Analysing the versatility of the 2-MMPP traffic model

A. Nogueira, R. Valadas
Universidade de Aveiro / Instituto de Telecomunicações, 3810-193 Aveiro, Portugal
e-mail: nogueira@av.it.pt; rv@ua.pt

Summary

Cell loss is an important measure of Quality of Service (QoS) in ATM networks. Cell loss performance of ATM elements handling bursty sources must be evaluated in order to provide guaranteed QoS to users and to dimension network resources correctly.

In this paper, we make a comparative study of two fitting algorithms for 2-MMPP ATM traffic models, both based on cell interarrival times: one fits the cumulative distribution and auto-covariance functions and the other fits the first three moments and auto-covariance function. The 2-MMPP is then evaluated as an approximation to 3-MMPP, 5-MMPP, IPP, IDP, Hyper-Exponential On-Off and Self-Similar traffic sources. The usefulness of both fitting procedures for cell loss prediction is evaluated and compared. The proposed algorithms can be used in the characterisation of ATM traffic streams and in connection admission control procedures.

1. Introduction

Efficient operation of ATM networks and provision of guaranteed Quality of Service (QoS) for many services to be supported in such networks will require proper modelling and control of ATM traffic. Cell loss performance of ATM elements, which is one of the fundamental measures of QoS, must be evaluated in order to provide guaranteed performance levels to users and to dimension network elements correctly.

On-Off models have been commonly used for representing bursty sources. Particularly, the IPP model has been found to adequately represent the behaviour of voice sources and it has been proposed as a model for other traffic types, including video, data and multimedia traffic. The MMPP model, with a variable number of states, is a non-renewal model that has also been widely used for modelling ATM traffic. However, findings in a number of packet-based network scenarios suggest that traffic in such networks presents a certain degree of self-similarity.

The proliferation of different traffic models and their application in various (and sometimes quite analogous) scenarios have motivated us to study the conditions of applicability for different traffic models (and their associated estimation procedures) to certain traffic scenarios. Specifically, here we make an exhaustive study of the 2-MMPP process and two of its associated fitting procedures, analysing their suitability in approximating various types of arriving flows. In particular, we have considered 3-MMPP, 5-MMPP, IPP, IDP, On-Off sources with hyper-exponential On and Off period distributions, which are more general than exponential On-Off sources but are still quite tractable, and Self-similar traffic models.

The rest of the paper is structured as follows: in section 2, we briefly summarise various commonly used point process superposition approximations based on the 2-MMPP process; in section 3, an outline of the methodology followed for this study is provided and in section 4 the main results are presented. Finally, in section 5, we will draw some conclusions and foresee the most relevant future tasks.

2. 2-MMPP fitting procedures

In order to characterise the input stream at an ATM multiplexer, two main approaches were suggested in the literature: point process and fluid flow approximations. We will focus on non-renewal point process approximations based on the MMPP process, since such approximations have been shown to capture both the cell scale and burst scale characteristics of traffic [2]. A number of approaches have been proposed in the literature in which the superimposed stream is modelled by a 2-MMPP. These approaches differ in the choice of the traffic characteristics used to match the 2-MMPP parameters and can be divided in two groups: methods based on cell counting statistics ([2], [3] and [4]) and methods based on inter-arrival time statistics ([1], [5] and [6]).

Assuming interval-stationary 2-MMPP processes, where X_i represents the interarrival time between the i^{th} and $(i+1)^{th}$ cells, the distribution of the interarrival time X_i is a second order *hyperexponential distribution* (H_2) with complementary CDF:

$$F_c(x) = qe^{-u_1 x} + (1-q)e^{-u_2 x},\ 0 < q < 1 \qquad (1)$$

and density function given by

$$f(x) = qu_1 e^{-u_1 x} + (1-q)u_2 e^{-u_2 x},\ 0 < q < 1. \qquad (2)$$

The three parameters of the hyperexponential distribution, u_1, u_2 and q, can be related with the 2-MMPP parameters (the state transition rates, (r_1, r_2)), and the Poisson arrival rates in each state (λ_1, λ_2)) by:

