

Analysis of CMPP Approach in Modeling Broadband Traffic

R.G. Garroppo, S. Giordano, S. Lucetti, and M. Pagano

Department of Information Engineering, University of Pisa
Via Diotisalvi 2 - 56126 Pisa - Italy
{r.garroppo, s.giordano, s.lucetti, m.pagano}@iet.unipi.it

Abstract. The CMPP (Circulant Modulated Poisson Process) modeling approach represents an appealing solution since it provides the integration of traffic measurement and modeling. At the same time, it maintains the Markovian hypothesis that permits analytical transient and steady-state analyses of queueing systems using efficient algorithms. These relevant features of CMPP approach has driven us to analyze in more details the fitting procedure when it is applied to actual broadband traffic. In the paper, investigating the estimation algorithm of model parameters, we emphasize the difficulty of CMPP in capturing the upper tail of marginal distribution of actual data, which leads to an optimistic evaluation of network performance. As shown in the paper, a simple relation exists between the number of significant eigenvalues obtained by the spectral decomposition and the peak rate that the CMPP structure is able to capture. The relation evidences the difficulties of CMPP to model actual traffic, characterized by long tailed distribution, as well as traffic data with the well accepted hypothesis of gaussian marginal.

1. Introduction

The CMPP approach for modeling arrivals process by means of a circulant modulated Poisson process, provides a technique for integration of traffic measurement and modeling [10], maintaining, at the same time, the Markovian hypothesis that permits analytical transient and steady-state studies of queueing systems using efficient algorithms [9]. The developed modeling theory has permitted to study the impact of power spectrum, bispectrum, trispectrum, and marginal distribution of the input process on queueing behavior and loss rate. These studies have highlighted the key role played on the queueing performance by the marginal distribution, especially in the low frequencies region [8]. The technique for the construction of a CMPP that matches marginal distribution and autocorrelation function of the observed process has been presented in [2,9], where the authors showed simulation results with measured traffic data to prove the goodness of this approach. In this paper, further analysis of CMPP fitting procedure will be presented, highlighting a limitation of the mentioned algorithm in matching accuracy for the marginal distribution of observed rate process. Moreover, the presented study determines the maximum peak rate captured by the CMPP model once the spectrum has been matched and emphasizes the necessity of a

CMPP structure containing a large number of effective eigenvalues to adequately capture even the light tail of a gaussian function, usually accepted as realistic for traffic distribution in the core network [1]. The relevance of these considerations is related to the impact of the marginal distribution tail of input traffic on queueing behaviour. Indeed, as shown in the numerical analysis Section, optimistic performance are estimated when the peak rate is not matched. On the other hand, the actual traffic rate has a marginal distribution that in some cases exhibits a tail heavier than Gaussian [3,5]; under such condition, the CMPP models result inadequate to estimate realistic queueing performance. Lastly, some advice to overcome the exposed limitation are briefly introduced.

2. Background on CMPP

The fitting procedure of a CMPP model mainly consists of three steps [9], which are briefly summarized in this section. In the first step, the autocorrelation function of the observed rate process is estimated and then matched by a sum of exponentials (with complex parameters λ_k) weighted by real and strictly positive power coefficients ψ_k . This matching is a non-linear problem and cannot be solved directly. An approximate, but quite accurate, solution is obtained by using the Prony algorithm [6] to express the autocorrelation function in terms of complex exponentials with complex coefficients, and then satisfying the constraints on the ψ_k 's (which must be real and strictly positive) by matching the power spectral density (PSD) using the nonnegative least square (NNLS) method.

The second step aims to design the transition frequencies matrix $\underline{\underline{Q}}$ of the underlying modulant continuous time Markov chain. In order to fit the PSD of the modeled process, the eigenvalues of $\underline{\underline{Q}}$ must contain all the λ_k 's obtained in the previous step; the use of a circulant matrix permits to solve the inverse eigenvalues problem. An efficient procedure to solve this problem is the Index Search Algorithm (ISA), presented in [1].

