan. 2004,  $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

commonly used, considering also that real-world constraints imply measurement finite bandwidth and duration.

Under this general hypothesis of power-law spectrum, first we point out that the infinite-time average in definition (1) converges for  $\alpha_i > -5$  (being MAVAR the variance of a moving average of data second difference). Then, by considering separately each term of the sum in (3) and letting  $P = 1$ ,  $\alpha = \alpha_i$  ( $-5 < \alpha \leq 0$ ), MAVAR is found to obey a simple power law of the observation interval  $\tau$  (ideally asymptotically for  $n \rightarrow \infty$ , keeping constant  $n\tau_0 = \tau$ , but in practice for  $n > 4$ ), i.e.

$$\text{Mod } \sigma_y^2(\tau) \sim A_\mu \tau^\mu \quad (4)$$

where  $\mu = -3 - \alpha$  [13], [14], [18], [19]. If  $P > 1$ , it is immediate to generalize (4) to summation of powers  $A_{\mu_i} \tau^{\mu_i}$ .

Therefore, if  $x(t)$  obeys (3) and assuming sufficient separation between components, a log-log plot of MAVAR looks ideally as a broken line made of  $P$  straight segments, whose slopes  $\mu_i$  give the exponent estimates  $\alpha_i = -3 - \mu_i$  of the power-law noise components prevailing in distinct ranges of  $\tau$ .

In papers [23], [13], [14], these estimates of  $\alpha_i$  were demonstrated to be very accurate, even better than those obtained by the Daubechies' wavelet logscale diagram technique [32], which is one of the best reputed and most widely adopted methods for analyzing LRD traffic. Moreover, nonstationary components of various kinds in the analyzed sequence (viz. polynomial drifts, periodic trends and steps) affect MAVAR negligibly or in a well recognizable way [13], [14]. In particular, data offset and linear drift are cancelled in MAVAR

results, being it based on data second difference. Therefore, we chose MAVAR as main tool to analyze traffic traces.

### III. GSM TRAFFIC DATA SETS AND ANALYSIS CRITERIA

We analyzed sets of real GSM telephone traffic data, collected in a Mobile Switching Centre (MSC) by the operator Telecom Italia Mobile (TIM) for billing and traffic monitoring, which were already studied in [12]. Data recorded are the initial (arrival) time and duration of GSM answered voice calls, originated or terminated in RM82D1, a wide-area cell located in Fiumicino (30 km from Rome, Italy). Time scale is discrete, with 1-second intervals, as standard in network management systems (i.e., events are cumulated throughout each second).

Unfortunately, unanswered call attempts (i.e., due to called party unavailability or busy line, which may account for 30%-40% of the total call attempts) could not be recorded. Moreover, no information was recorded to trace user mobility between cells. However, we notice that the large cell size makes handover sporadic.

The exclusion of all unsuccessful repeated call attempts, the negligible impact of handovers and the seeming absence of congestion, as observable forth in Figs. 2 and 4, would lead to expect that call arrivals in data analyzed tend to behave as Poisson events, in contrast with results [12], where the distribution of call interarrival times is estimated to be not negative-exponential. On the other hand, since the cell where telephone traffic was recorded is located close to the most important Italian airport and to an important highway, it

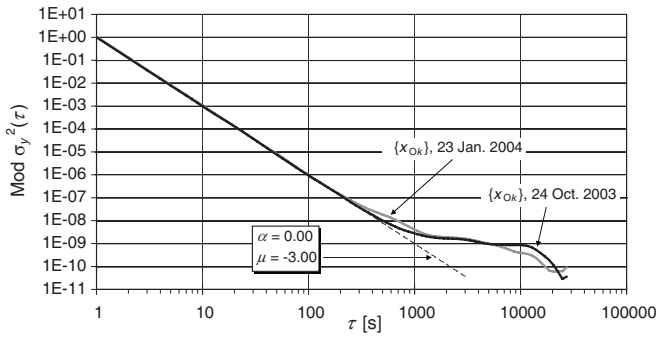


Fig. 5. MAVAR computed on sequences  $\{x_{Ok}\}$  of 24 Oct. 2003 and 23 Jan. 2004 ( $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

would be well-founded to conjecture some correlation between calls as well. Therefore, since modelling distributions of interarrival times in empirical data can be thorny, although commonly pursued, we sought possible time-correlation of call arrivals directly in the time domain by MAVAR, owing to its superior sensitivity and robustness to spurious components (e.g., trends) in experimental data.