$$u_1 = \frac{\lambda_1 + \lambda_2 + r_1 + r_2 - \delta}{2}; \qquad (3a)$$

$$u_2 = \frac{\lambda_1 + \lambda_2 + r_1 + r_2 + \delta}{2}; \qquad (3b)$$

$$q = \frac{\lambda_2 r_1^2 + \lambda_1^2 r_2}{(\lambda_2 r_1 + \lambda_1 r_2)(u_1 - u_2)} - \frac{u_2}{u_1 - u_2} \qquad (3c)$$

where $\delta = \sqrt{(\lambda_1 - \lambda_2 + r_1 - r_2)^2 + 4 r_1 r_2}$. The auto-covariance function, C[k], $k \geq 1$, is given by ([6]):

$$C[k] = E[(X_1 - E\{X_1\})(X_{k+1} - E\{X_{k+1}\})] = A\sigma^k \qquad (4)$$

where $A = \dfrac{(\lambda_1 - \lambda_2)^2 r_1 r_2}{(\lambda_2 r_1 + \lambda_1 r_2)^2 (\lambda_1 \lambda_2 + \lambda_2 r_1 + \lambda_1 r_2)}$ (5)

and $\sigma = \dfrac{\lambda_1 \lambda_2}{(\lambda_1 \lambda_2 + \lambda_2 r_1 + \lambda_1 r_2)}$ (6)

We can see from equation 6 that when one of the cell arrival rates is zero, the correlation is null. From equation 5, we see that the higher the difference between the state arrival rates the higher is A and consequently the correlation value.

The algorithm presented in [1] (called the *moments* method), estimates the second-order hyper-exponential parameters u_1, u_2, q by fitting the empirical and theoretical first three moments of the interarrival time process and estimates the parameter σ by fitting the empirical and theoretical auto-covariance functions.

In the fitting procedure proposed in [6], the following characterising statistics of the arrival stream are matched to the 2-MMPP parameters: (i) the complementary distribution function of the inter-arrival times; and (ii) the covariance function of the inter-arrival times. The complementary distribution function of the inter-arrival times, $F_c(x)$, is characterised based on three quantities measured from experimental data: the mean interarrival time, the initial slope of $F_c(x)$ and the asymptotic slope of $F_c(x)$.

In [1], a variant of this method (designated there by *cdf method*) was implemented, fitting the empirical and theoretical complementary CDFs and the empirical and theoretical auto-covariance functions using the *NonlinearFit* function of the MATHEMATICA package. Here, the complementary distribution function of the inter-arrival times, $F_c(x)$, is characterised in a one-step procedure, using the above mentioned non-linear fitting scheme.

Note that in both estimation procedures, only the decaying rate of the auto-correlation function, σ, is fitted, so we are not concerned with its amplitude, A.

Most of the above methods are based on counting rather than inter-arrival statistics, and one of the reasons for this lies in the difficulty of capturing inter-arrival statistics from real traffic data. The measurement equipment must have high resolution and huge buffer spaces in order to capture and store every arriving cell instant.

In order to obtain cell loss in an ATM multiplexer, the matched 2-MMPP needs to be applied as the arrival process. The resulting queueing model is a 2-MMPP/D/1/K queue. There exists exact and approximate analytical methods for evaluating cell loss in such queues using the matrix geometric approach: exact methods [7] become computationally expensive when dealing with large buffer sizes (the number of required operations is $O(k^3)$, where k denotes the buffer size), while approximate methods can reduce the computational complexity by two orders of magnitude by using a truncated version of the matrix generating function [8]. Cell loss can also be evaluated by simulation.

3. Methodology used

An effective traffic model must reproduce the first and second order statistics of the original traffic sample. The distribution function defines the first order statistics whereas the second order statistics can be accounted for by the auto-correlation function. The second order statistics play an important role in traffic modelling, because traffic auto-correlation is an important factor in ATM cell losses due to buffer and bandwidth limitations.

Our analysis of the 2-MMPP model (and its associated inference procedures) as representative of different traffic types includes some of the most well-known and studied traffic models: the MMPP process (with different numbers of states), the IPP (a special case of the 2-MMPP process) and IDP processes, the second-order Hyper-exponential (H_2) On-Off process (Figure 1) and Self-similar traffic generated according to methods described in [9]. The H_2 On-Off process was included in this comparative study because the three parameters of the hyper-exponential distribution give more freedom in matching the first three moments of the On or Off period durations for the On-Off sources than it is possible using exponential distributions.