The last step is then the estimation of a vector γ associated to the Poissonian generation of arrivals in each state of the modulating Markov chain, such that the model matches the cumulative distribution function (CDF), $F(x)$, of the observed rate process. In more details, the fitting procedure starts considering that the autocorrelation function of a CMPP model with N states is expressed by the following:

$$R_{CMPP}(\tau) = \psi_0 + \sum_{l=1}^{N-1} \psi_l \cdot \exp(\lambda_l \cdot |\tau|) \quad (2.1)$$

with positive real ψ_l 's. The Fourier transform of (2.1) can be expressed by:

$$S_{CMPP}(\omega) = 2 \cdot \pi \cdot \psi_0 \cdot \delta(\omega) + \sum_{l=1}^{N-1} \psi_l \cdot b_l(\omega) \quad (2.2)$$

where $b_l(\omega) = F[\exp(\lambda_l \cdot |\tau|)] = \frac{-2 \cdot \lambda_l}{\omega^2 + \lambda_l^2}$, and $\int_{-\infty}^{\infty} b_l(\omega) d\omega = 1$; hence, the ψ_l 's

represent the power associated to each λ_l . The λ_l 's are the eigenvalues of the transition matrix, which must include all the "effective" ones that derive by the exponential decomposition of $R(\tau)$, the autocorrelation function of the measured rate process. Using the Prony method, the estimated $R(\tau)$ can be written as

$$R(\tau) \cong \sum_{k=0}^p \psi_{P,k} \cdot \exp(\lambda_{P,k} \cdot |\tau|) \tag{2.3}$$

The presence of a constant term in $R_{CMPP}(\tau)$ requires $\lambda_{P,0}$ to be imposed equal to zero, and consequently $\psi_0 = \psi_{P,0}$: this is simply obtained applying the Prony method to the autocovariance function $C(\tau) = R(\tau) - \bar{\gamma}^2$ ($\bar{\gamma}$ is the mean value of the observed traffic rate), since $R(\tau \rightarrow \infty) \rightarrow \bar{\gamma}^2$, and from (2.3) $\psi_{P,0} = \bar{\gamma}^2$. After the NNLS matching, the expression (2.3) remains substantially unchanged and can be rewritten as

$$R(\tau) \cong \psi_{P,0} + \sum_{k=1}^p \psi_{P,k} \cdot \exp(\lambda_{P,k} \cdot |\tau|) \tag{2.4}$$

being aware that p , the $\psi_{P,k}$'s and the $\lambda_{P,k}$'s may not be the same as those of (2.3) (they surely will not be in the case of complex eigenvalues). The order p of the exponential decomposition may be much less than the order N of the model, and thus in the construction of the transition matrix only few λ_l 's will be imposed equal to the $\lambda_{P,k}$'s. Indicating with \underline{i} the vector of indices (of dimension p) such that $\lambda_{i[k]} = \lambda_{P,k}$, the relation $\psi_{i[k]} = \psi_{P,k}$ consequently holds. On the other hand, in order to obtain $R_{CMPP}(\tau) \cong R(\tau)$, all the other ψ_l 's will be imposed equal to zero.

After having determined the transition matrix \underline{Q} (note that many solutions are possible for each set of eigenvalues, since the order N of the matrix is higher than the number p of desired eigenvalues), the third step, i.e. the design of the rate vector $\underline{\gamma}$ such that $F_{CMPP}(x) \cong F(x)$, involves the minimisation of the distance between $F_{CMPP}(x)$ and $F(x)$, which is obtained by using the Nelder-Mead Simplex Search method. Since $F_{CMPP}(x)$ is a piecewise step function, which jumps by $1/N$ at each value γ_i in $\underline{\gamma}$, the task is to determine the optimal vector $\underline{\gamma}$ which minimises the quantity

$$\sum_{i=0}^{N-1} |\gamma_i - \gamma_i| \tag{2.5}$$

where $\underline{\gamma}$ is obtained by the quantization of $F(x)$ in levels, whose amplitude is $1/N$.

Defining $\beta_i = \sqrt{\psi_i} \cdot \exp(j\vartheta_i)$, $i=0,1,\dots,N-1$, the vector $\underline{\beta} = [\beta_0, \beta_1, \dots, \beta_{N-1}]$ represents the Discrete Fourier Transform of $\underline{\gamma}$, and its Inverse can be expanded as

$$\gamma_i = \bar{\gamma} + \sum_{l=1}^{N-1} \sqrt{\psi_l} \cdot \exp\left[-j\left(\frac{2 \cdot \pi \cdot i \cdot l}{N} - \vartheta_l\right)\right], \text{ for } i=0,1,2,\dots,N-1$$

where the expression of β_l has been substituted.