Traffic data analyzed were collected continuously over the 24 hours in 6 distinct days: 18 July 2003, 11 Aug. 2003, 24 Oct. 2003, 26 Oct. 2003, 23 Jan. 2004 and 25 Jan. 2004. To summarize, we analyzed  $6 \times 2$  traffic files (originated and terminated calls), each one further divided in 24 segments (one per hour), listing answered calls with arrival time and duration (time quantized to 1 s). Processing this raw traffic data, we produced four new sets of  $24 \times 6$  traffic data series, namely for a given time frame:

- $\mathbf{x}_O$  and  $\mathbf{x}_T$ , whose items  $x_{Ok}$  and  $x_{Tk}$  are the number of originated and terminated, respectively, new calls in the  $k$ -th second (i.e., the answered call arrival rate in the  $k$ -th second);
- $\mathbf{n}_O$  and  $\mathbf{n}_T$ , whose items  $n_{Ok}$  and  $n_{Tk}$  are the number of originated and terminated, respectively, simultaneous active calls in the  $k$ -th second.

Then, we sought possible time-correlation of call arrivals in the time domain, by computing MAVAR on time series  $\{x_{Ok}\}$  and  $\{x_{Tk}\}$ . Moreover, we evaluated the applicability of a Poisson model to describe the statistical properties of the call arrival process.

#### IV. ANALYSIS RESULTS

All traffic sequences  $\mathbf{x}_O$  and  $\mathbf{x}_T$  examined are strongly nonstationary and exhibit a diurnal pseudoperiodic average trend, with peak hours in the morning and late afternoon. For example, Figs. 1, 2 and Figs. 3, 4 plot the originated call arrivals  $\{x_{Ok}\}$  (new calls/s) and simultaneous active calls  $\{n_{Ok}\}$ , recorded on 24 Oct. 2003 and 23 Jan. 2004, respectively (one black pixel plotted for each data value). Figs. 2 and 4 show no evidence of congestion. Similar trends were observed in all other days.

##### A. Study of Time-Correlation of Call Arrivals

We computed MAVAR on the  $6 \times 2$  sequences  $\{x_{Ok}\}$  and  $\{x_{Tk}\}$  over the whole measurement period  $T = 24$  h ( $N =$

86400,  $\tau_0 = 1$  s). For instance, Fig. 5 plots MAVAR computed on the same sequences of Figs. 1 and 3 (originated call arrivals  $\{x_{Ok}\}$  recorded on 24 Oct. 2003 and 23 Jan. 2004).

By inspection of Fig. 5, we notice that  $\text{Mod } \sigma_y^2(\tau)$  is almost perfectly linear (in the log-log plot) for  $\tau$  up to 200 s or 500 s, with slope  $\mu \cong -3.0$  corresponding to  $\alpha \cong 0.0$ . This means that in the short term (i.e., on observation intervals up to few hundreds of seconds), the deviation of the data sequence from a linear trend is purely *random white* with excellent approximation (i.e., with no memory), whilst average drifts of order two and above are negligible (data offset and linear drift are cancelled anyway in MAVAR results).

For  $\tau$  longer than  $\sim 500$  s or so, conversely,  $\text{Mod } \sigma_y^2(\tau)$  departs from the linear trend, capturing the diurnal pseudoperiodic variation of the arrival rate evident in Figs. 1 and 3. For long  $\tau$ , the higher slope  $\mu$  reflects the slower wander of the data sequence averaged in the long term. Moreover, we should also consider the poor statistical confidence of  $\text{Mod } \sigma_y^2(\tau)$  values for longest  $\tau$ .

In all 24 hours sequences, for both originated and terminated calls, we observe the same behaviour with little variation. In all cases,  $\text{Mod } \sigma_y^2(\tau)$  is linear for  $\tau < 10^2 \div 10^3$  s, with slope  $\mu \cong -3.0$  ( $\alpha \cong 0.0$ , maximum deviation  $\pm 0.02$ , evaluated by linear regression). Therefore, the number of call arrivals in a second has been always found perfectly uncorrelated to the number of arrivals in other seconds. Non-negligible time-correlation may be found only averaging on long observation intervals (say, at least 500 s), due to the diurnal variation of the average arrival rate.

##### B. Poisson Model of Call Arrivals

From such results, it comes natural to infer that both originated and terminated call arrivals may be modelled as a classic (non-homogeneous) Poisson random process with slowly variable arrival rate  $\lambda(t)$ , which follows a diurnal trend such as that in Figs. 1 and 3. Given the data set of a particular day,  $\lambda(t)$  can be estimated (e.g., by a moving average or some more sophisticated method) and becomes, therefore, a deterministic term in the Poisson model of the random arrival process. However, the pseudoperiodic arrival rate  $\lambda(t)$  changes randomly day by day, although to a limited extent. Statistical characterization of  $\lambda(t)$  is beyond the scope of this paper and, however, would need measuring traffic data over many more days (i.e., years) to be significant.