The suitability of each inference procedure to describe various traffic flows was tested by comparing the CLR obtained when driving a buffer with finite capacity and constant service rate with (i) traffic streams of three hundred thousand cells each generated (through simulation) by the original traffic model and (ii) data generated by the 2-MMPP model inferred (using *moments* and *cdf* fitting procedures) from data of step (i). Here, 10 replicas were generated for each inferred model. We also compared some statistics of the original and modelled data, namely: mean, standard deviation and Hurst parameter.

We tested the 2-MMPP model and its inference procedures as an approximation to 3-MMPP, 5-MMPP, IPP, IDP, H_2 On-Off and Self-similar (with H=0.8) traffic models. The Hurst parameter was considered in order to assess the self-similar behaviour of the generated traffic streams. This parameter was measured using the methods described in [10].

4. Results and discussion

A comparison of the main statistics of the original and modelled traffic is presented in Table 1. For the modelled traffic, the table shows the mean and 95% confidence intervals corresponding to 10 replicas.

In order to provide a reference for our implementations of the fitting procedures, the first original model considered was the 2-MMPP process. From Table 1, we see that both estimation procedures can capture all major statistics of the original traffic. Regarding the auto-correlation function, we can see from Figure 2a that the fitting procedures slightly overestimate the amplitude of the original traffic auto-correlation function, but the results are very close to each other and the decaying behaviour is accurately replicated. Comparing CLR values (Figure 2b), the agreement between original and simulated values is still very good, indicating that these 2-MMPP fitting procedures can accurately estimate the parameters of the original 2-MMPP process.

For the 3-MMPP model, we see that both fitting procedures do not accurately estimate the main statistics of the original traffic, which are, however, near the confidence intervals of the modelled traffic. Comparing the auto-correlation functions (Figure 3a), we notice that both fitting methods capture very accurately the original auto-correlation but they slightly overestimate the CLR values caused by the original traffic (Figure 3b). It was verified that both estimated 2-MMPP models have longer bursts than the original 3-MMPP model, resulting in higher cell losses in queue. Comparing directly the behaviour of both fitting procedures, we see that they behave quite similarly for this particular model, with a light advantage for the *moments* method.

In the attempt to approximate a 5-MMPP by a 2-MMPP model, both fitting procedures perform in a satisfactory way (Figures 4a and 4b). The *moments* and the *cdf* inference procedures both overestimate the main statistics of the original traffic and its CLR values. Nonetheless, the deviation from the original values is higher for the *moments* approach, reaching about 50% for CLR values. Regarding the auto-correlation function, both fitting methods capture the mean decaying tendency of the original function, but they tend to underestimate its amplitude for higher lags.

From the above analysis, we can conclude that as the number of states of the original Markov chain increases, the accuracy of the approximation by a 2-MMPP process begins to degrade, for both estimation approaches. However, for a number of states up to 5, the error of this approximation is not too penalising. Of course, these conclusions do not apply to any set of parameters selected for each model. Anyway, if the differences between (some of) the arrival rates in each state (for the 3 and 5-MMPP models) are not too significant, or if the process does not change state too frequently, the error of approximating three or five states by only two states is not too much penalising.

The next process under consideration was the frequently used IPP model. As known, this On-Off process possesses no correlation [1] and so, in a first glance, we can not expect a good approximation of this model by a 2-MMPP. From Table 1, we see that the *moments* and the *cdf* methods clearly overestimate the main statistics of the original traffic. Obviously, the null auto-correlation of the original trace (for our generator, the auto-correlation is not null but is still very small – Figure 5a) is not captured by the approximating 2-MMPP processes, in spite of the very small values presented by the process estimated using the moments method. This fact has direct implications in the CLR behaviour, and we see that traffic generated according to the selected original IPP process generates very small CLR values, inclusively falling to zero for buffer sizes larger than 30 cells (Figure 5b). The 2-MMPP process estimated using the *moments* method generates CLR values one order of magnitude larger than the original process, whereas for the *cdf* approach this difference extends itself to three orders of magnitude.