In order to obtain real γ_i , $\underline{\beta}$ must exhibit the Hermitian property (i.e. $\beta_{N-l} = \beta_l^*$, which corresponds to $\psi_{N-l} = \psi_l$ and $\vartheta_{N-l} = -\vartheta_l$). Indeed, if $\underline{\beta}$ does not satisfy the Hermitian property, its Inverse Finite Fourier Transform $\underline{\gamma}$ cannot be real. Under the condition of Hermitianity on $\underline{\beta}$, the above relation assumes the following expression

$$\gamma_i = \bar{\gamma} + \sum_{l=1}^{N-1} \sqrt{\psi_l} \cdot \cos\left(\frac{2 \cdot \pi \cdot i \cdot l}{N} - \vartheta_l\right), \text{ for } i=0,1,2,\dots,N-1 \tag{2.6}$$

that permits to estimate $\underline{\gamma}$ by applying the Nelder-Meade Simplex Search method to (2.5) as a function of $\underline{\vartheta}$. The Hermitian conditions on β_l are automatically satisfied for those power coefficients related to conjugated complex pairs of eigenvalues, but cannot stand for real ones, since only one ψ_i is associated to each of them. To overcome this problem, each real eigenvalue needs to be considered twice. In order to maintain the same correlation structure (or equivalently the same PSD), the corresponding power coefficients will be assumed equal to half of the original ψ_i 's.

3. Analysis of Fitting Procedure

The investigation presented in this work involves the last step of the fitting procedure and evidences a relevant limitation on the tail behavior of the marginal distribution of CMPP models. This limitation may considerably affect the evaluation of queueing performance of actual traffic, leading to an underestimation of network resources needed to guarantee the target QoS expressed in terms of loss probability.

The first observation on the fitting procedure is that, putting $\tau=0$ in (2.4), the variance σ^2 of the rate process can be expressed as the sum of the ψ_l for $l=1, 2, \dots, N-1$. On the other hand, the maximum theoretical rate achievable by the model is derived by (2.6) putting all the cosines equal to +1. In this case, a second relation involving ψ_l 's can be simply obtained:

$$\gamma_{MAX} - \bar{\gamma} = \sum_{l=1}^{N-1} \sqrt{\psi_l} \tag{3.1}$$

The maximum rate deviation from the mean value is then limited by (3.1), under the constraint $\sum_{l=1}^{N-1} \psi_l = \sigma^2$. As we stated before, only the p ψ_l 's associated to the effective eigenvalues are non zero. Among these p power coefficients, some are

related to real λ_i , hence each of them needs to be split into two terms with halved magnitude. Therefore, the resulting set of couples (λ_i, ψ_i) after this operation consists of q elements, with $p \leq q \ll N$. In the remaining of the paper, we will refer to (λ_i, ψ_i) as elements of this set; consequently if $\lambda_i \in \mathfrak{R} \Rightarrow \exists \lambda_{N-i} = \lambda_i$ and $\psi_{N-i} = \psi_i$, with ψ_i equal to half of the original power coefficient obtained by NNLS algorithm. Thus relation (3.1) can be rewritten as

$$\gamma_{MAX} - \bar{\gamma} = \sum_{k=1}^q \sqrt{\psi_{i[k]}} \tag{3.2}$$

where \underline{i} is now a vector of indices of dimension q .

In most actual cases, the measured peak rate is quite higher than the mean value and then, in order to capture the long tailed behavior of the rate distribution, the sum in equation (3.2) should be as large as possible. The problem of maximizing (3.2) with the constraint $\sum_{k=1}^q \psi_{i[k]} = \sigma^2$ can be easily solved using the Lagrange-Multipliers method, leading to the solution

$$\psi_{i[k]} = \sigma^2 / q . \tag{3.3}$$

Consequently

$$\max_{\underline{\psi}} \{ \gamma_{MAX} - \bar{\gamma} \} = \sum_{k=1}^q \sqrt{\psi_{i[k]}} \cong \sqrt{q} \cdot \sigma \tag{3.4}$$

holds.