These MAVAR results ensure the absence of correlation between samples of time series  $\{x_{Ok}\}$  and  $\{x_{Tk}\}$ , but give no insight on intervals shorter than  $\tau_0 = 1$  s. If the call arrival process is ideal Poisson, on the other hand, then time correlation is null even on infinitesimal intervals and, in particular, the number of arrivals  $x_k$  in  $\tau$  seconds is a random variable distributed as

$$P(x_k = i) = \frac{(\lambda\tau)^i}{i!} e^{-\lambda\tau} \quad (5)$$

with both mean  $m_x$  and variance  $\sigma_x^2$  equal to  $\lambda\tau$ .

Hence, we studied the distribution of samples  $x_{Ok}$  and  $x_{Tk}$  in all traffic series. Due to the severe nonstationarity of sequences, the distribution of the  $N = 84000$  samples over

TABLE I  
MEAN, VARIANCE AND  $\chi^2$ -TEST (TO SAME-MEAN POISSON DISTRIBUTION) OF SUBSEQUENCES  $\{x_{Ok}\}$  RECORDED ON 24 OCT. 2003 AND 23 JAN. 2004 ( $\tau_0 = 1$  s).

Time Interval		$m_x$	$\frac{\sigma_x^2}{m_x}$	$\frac{\sigma_x^2}{m_x}$	$Q(\chi^2)$
24 Oct. 2003					
0.00 - 24.00	86400	0.334	0.408	1.221	0
9.00 - 14.00	18001	0.467	0.468	1.001	0.9610
9.00 - 11.00	7201	0.558	0.540	0.968	0.7988
11.00 - 13.00	7201	0.472	0.462	0.978	0.9979
16.00 - 18.00	7201	0.680	0.700	1.029	0.9475
18.00 - 20.00	7201	0.720	0.739	1.026	0.5262
23 Jan. 2004					
0.00 - 24.00	86400	0.353	0.431	1.220	0
9.00 - 14.00	18001	0.476	0.483	1.015	0.9706
9.00 - 11.00	7201	0.489	0.495	1.013	0.9833
11.00 - 13.00	7201	0.506	0.514	1.015	0.9998
16.00 - 18.00	7201	0.783	0.799	1.021	0.7593
18.00 - 20.00	7201	0.640	0.641	1.001	0.99997

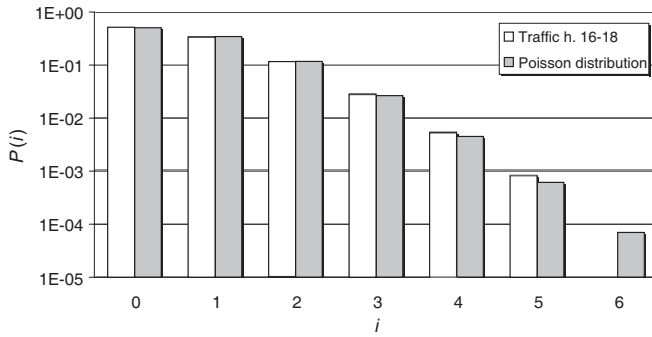


Fig. 6. Distribution of samples  $x_{Ok}$  (16.00-18.00, 24 Oct. 2003,  $T = 2$  h,  $N = 7201$ ,  $\tau_0 = 1$  s) compared to the Poisson distribution with same mean.

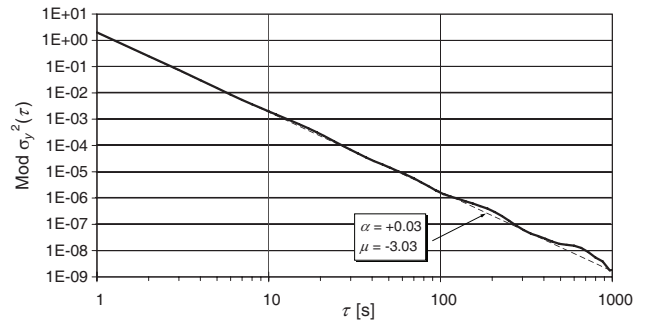


Fig. 7. MAVAR of sequence  $\{x_{Ok}\}$  (19.00-20.00 24 Oct. 2003,  $T = 1$  h,  $N = 3600$ ,  $\tau_0 = 1$  s).

the whole period ( $T = 24$  h) does *not* obey (5) ( $\tau = 1$  s). Also their mean and variance are different.

The right approach, by good practice in telephone traffic engineering, is restricting evaluation of statistics to peak hours, where stationarity holds at best and the number of calls per second is maximum (quantization effects are minimum, as evident in Figs. 1 and 3). Thus, we computed mean, variance and distribution of samples  $x_{Ok}$  and  $x_{Tk}$ , in all six days, separately in four peak-hour intervals 9.00-11.00, 11.00-13.00, 16.00-18.00 and 18.00-20.00 ( $N = 7201$ ). In all cases, mean and variance resulted nearly equal ( $\sigma_x^2/m_x = 1 \pm 0.04$ ) and the distribution very close to (5).