Regarding the IDP process, our first observation is that the *moments* method could not estimate a 2-MMPP model from IDP traffic. We tested different sets of IDP parameters and for all of them it was impossible to estimate a valid transition rate matrix for the 2-MMPP underlying Markov chain, that is, a matrix that satisfies the general format of the 2-MMPP infinitesimal generator. This is mainly due to the estimated values of the second and third moments. The *cdf* method, however, estimates a 2-MMPP process from the original trace. The fitting is not very accurate: the main statistics of the original traffic are not accurately replicated by the estimated 2-MMPP process, as well as the CLR behaviour (Figure 6b). The auto-correlation function decaying behaviour (Figure 6a), on the other hand, is captured with acceptable accuracy. The ACF function of the original IDP process has a certain periodicity, corresponding to the cell bursts generated in the On periods. In our simulations, however, we have chosen an IDP process with frequent state transitions, causing relatively small burst lengths.

The On and Off periods of the H_2 On-Off traffic model have heavy-tailed distributions. Although the heavy-tailed property is not a necessary condition for self-similarity, some authors suggest that the long range dependency property exhibited by some real traffic flows is directly related to this behaviour of the individual sources. However, it was essentially the fact that these second-order hyper-exponential distributions (which are represented by three parameters) allow more degrees of freedom for matching the first three moments of the distributions of the On and Off states that lead us to include this model in our study.

Based on this knowledge, one can expect an original auto-correlation function with significant values at various time scales, and this behaviour cannot be obviously replicated by any 2-MMPP model. Looking at the auto-correlation plots (Figure 7a) we, effectively, see that the original auto-correlation behaves almost in the same way regardless of the considered lag values (although we do not cover many time scales in these plots): in different time scales, we still find very significant auto-correlation coefficients, which is one of the most striking visual characteristics of heavy-tailed distributions. This disparity in the correlation behaviours is directly reflected in the CLR performance exhibited by the arrival traces (Figure 7b): the original trace causes significant CLR values, for buffer sizes up to 100 cells, while the estimated traces cause much smaller and rapidly decaying CLR values. Comparing directly both estimation approaches, the results obtained using the *moments* method are better.

Considering an original self-similar traffic model (with Hurst parameter $H = 0.8$), the accuracy obtained using both fitting procedures is not very promising: the main statistics of the original traffic are not accurately estimated and the auto-correlation function amplitude (Figure 8a) of the original traffic is clearly underestimated. The CLR values are clearly underestimated (by more than two orders of magnitude) by both inference approaches (Figure 8b). The self-similar generator used has a low variability, so our results tell us that these modelling approaches cannot be applied to self-similar traffic with low rate variation.

5. Conclusions and further work

In this paper, we made a comparative study of two fitting algorithms for 2-MMPP ATM traffic models, both based on cell interarrival times: one that fits the complementary cumulative distribution and auto-covariance functions and the other that fits the first three moments and the auto-covariance function.

The 2-MMPP process and these associated inference procedures were evaluated as approximations to 3-MMPP, 5-MMPP, IPP, IDP, Hyper-Exponential On/Off and Self-Similar traffic sources. The comparison of the fitting approaches and the accuracy of the approximations made were assessed through cell loss predictions obtained using the original and modelled traffic streams.

The proposed and studied fitting algorithms can be used in the characterisation of ATM traffic streams, in network planning and dimensioning and in connection admission control procedures.

This work is now being complemented through a comparative study of other fitting algorithms for the 2-MMPP ATM traffic model. Only such an exhaustive task could give us a complete understanding about the potentials of the 2-MMPP model and its inference procedures in the characterisation of multiple ATM traffic types, as well as the applicability conditions of each fitting approach.

6. References

[1] A. Nogueira, P. Salvador and R. Valadas, “Fitting Algorithms for ATM Traffic Models”, Broadband Access Conference-BAC’99, Cracow, Poland, October 11-13 1999.

[2] Baiocchi A, Blefari-Melazzi N, Listanti M., Roveri A. and Winkler R., “Modelling Issues of an ATM Multiplexers within a Bursty Traffic Environment”, Infocom 91, 1991.

[3] Heffes H. and Lucantoni D.M., “A Markov Modulated Characterisation of Packetized

Voice and Data Traffic and Related Statistical Multiplexer Performance", IEEE Journal of Selec. Areas on Comm., Vol. 4, No. 6, 1986.