This equation represents the intrinsic limitation of CMPP in terms of maximum achievable rate as a function of the number q of effective eigenvalues, in the hypothesis of evenly distributed power coefficients. Considering that the distribution of the amplitudes of the dominant ψ_i 's will hardly be like (3.3), the γ_{MAX} value obtained from the above relation represents only an upper bound, actually difficult to reach. However, (3.4) gives an indication on the minimum number of exponentials required to reach a target peak rate, fixed the variance and the mean value of the observed rate process. As an example, suppose that the observed process presents a gaussian marginal distribution with mean \bar{x} and variance σ^2 ; hence, the CDF is

$$F(x) = 1 - Q\left(\frac{x - \bar{x}}{\sigma}\right) \text{ where } Q(y) = \frac{1}{2\pi} \int_y^\infty \exp\left(-\frac{x^2}{2}\right) dx \tag{3.5}$$

Using a 500 state CMPP to model this rate process, the CDF results divided in intervals whose heights are $1/500 = 2 \cdot 10^{-3}$ (the high number of states has been chosen in order to obtain a fine quantization of CDF). The maximum level of the quantized CDF is then limited to the value 0.998, corresponding to $(x - \bar{x})/\sigma$ equal to 3.09. The

comparison of this relation with (3.4) leads to $q=3.09^2 \cong 9.55$. Therefore, the original autocovariance should be decomposed in, at least, 10 exponentials; more likely they will not be sufficient, since the assumption that all power coefficients ψ_i 's are of equal magnitude is not easily verified. Indeed, a more realistic scenario is that few ψ_i 's (around 6 or 8, as supported by the analysis in the Numerical Section) will be dominant with respect to the others and consequently the reproduced peak rate will be such that $\gamma_{MAX} - \bar{\gamma} = \sqrt{6} \cdot \sigma$, in correspondence of whom the original CDF will assume the value $1 - Q(\sqrt{6}) \cong 0.993$. Hence, the tail of the model will be shorter than the one of the observed rate process. The relevance of this drawback can be pointed out envisaging that, especially for traffic whose power spectrum is concentrated in the lower region of the frequencies [7] (this assumption is supported by the self similar nature highlighted by the recent modeling results based on the analysis of acquired traffic data [11]), the tail of the marginal distribution has a deep impact on the network resources required to guarantee a target loss probability.

Note that from (2.6) $\bar{\gamma} - \gamma_{MIN} = \sum_{k=1}^q \sqrt{\psi_{i[k]}}$ can also be derived, which, together with (3.2), implies the following general relation

$$|\gamma - \bar{\gamma}|_{MAX} = \sum_{k=1}^q \sqrt{\psi_{i[k]}} \tag{3.6}$$

4. Numerical Results

To test the relevance of the presented analysis, we consider two sets of simulations: the first one is carried out applying the CMPP fitting procedure on synthetic data with a well defined and known spectral decomposition, whereas the second one refers to actual traffic data. In the first simulation scenario, we have generated two traces having the same mean value, variance and power spectrum, differing only by their probability density functions (one is gaussian and the other one is triangular). The values of the mean and the variance of the data ($\bar{\gamma}=4500$ cells per second, $\sigma^2=10^6$ cps²) have been chosen in order to obtain a negligible probability of having negative values in the gaussian rate trace. In order to build the CMPP models of the two traces, we have first decomposed the autocovariance functions into a sum of complex exponentials. Then, applying the NNLS algorithm, only the three couples of eigenvalues shown in Table 1.(a) (here and in the remaining of the paper, the “**”) reminds that the real eigenvalue is considered twice and that the corresponding ψ has been already halved in magnitude, according to the procedure described in the previous sections) have turned out to be associated to power coefficients ψ different from zero (note that $\sum_i \psi_i$ equals the imposed variance σ^2).

Table 1. Eigenvalues with non-zero power coefficient deriving from the NNLS algorithm: (a) Synthetic Gaussian trace; (b) trace “Videoconference”; (c) trace “October89”

ψ [cps ²]	$\hat{\lambda}$ [rad/sec]	ψ [cps ²]	$\hat{\lambda}$ [rad/sec]	ψ [cps ²]	$\hat{\lambda}$ [rad/sec]
$1.72 e+5$	-0.78 ± 6.92	$9.72 e+4$	-2.85 ± 6.57	$1.46 e+5$	-1.23 ± 6.03
$2.62 e+5$	-1.26 ± 3.00	$1.25 e+5$	$-5.92 **$	$2.74 e+6$	-2.35 ± 0.554
$6.6e+4$	$-0.14 **$	$6.40 e+5$	$-0.628 **$	$5.25 e+5$	$-0.033 **$
(a)		(b)		(c)	

According to (3.6), the maximum deviation achievable using this set of eigenvalues is equal to

$$|\gamma - \bar{\gamma}|_{MAX} = \sum \sqrt{\psi} \cong 2367 \text{ cps} . \tag{4.1}$$

The obtained value is quite close to the limit $\sqrt{6} \cdot \sigma \cong 2450$ cps, since the power coefficient magnitudes have a quasi-uniform distribution.