For example, numerical results obtained on subsequences  $\{x_{Ok}\}$  of 24 Oct. 2003 and 23 Jan. 2004 are summarized in Table 1 (cf. Figs. 1 and 3). In Fig. 6, moreover, the normalized distribution of samples  $x_{Ok}$  in interval 16.00-18.00 of 24 Oct. 2003 is compared to the Poisson distribution (5) having same mean. The two distributions match very accurately, even where confidence is little.

To have a quantitative measure of how the Poisson dis-

tribution and the empirical distributions of samples  $x_{Ok}$  and  $x_{Tk}$  do match, we evaluated the standard *chi-square test* [33]. In peak-hour intervals of all days, the values of the chi-square probability  $Q(\chi^2)$  are above 0.75.<sup>1</sup> Precise values are summarized in Table 1 as well.

In conclusion, our analysis of traffic data, by both MAVAR and  $\chi^2$  test evaluation, demonstrates excellent consistency between the empirical statistics of new call arrivals and the ideal non-homogeneous Poisson model with variable rate  $\lambda(t)$ .

### C. More about Stationarity: Whiteness of Call Arrivals

Since all traffic sequences exhibit a diurnal trend in the average arrival rate, we investigated whether also the whiteness of short-term random fluctuations (cf. Sec. IV-A) is

<sup>1</sup>A small value of  $Q(\chi^2)$  indicates a significant difference between distributions. In other words, it *disproves* (to a certain level of significance) the null hypothesis that two data sets are drawn from the same population distribution function. Disproving the null hypothesis, in effect, proves that the data sets are from different distributions. Failing to disprove the null hypothesis, on the other hand, only shows that the data sets can be *consistent* with a single distribution function. See [33] for details.

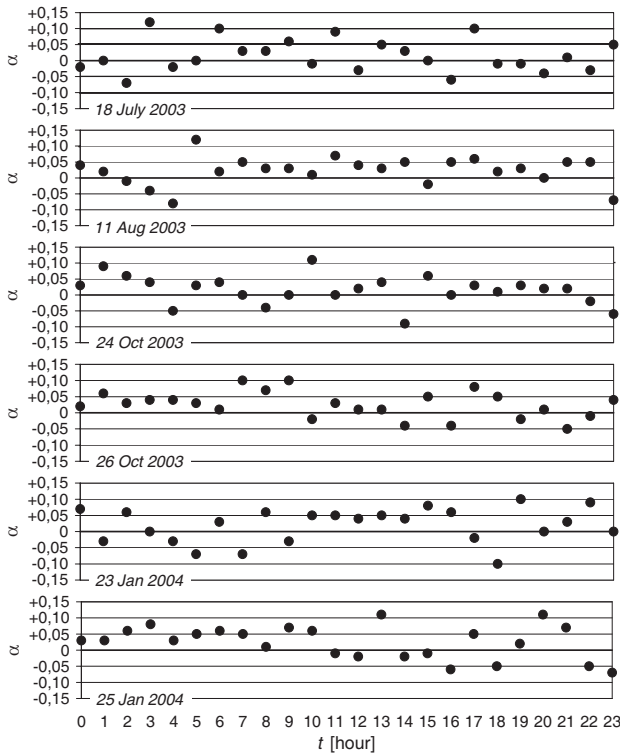


Fig. 8. Values of  $\alpha$  over the 24 hours, estimated by MAVAR in originated call arrivals/s sequences  $\{x_{Ok}\}$  ( $T = 1$  h,  $N = 3600$ ,  $\tau_0 = 1$  s,  $\tau \leq 100$  s).

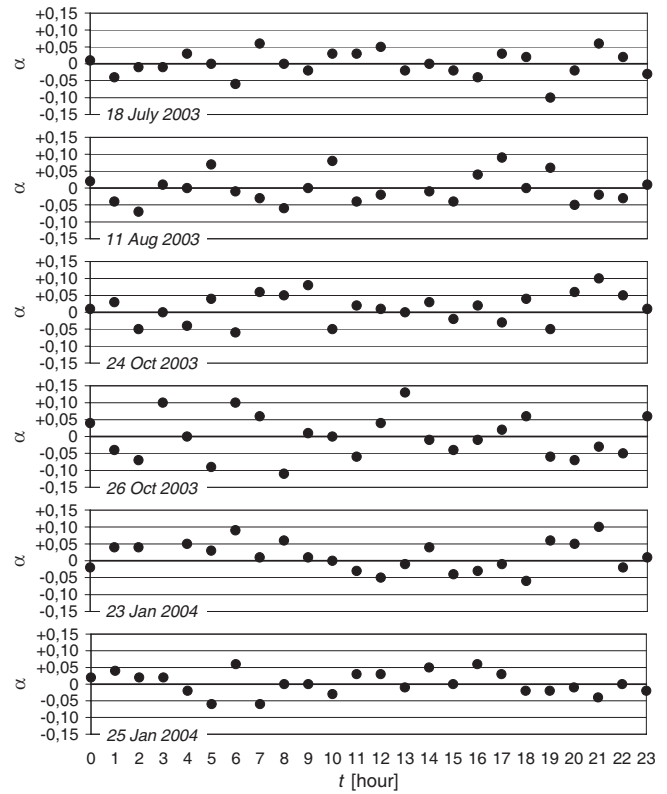


Fig. 9. Values of  $\alpha$  over the 24 hours, estimated by MAVAR in terminated call arrivals/s sequences  $\{x_{Tk}\}$  ( $T = 1$  h,  $N = 3600$ ,  $\tau_0 = 1$  s,  $\tau \leq 100$  s).