[4] Wang S.S. and Silvester J.A., "An Approximate Model for Performance Evaluation of Real-Time Multimedia Communication Systems", Performance Evaluation 22, 1995.

[5] Zukermann M., "Bandwidth Allocation for Bursty Isochronous Traffic in a Hybrid Switching System", IEEE Transactions on Comm., Vol. 37, No. 12, 1989.

[6] Kang S.H. and Sung D.K., "Two-State MMPP Modelling of ATM Superposed Traffic Streams Based on the Characterisation of Correlated Interarrival Times", Globecom'95, 1995.

[7] Blondia C., "The N/G/1 Finite Capacity Queue", Commun. Statist. - Stochastic Models, Vol. 5, No. 2, 1989.

[8] Baiocchi A. and Blefari-Melazzi N., "Analysis of the Loss Probability of the MAP/G/1/K Queue Part II: Approximations and Numerical Results", Commun. Statist. - Stochastic Models, Vol. 10, No. 4, 1994

[9] Z.M. Yin, "New Methods for Simulation of Fractional Brownian Motion", Journal of Computational Physics, v.127 n.1, August 1996, pp. 66-72.

[10] Mine Caglar, K.R. Krishnan and Iraj Saniee, "Estimation of Traffic Parameters in High-Speed Data Networks", ITC 16, 1999, pp. 867-876.

		Mean	**St. Deviation**	**Hurst Param.**
2-MMPP	Original	0.00731874	0.00936198	0.7758
	moments	0.007537± 1.1e-4	0.009455± 3.5e-4	0.7690± 4.8e-3
	cdf	0.007335± 2.7e-4	0.009058± 5.2e-4	0.7597± 2.7e-3
3-MMPP	Original	0.00100191	0.0124362	0.6996
	moments	0.000704± 1.7e-4	0.000702± 8.3e-5	0.7040± 1.9e-2
	cdf	0.001822± 5.7e-4	0.001826± 1.7e-4	0.6968± 3.7e-2
5-MMPP	Original	0.000439233	0.00207648	0.6945
	moments	0.000807± 0.5e-4	0.000809± 1.7e-5	0.687± 2.1e-2
	cdf	0.000958± 2.2e-4	0.000954± 3.2e-4	0.708± 1.6e-2
IPP	Original	0.00249274	0.0613735	0.6282
	moments	0.004478± 2.2e-4	0.084272± 5.7e-3	0.6878± 1.3e-1
	cdf	0.007026± 1.8e-4	0.020571± 4.4e-3	0.7001± 9.8e-2
IDP	Original	0.0856788	0.0953909	0.5362
	moments	**Not-applicable**		
	cdf	0.094404± 2.4e-3	0.07248± 7.7e-3	0.6532± 2.7e-2
H_2 On-Off	Original	0.00196964	0.0716574	0.6100
	moments	0.0010624± 5.7e-4	0.001063± 9.7e-4	0.6720± 3.8e-2
	cdf	0.0130197± 1.7e-2	0.0129845± 2.9e-3	0.6530± 1.7e-2
S.S (H=0.8)	Original	0.000977045	0.000090313	0.7967
	moments	0.000492± 3.1e-4	0.0004916± 4.1e-4	0.6989± 1.1e-2
	cdf	0.000653± 4.7e-4	0.0006520± 3.7e-4	0.6780± 1.5e-2

Table 1: Statistics of original and modelled (by the moments and cdf fitting procedures) traffic streams.

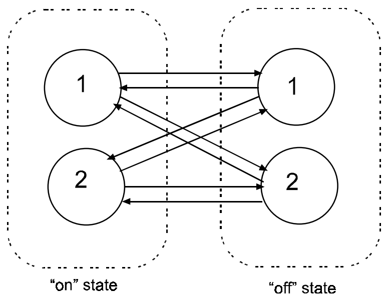

Figure 1: Second-stage hyper-exponential (H_2) On-Off source (transitions are only allowed between states residing in the different groups and not between the states belonging to the same group. As a result, the source is only allowed to alternate between the "on" and "off" states).