Two CMPP models have been built from these eigenvalues to match the two different CDFs. The γ_{MAX} ’s for the two models have resulted equal to 6660 cps and 6500 cps for the triangular and the gaussian hypothesis, respectively. These limits do not consider the exponential and I.I.D. generation of arrivals (Poissonian microdynamics) in each state of the CMPP model, which affects the upper tails of marginal distributions, as well as the peak rates of generated traces. In particular, the latter ones are higher than the respective γ_{MAX} ’s, (see the third column of Table 2, which summarizes the main statistics of the analyzed data traffic and of the related traces generated by the corresponding CMPP models). The results of the fittings in terms of PSD and complementary probability (CP) of the generated traces in the gaussian hypothesis are shown in Fig. 1. The excellent matching of the PSD is not accompanied by an equivalently good fitting of the distribution function. This mismatch is not easily revealed by the CDF’s comparison, hence a CP plot is always required to appreciate possible differences in the upper tails. In the triangular case, we observed good fitting results for both PSD and CP plots, but we do not report the relative figures here for sake of simplicity.

In order to estimate the errors in the evaluation of queueing performance introduced by the mismatching of the CDF tail, we have analyzed the results obtained by means of discrete-event simulations. The simulations have been carried out feeding a FCFS G/D/1/K queueing system with the four traces; in the remaining of the paper we will indicate with μ the servant constant cell rate. The size K of the buffer has been fixed equal to 250 cells, corresponding to a maximum delay introduced by the queue, variable with the normalized offered load, around 50msec. The traffic is completely contained in the LF and the MF region of the queue [7]. Therefore, no further filtering of traffic is possible without losing a portion of its spectrum, i.e. MF components, which would affect the queueing performance evaluation. The results are shown in the last two columns of Table 4 and emphasize as the reduced peak of the trace related to the gaussian case leads to a slightly optimistic resources allocation (nearly 6%). It is

important to highlight that the CMPP designs have been repeated using different values for N , ranging from 200 to 500, but no appreciable difference has been noticed.

In the second set of simulations, two real traffic traces have been considered: the two data sets are related to a videoconference service and a LAN traffic trace. The description of the characteristics of the videoconference traffic data are described in [4], whereas the second trace is the well known ‘‘October89’’ trace collected at Bellcore Labs. In both analyses, the data (i.e. the estimates of the rate processes) refer to the equivalent number of ATM cells (calculated as the number of transmitted bytes divided by 48) per second observed in non overlapped time intervals of length T_u . In the first case, we assume T_u equal to the frame period, i.e. 40 ms, whereas for the second traffic trace the value of 200 ms has been chosen.

The particular shape of the autocovariance function of the videoconference trace has led to a spectral decomposition characterized by the eigenvalues reported in Table 1.(b). In this real scenario, we have obtained only few eigenvalues with non-zero power coefficients, enforcing the hypothesis introduced in the previous section regarding the number of dominant ψ 's. In this case the maximum achievable deviation is equal to $|\gamma - \bar{\gamma}|_{MAX} = \sum \sqrt{\psi} \cong 5950$ cps and the mean value is approximately 4350 cps. Hence, we should expect a CMPP peak rate of about 10300 cps, against a measured peak rate of about 21000 cps. The peak rate error is very high (nearly 43%) due to the very unfavorable condition on the power coefficient distribution. Indeed, the power spectral decomposition of the considered trace presents only a single couple of dominant effective real eigenvalues and, at the same time, its marginal distribution exhibits an upper tail behavior heavier than gaussian [4]. The modeling results are shown in Fig. 2.(a) and 2.(b), which represent the PSD and the complementary probability matching, respectively. In particular, Fig. 2.(b) shows clearly that the obtained CMPP peak rate is much lower than the observed 21000 cps, as expected. This large difference in terms of peak rate leads to a very optimistic evaluation of the queueing behavior, as pointed out by the simulations results, see Table 2, columns 4 and 5, which contain the servant cell rate required in order to reach a cell loss probability of 10^{-4} and 10^{-5} .