affected similarly. To this aim, we analyzed by MAVAR the 1 hour traffic data series separately. From each of the  $24 \times 6 \times 2$  subsequences, we estimated the exponent  $\alpha$  of the underlying fractional noise dominant in the short term, by linear regression of  $\text{Mod } \sigma_y^2(\tau)$  on interval  $\tau \leq 100$  s.

For example, Fig. 7 plots MAVAR computed on one of those subsequences (originated call arrivals/s, recorded on 19.00-20.00, 24 Oct. 2003,  $T = 1$  h,  $N = 3600$ ,  $\tau_0 = 1$  s). As expected,  $\text{Mod } \sigma_y^2(\tau)$  follows approximately a linear trend on its whole length ( $\mu \cong -3.03$ ,  $\alpha \cong +0.03$  for  $\tau \leq 100$  s). Compared to 24-hours results (cf. Fig. 5), we notice that here the line is slightly more irregular, due to less confidence.

The sequences of values  $\{\alpha_{i,j}\}$  ( $i = 1, 2, \dots, 6; j = 1, 2, \dots, 24$ ), estimated in each hour of the 6 days, are plotted in Figs. 8 and 9, for originated and terminated calls, respectively. We notice, on the one hand, that no periodicity is evident and, on the other, that estimated values  $\alpha_{i,j}$  have small uncertainty around their mean  $m_\alpha \cong 0$ . Therefore, we conclude that call arrival sequences  $\{x_{Ok}\}$  and  $\{x_{Tk}\}$  result uniformly and accurately white in all hours, regardless the time of the day.

D. Simultaneous Active Calls

We repeated the analysis of Sec. IV-A on the sequences  $\{n_{Ok}\}$  and  $\{n_{Tk}\}$ , i.e. on the number of simultaneous active calls ( $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s). For instance, Fig. 10 plots MAVAR computed on the same sequences of Figs. 2 and 4 (originated active calls  $\{n_{Ok}\}$  recorded on 24 Oct. 2003 and 23 Jan. 2004). Moreover, Fig. 11 plots the PSD estimated on the sequence of Fig. 2 (24 Oct. 2003). Also in this case, as obvious, a similar behaviour was observed in all days.

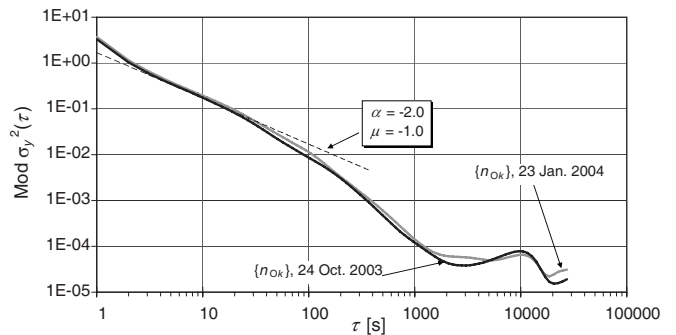


Fig. 10. MAVAR computed on sequences  $\{n_{Ok}\}$  of 24 Oct. 2003 and 23 Jan. 2004 ( $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

The random process  $n(t)$  (continuous-time version of sequence  $\{n_k\}$ ) is given by the summation of rectangles starting at each call arrival and having random length (call duration). Analysis of this problem is outside the scope of this paper, but is equivalent to the study of the number of active servers in an  $M/G/\infty/0$  queue (telephone multiplexer). If calls begin as Poisson events and have random duration distributed as a negative exponential, the PSD of  $n(t)$  (neglecting its continuous component) results having form

$$S_n(f) = \frac{A}{B + f^2} \tag{6}$$

as derived in the Appendix.

Observing Figs. 10 and 11, we notice that the experimental results do match this analytical model, although the call duration distribution is different than negative-exponential [12].