The last analysis refers to the first 1000 s of the above mentioned LAN traffic trace; we do not consider the entire data set since a shift of the mean value has been noticed out of this time period. The relevant eigenvalues obtained after the NNLS matching of the spectrum are reported in Table 1.(c), and correspond to a theoretical maximum deviation $|\gamma - \bar{\gamma}|_{MAX} \cong 5524$ cps. Adding the estimated mean (about 6016 cps) to this value, the limit 11540 cps is obtained, with respect to a model $\gamma_{MAX} = \max_i(\gamma_i)$ of 11145 cps and of a data set peak of 17800 cps. Fig. 3.(a) and 3.(b) present the matching of PSD and complementary probability respectively: the former shows the good fitting of the considered second order statistic, whereas the latter confirms the limitation of the model in capturing the upper tail behavior of marginal distribution.

The last two columns of Table 2 clearly evidence the entity of the relative error in resources allocation when the CMPP models does not capture the upper tail of marginal distribution (and particularly the peak cell rate).

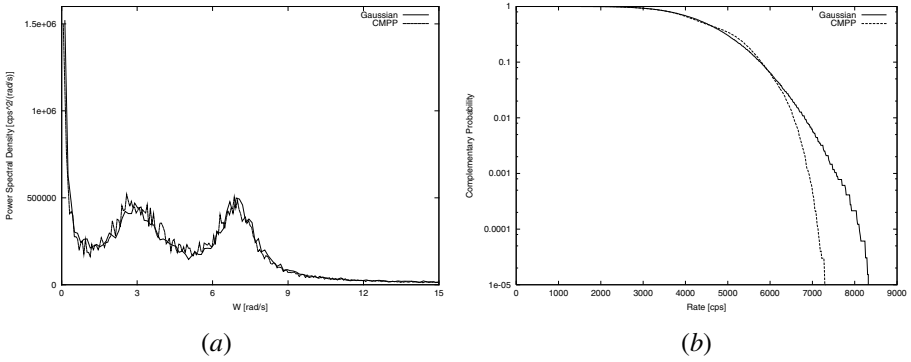


Fig. 1. Comparison of (a) Power Spectral Density and (b) Complementary Probability of Synthetic Traffic Data – Gaussian Case

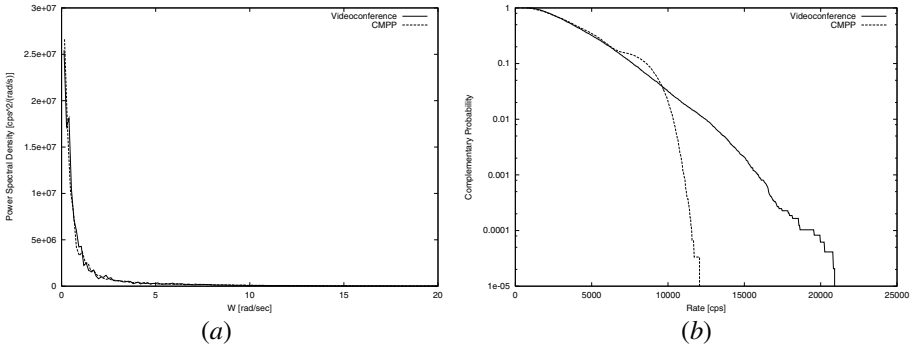


Fig. 2. Comparison of (a) Power Spectral Density and (b) Complementary Probability of “Videoconference” Trace

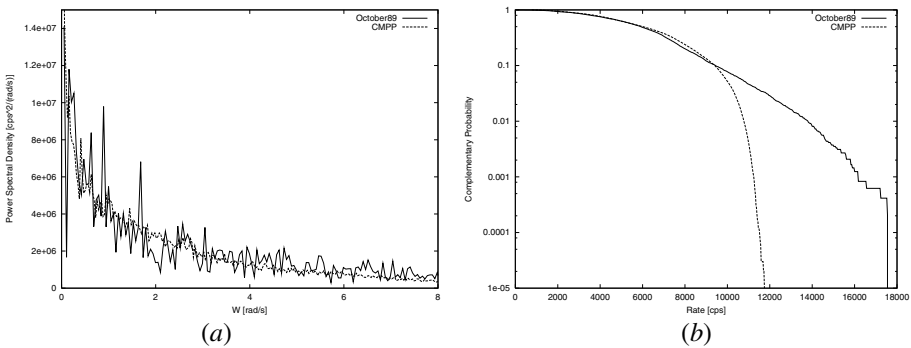


Fig. 3. Comparison of (a) Power Spectral Density and (b) Complementary Probability of “October89” Trace