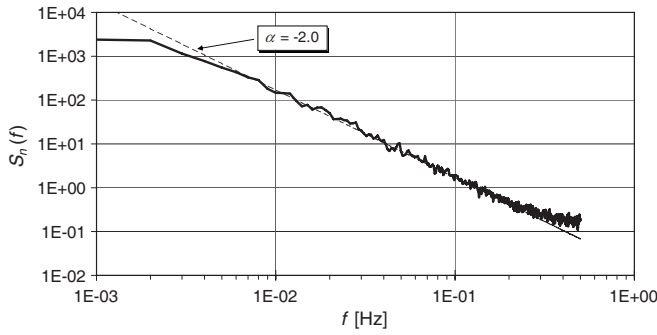


Fig. 11. PSD estimated as periodogram (Hann windowing on segments of 1000 data points) on sequence  $\{n_{Ok}\}$  of 24 Oct. 2003 ( $T = 24$  h,  $N = 86400$ ,  $\tau_0 = 1$  s).

In Fig. 11, the PSD behaves as  $\sim 1/f^2$  for  $f > 10^{-2}$  Hz. In Fig. 10, the trend of  $\text{Mod } \sigma_y^2(\tau)$  is somehow consistent with such a PSD, if we notice that:

- in the short term ( $\tau < 10^2$  s), the average slope is approximately  $\mu \cong -1$ , corresponding to  $\alpha \cong -2$ ;
- in the long term ( $\tau > 10^2$  s), besides the effect of diurnal wander for  $\tau > 10^3$  s, the average slope increases, due to flattening of  $S_n(f)$  for  $f \rightarrow 0$  (cf. eq. (6));
- the slope of  $\text{Mod } \sigma_y^2(\tau)$  for  $n < 4$  may be not significant (Sec. II-B).

## V. CONCLUSIONS

In literature, several studies contradicted the ubiquitous likelihood of the classic Poisson model for telephone call arrivals in wireless cellular networks, due for example to access congestion, user mobility and possible correlation between nearby users. Nevertheless, this traffic model is still commonly assumed in most works, mainly for the sake of simplicity, when cellular network performance is evaluated.

In this paper, we investigated possible time-correlation of both originated and terminated answered call arrivals in sets of real GSM telephone traffic data. Results obtained by MAVAR time-domain analysis have been emphasized, owing to its superior sensitivity and capability of discriminating power-law noise types, to overcome the thorniness of modelling the distribution of call interarrival times in empirical data. The main findings can be summarized as follows.

- All traffic sequences examined are strongly nonstationary. The call arrival rate exhibits a diurnal pseudoperiodic trend, with peak hours in the morning and late afternoon (see Figs. 1, 2, 3, 4).
- Besides the diurnal variation of the average arrival rate, the number of call arrivals in a second has been found perfectly uncorrelated to the number of arrivals in other seconds (Secs. IV-A and IV-C).
- Restricting evaluation of statistics to short time intervals ( $\sim 1$  h) or better to peak hours, to ensure stationarity of the arrival rate, the number of call arrivals in a second has been found having same mean and variance ( $\sigma_x^2/m_x = 1 \pm 0.04$ ) and distributed, with good consistency by  $\chi^2$ -test evaluation, as the ideal Poisson probability distribution (Sec. IV-B).

- Uniform and accurate whiteness of call arrivals per second has been verified in all hours, regardless the time of the day (Sec. IV-C).

In conclusion, call arrivals proved excellent consistency, by MAVAR analysis and  $\chi^2$ -test evaluation, with the non-homogeneous Poisson model with diurnal variable rate  $\lambda(t)$ , as expected considering that handover and congestion effects are negligible in experimental data analyzed. On the contrary, attempting to model empirical distributions of call interarrival times is difficult and may give ambiguous results (cf. also results [12] on same experimental data).

These results confirm, at least to the limited extent of these empirical data, that the Poisson model is still adequate to describe realistically telephone traffic in cellular networks, unless focusing specifically on particular issues as small user population, access congestion and very frequent handovers.

## APPENDIX

### PSD OF PROCESS $n(t)$

Let the process  $n(t)$  be defined as

$$n(t) = \sum h_T(t - t_j) \quad (7)$$

where  $t_j$  are Poisson instants with average rate  $\lambda$  and  $h_T(t)$  is a pulse response dependent on the random variable  $T$  (shot noise). Its PSD is given by [34]

$$S_n(f) = \lambda \int |H_T(f)|^2 p(T) dT \quad (8)$$

where  $p(T)$  is the probability density function of  $T$ .

If  $h_T(t)$  is the unitary rectangle in interval  $(0, T)$  and  $p(T)$  is the negative-exponential function with mean  $1/\mu$ , then

$$S_n(f) = \lambda \int_0^\infty \left( \frac{\sin \pi f T}{\pi f} \right)^2 \mu e^{-\mu T} dT = \frac{2\lambda}{\mu^2 + 4\pi^2 f^2} \quad (9)$$

having neglected the continuous component  $\lambda/\mu$ .

## ACKNOWLEDGEMENT

The authors wish to thank Telecom Italia Mobile (TIM), for providing the traffic data analyzed in this study, Achille Pattavina, Marco Ferrari and Sandro Bellini, for his help on the Appendix.

## REFERENCES

- [1] R. Syski, *Introduction to Congestion Theory in Telephone Systems*, 2nd ed. Amsterdam, The Netherlands: Elsevier Science Publishers B.V., 1986.
- [2] R. I. Wilkinson, "Theories for toll traffic engineering in the USA," *Bell System Tech. J.*, vol. 35, no. 2, pp. 421–514, Mar. 1956.
- [3] C. W. Pratt, "The concept of marginal overflow in alternate routing," in *Proc. 5th International Teletraffic Congress (ITC)*, 1967.
- [4] S. Katz, "Statistical performance analysis of a switched communications network," in *Proc. 5th International Teletraffic Congress (ITC)*, 1967.
- [5] F. Barcelò and J. Jordan, "Channel holding time distribution in public telephony system (PAMR and PCS)," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1615–1625, Sept. 2000.
- [6] C. Jedrzycky and V. C. M. Leung, "Probability distribution of channel holding time in cellular telephony system," in *Proc. IEEE Veh. Technol. Conf.*, May 1996.
- [7] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Trans. Comput.*, vol. 47, no. 6, pp. 679–692, June 1998.

- [8] Y. Fang, "Hyper-Erlang distributions and traffic modeling in wireless and mobile networks," in *Proc. IEEE Wireless Commun. Networking Conference (WCNC)*, Sept. 1999.
- [9] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling of PCS networks," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1062–1072, July 1999.
- [10] J. A. Barria and B. H. Soong, "A Coxian model for channel holding time distribution for teletraffic mobility modelling," *IEEE Commun. Lett.*, vol. 4, no. 12, pp. 402–404, Dec. 2000.
- [11] F. Barcelò and S. Bueno, "Idle and inter-arrival time statistics in public access mobile radio (PAMR) system," in *Proc. IEEE Globecom*, Nov. 1997.
- [12] A. Pattavina and A. Parini, "Modelling voice call interarrival and holding time distributions in mobile networks," in *Proc. 19th International Teletraffic Congress (ITC)*, Aug. 2005.
- [13] S. Bregni and L. Jmoda, "Improved estimation of the Hurst parameter of long-range dependent traffic using the modified Hadamard variance," in *Proc. IEEE ICC*, 2006.
- [14] S. Bregni and L. Jmoda, "Accurate estimation of the Hurst parameter of long-range dependent traffic using modified Allan and Hadamard variances," to be published in *IEEE Transactions on Communications*.
- [15] D. W. Allan and J. A. Barnes, "A modified Allan variance with increased oscillator characterization ability," in *Proc. 35th Annual Freq. Contr. Symp.*, 1981.
- [16] P. Lesage and T. Ayi, "Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate," *IEEE Trans. Instrum. Meas.*, vol. 33, no. 4, pp. 332–336, Dec. 1984.
- [17] L. G. Bernier, "Theoretical analysis of the modified Allan variance," in *Proc. 41st Annual Freq. Contr. Symp.*, 1987.
- [18] D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, eds., "Characterization of clocks and oscillators," NIST Technical Note 1337, Mar. 1990.
- [19] S. Bregni, Chapter 5: Characterization and Modelling of Clocks, in *Synchronization of Digital Telecommunications Networks*. Chichester, UK: John Wiley & Sons, 2002, pp. 203–281.
- [20] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen Jr., W. L. Smith, R. L. Sydnor, R. F. C. Vessot, and G. M. R. Winkler, "Characterization of frequency stability," *IEEE Trans. Instrum. Meas.*, vol. 20, no. 2, pp. 105–120, May 1971.
- [21] "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology," IEEE Std. 1139, approved 1988, revised 1999.
- [22] ITU T Rec. G.810 "Definitions and Terminology for Synchronisation Networks," Rec. G.811 "Timing Characteristics of Primary Reference Clocks," Rec. G.812 "Timing Requirements of Slave Clocks Suitable for Use as Node Clocks in Synchronization Networks," Rec. G.813 "Timing Characteristics of SDH Equipment Slave Clocks (SEC)," Geneva 1996-2003.
- [23] S. Bregni and L. Primerano, "The modified Allan variance as time-domain analysis tool for estimating the Hurst parameter of long-range dependent traffic," in *Proc. IEEE GLOBECOM*, 2004.
- [24] S. Bregni and W. Erangoli, "Fractional noise in experimental measurements of IP traffic in a metropolitan area network," in *Proc. IEEE GLOBECOM*, 2005.
- [25] P. Lesage and C. Audoin, "Characterization of frequency stability: uncertainty due to the finite number of measurements," *IEEE Trans. Instrum. Meas.*, vol. 22, no. 2, pp. 157–161, June 1973. "Comments on '—'," *IEEE Trans. Instrum. Meas.*, vol. 24, no. 1, p. 86, Mar. 1975. "Correction to '—'," *IEEE Trans. Instrum. Meas.*, vol. 25, no. 3, p. 270, Sept. 1976.
- [26] S. R. Stein, "Frequency and time—their measurement and characterization," in *Precision Frequency Control*, vol. 2, E. A. Gerber and A. Ballato, eds. New York: Academic Press, 1985, pp. 191–232.
- [27] C. A. Greenhall and W. J. Riley, "Uncertainty of stability variances based on finite differences." [Online]. Available: <http://www.wiley.com>
- [28] C. A. Greenhall, "Recipes for degrees of freedom of frequency stability estimators," *IEEE Trans. Instrum. Meas.*, vol. 40, no. 6, pp. 994–999, Dec. 1991.
- [29] W. J. Riley, "Confidence intervals and bias corrections for the stable32 variance functions," Hamilton Technical Services, 2000. [Online]. Available: <http://www.wiley.com>
- [30] J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress," *Proc. IEEE*, vol. 66, no. 9, pp. 1048–1075, Sept. 1978.
- [31] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "The multiscale nature of network traffic," *IEEE Signal Processing Mag.*, vol. 19, no. 3, pp. 28–46, May 2002.
- [32] P. Abry and D. Veitch, "Wavelet analysis of long-tange dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no.1, pp. 2–15, Jan. 1998.
- [33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Chapter 14: Statistical Description of Data, in *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge University Press, 2002, pp. 609–655.
- [34] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed. New York: McGraw Hill, 1984.