Table 2. Relevant statistics of analysed traces and corresponding CMPP models

	Mean [cps]	Variance [cps ²]	Peak [cps]	$\mu (P_{\text{loss}}=10^{-4})$ [cps]	$\mu (P_{\text{loss}}=10^{-5})$ [cps]
Triangular (CMPP model)	4493 (4506)	1.00e+6 (1.00e+6)	7000 (7500)	5930 6025 (+1.6%)	6033 6170 (+2.3%)
Gaussian (CMPP model)	4490 (4503)	1.00e+6 (0.99e+6)	8344 (7330)	6110 5855 (-4.2%)	6390 6010 (-5.9%)
October89 (CMPP model)	6016 (6030)	7.48e+6 (6.03e+6)	17600 (11780)	15360 10330 (-32.7%)	16120 10590 (-34.3%)
Videoconference (CMPP model)	4350 (4410)	6.15e+6 (6.10e+6)	20975 (12100)	16400 9670 (-41.0%)	18520 9820 (-47.0%)

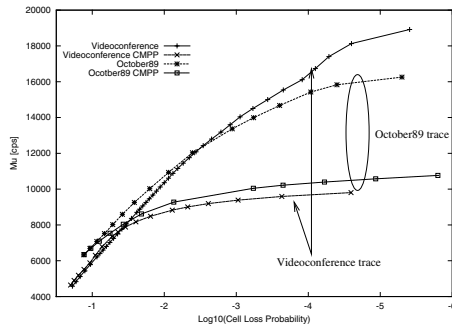


Fig. 4. Performance Comparison

In the cases of the videoconference and LAN traces, a more complete analysis of modeling performance is shown in Figure 4, where the target CLP is plotted versus the servant cell rate needed to guarantee it. The analysis of Table 2 points out two important results: the first one concerns the critical behavior of the CMPP model even in the gaussian case, which implies an optimistic resource allocation (4 to 6%). The second result highlights as actual traffic exhibits a slower decay of marginal distribution with respect to the gaussian hypothesis leading to a less suitable environment for the CMPP approach, evidenced by the large errors suffered (in terms of peak rate, matching of the upper tail of marginal distribution and consequently of the network resources needed to guarantee a target cell loss probability).

5. Improvement Proposals to Overcome the CMPP Drawback

We have observed that, under the same conditions on q , the uniform distribution of ψ 's magnitudes is the only one that guarantees the maximum model peak γ_{MAX} . Thus, it is desirable to have a spectral decomposition with power coefficients exhibiting this feature. To the aim of coming close to the uniform distribution and increasing the parameter q , a straight solution can be to split each dominant exponential in the sum of two or more terms. The j -th term of the exponential decomposition is represented by

the parameters λ_j and ψ_j , as described in Section 2. Our suggestion is to obtain an equivalent contribution to the autocorrelation function using a number B of power coefficients (associated to the given exponential). This procedure results in having B terms $\lambda_{j,i}, \psi_{j,i}$ for each couple λ_j, ψ_j , where $\lambda_{j,i} = \lambda_j$ and $\psi_{j,i} = \psi_j / B$, $i=1,2,\dots,B$.

Consequently, the autocorrelation function remains unchanged, whereas the contribution of the decomposed couple λ_j, ψ_j to the peak rate becomes

$$\sum_{i=1}^B \sqrt{\psi_{j,i}} = \sum_{i=1}^B \sqrt{\frac{\psi_j}{B}} = \sqrt{B} \cdot \sqrt{\psi_j}, \text{ i.e. } \sqrt{B} \text{ times the original one (i.e. } \sqrt{\psi_j} \text{)}.$$

Unfortunately, this solution also presents a drawback, whose relevance needs further investigation: the increased number of eigenvalues to be assigned to \underline{Q} can cause the increment of the minimum number of the model states N that permits the solution of the ISA problem. As a consequence, an higher N means that the model parameters derivation takes a longer time, thus limiting the use of CMPP approach in a real time performance estimator.