**Stefano Bregni** (M'93-SM'99) was born in Milano, Italy, in 1965. In 1990, he graduated (Laurea and Master degrees) in telecommunications engineering at Politecnico di Milano. Since 1991, he worked on SDH and network synchronization issues, with special regard to clock stability measurement, first with SIRTI S.p.A (1991-1993) and then with the CEFRIEL consortium (1994-1999). He served on ETSI and ITU-T committees on digital network synchronization. In 1999, he joined Politecnico di Milano as a tenured Assistant Professor, becoming Associate Professor in 2004. He teaches telecommunications networks. He is the author of about 60 papers, mostly in IEEE conferences and journals, and of the books *Synchronization of Digital Telecommunications Networks* (John Wiley & Sons, 2002; translated and published in Russia by MIR Publishers, 2003) and *Sistemi di Trasmissione PDH e SDH - Multiplazione* (i.e. "PDH and SDH Transmission Systems-Multiplexing," McGraw-Hill, Italy, 2004). His current research interests focus mainly on traffic modelling and optical networks. Prof. Bregni is a Distinguished Lecturer of the IEEE Communications Society. He is Director of Education, Chair of the Transmission, Access and Optical Systems (TAOS) Technical Committee and voting member of the Globecom/ICC Technical Content (GITC) committee of the IEEE Communications Society. He is Symposia Chair of IEEE Globecom 2009 and Symposium Chair in six other IEEE Globecom and ICC conferences. He is Editor of IEEE Global Communications Newsletter and an Associate Editor of *IEEE Communications Surveys and Tutorials*. He was tutorial lecturer in four IEEE Globecom and ICC conferences.



**Roberto Cioffi** was born in Milano, Italy, in 1983. In 2008, he graduated in telecommunications engineering at Politecnico di Milano with a thesis on traffic analysis and characterization.



**Maurizio Décina** (M'73-SM'85-F'87) is Professor of telecommunications at the Politecnico di Milano (Milan, Italy), Department of Electronics and Information. A 1943 native of Pescasseroli, Italy, Décina received the Dott. Ing. degree in electronic engineering from the University of Rome, Italy, in 1966.

He began his career at the Ugo Bordoni Foundation and at the SIP/Telecom Italia Headquarters in Rome, where he was involved in the introduction of PCM transmission facilities. In 1976 he joined the University of Rome, where he worked on the design of customer access to the ISDN. From 1983 to 1987 he was the executive Director of R&D of Italtel in Milan, and contributed to the development of the Linea UT digital switching system. In 1988 he joined the Politecnico di Milano, where he was the founder and Director of the CEFRIEL research and education consortium until 2003, and has been active in ATM and IP switching research. In the early 80s and 90s he worked also as a scientific consultant for A&T Bell Laboratories in Naperville, Illinois, USA, in the field of exploratory broadband packet voice and photonic switching.

Prof. Décina was the President of the IEEE Communications Society for the years 1994-1995. In 1986 he was appointed Fellow of the IEEE, in 1997 he received the IEEE Award in International Communications, and in 2000 the IEEE Third Millennium Medal Award. Prof. Décina was a non-executive Member of the Board of Directors of several information and communication technology companies such as: Telecom Italia, Italtel, and Tiscali. He was also the founder of some start-ups, such as ICT Consulting and Securmatrics.