Another possible approach to overcome the presented limitation is to increase the value of T . In such a way, the peaks of the traffic data will be reduced, hence making easier to capture them by the CMPP model. Unfortunately, also this procedure presents some drawbacks. First of all, a limit exists on the maximum value of the time quantum T . Indeed, using higher T can cause the loss of a significant portion of information associated to the traffic data [7], leading to inaccurate queueing performance evaluation. Furthermore, increasing T reduces the variance σ^2 of the resulting trace. Consequently, the constraint on the ψ_i 's sum ($\sum_i \psi_i = \sigma^2$) produces a set of power coefficients of reduced magnitude, vanishing the advantage gained by the trace peak reduction. Further study is then needed to highlight the trend of the peak rate decay with respect to the time quantum T in different simulation scenarios, and to derive a relation with the tail behavior.

6. Conclusions

The paper presents an analysis of the algorithm for the measurement-based parameters estimation of CMPP models, raising some warnings to be aware of in the use of this modeling approach. In particular, the main result of the paper is the analytical derivation of an intrinsic limitation of the fitting procedure in the modeling of traffic characterized by long-tailed marginal distribution. Furthermore, the analysis shows that even in the gaussian hypothesis, a CMPP structure containing a large number of effective eigenvalues is necessary to adequately fit the CDF.

In general, the limitation manifests when the difference between the peak and the mean rate of the traffic data set exceeds few times its standard deviation. In this condition the CMPP model cannot match with sufficient accuracy the behavior of the CDF upper tail, leading to optimistic prediction of the network resources needed to guarantee the target QoS (in terms of cell loss probability).

Discrete-event simulations driven by actual traffic data and synthetic traces, generated according to the corresponding CMPP models, have confirmed the results of the proposed analytical study. Moreover, the analysis of the simulation results shows the practical relevance of the CMPP limitation in a single server queueing system, a relevant case study in performance comparison. In particular, the analyzed cases highlight the optimistic resources allocation produced by the fitting errors on the CDF. The presented limitation reduces the field of applicability of the CMPP modeling approach, hence further studies are needed to overcome this drawback. To this aim, some advice are presented as possible solutions to be investigated.

References

1. R. G. Addie, M. Zuckerman, T.D. Neame, "Broadband Traffic Modeling: Simple Solutions to Hard Problems" IEEE Communications Magazine, August 1998, pp.88-95
2. H. Che, San-qi Li, "Fast Algorithms for Measurement-Based Traffic Modeling", IEEE Journal on Selected Areas in Communications, June 1998, pp. 612-625
3. R. G. Garroppo, S. Giordano, S. Porcarelli, G. Procissi, "Testing α -stable processes in modelling broadband teletraffic" Proc. of IEEE ICC 2000, New Orleans, Louisiana, USA, 18-22 June, 2000
4. R. G. Garroppo, S. Giordano, M. Pagano, "Stochastic Features of VBR Video Traffic and Queueing Working Conditions: a Simulation Study using Chaotic Map Generator" in Proc. of IFIP Broadband Communications '99, Hong Kong, November 1999
5. A. Karasaridis, D. Hatzinakos, "A Non-Gaussian Self-Similarity Processes for Broadband Heavy-Traffic Modeling", in Proc. of GLOBECOM 98, pp. 2995-3000, Sidney, 1998
6. Steven M. Kay, "Modern Spectral Estimation: Theory & Application", Prentice-Hall, 1988.
7. Y. Kim, San-qi Li, "Timescales of Interest in Traffic Measurement for Link Bandwidth Allocation Design", Proc. IEEE, Infocom '96, March 1996, pp. 738-748
8. San-qi Li, Chia-Lin Hwang, "Queue Response to Input Correlation Functions: Continuous Spectral Analysis", ACM/IEEE Transactions on Networking, December 1993, pp. 678-691
9. San-qi Li, Chia-Lin Hwang, "On the Convergence of Traffic Measurement and Queueing Analysis: A Statistical-Matching and Queueing (SMAQ) Tool", IEEE/ACM Transactions on Networking, February 1997, pp. 95-110
10. San-qi Li, S. Park, D. Arifler, "SMAQ: A Measurement-Based Tool for Traffic Modeling and Queueing Analysis. Part I: Design Methodologies and Software Architecture", IEEE Communications Magazine, August 1998, pp. 56-65
11. W. Willinger, M.S. Taqqu, A. Erramilli, "A bibliographical guide to self-similar traffic and performance modelling for modern high speed networks", in F.P. Kelly, S. Zachary and I. Ziedins eds., Stochastic networks: Theory and Applications in Telecommunication Networks, Vol. 4 of Royal Statistical Society Lecture Notes Series, pp. 91-104, Oxford University Press, Oxford, 1996