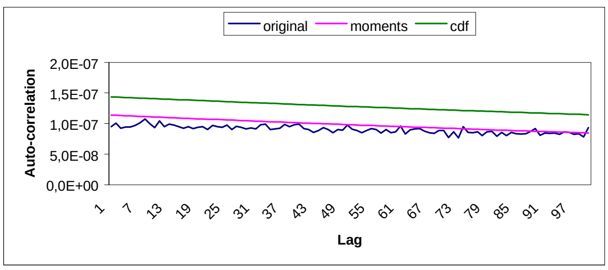

Figure 2a: Auto-correlation function (ACF) for the 2-MMPP original (O) and modelled traffic (M)

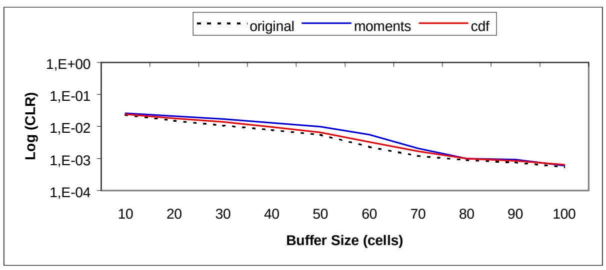

Figure 2b: CLR comparison for the 2-MMPP O and M traffic

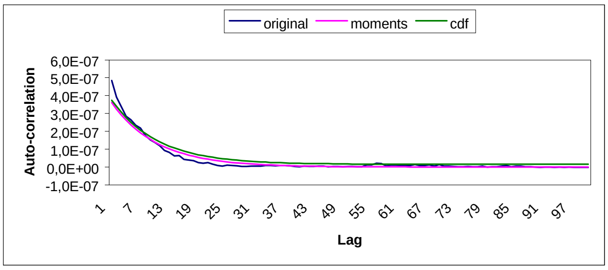

Figure 3a: ACF for the 3-MMPP O and M traffic

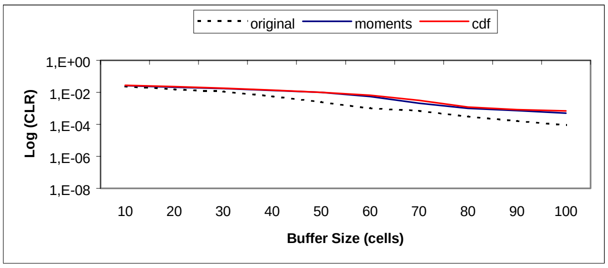

Figure 3b: CLR comparison for the 3-MMPP O and M traffic

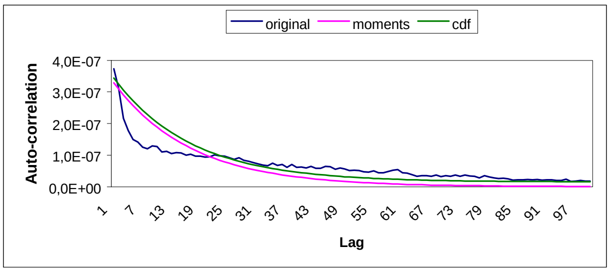

Figure 4a: ACF for the 5-MMPP O and M traffic

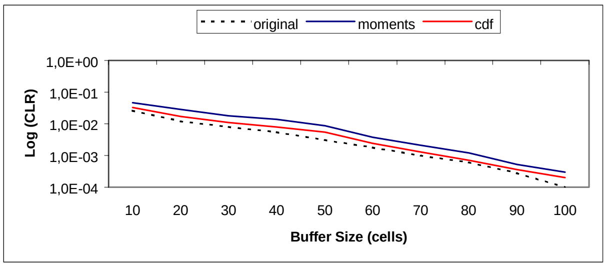

Figure 4b: CLR comparison for the 5-MMPP O and M traffic

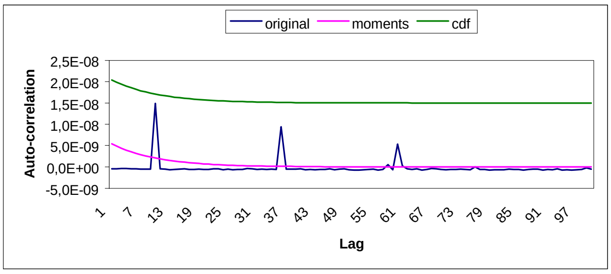

Figure 5a: ACF for the IPP O and M traffic.

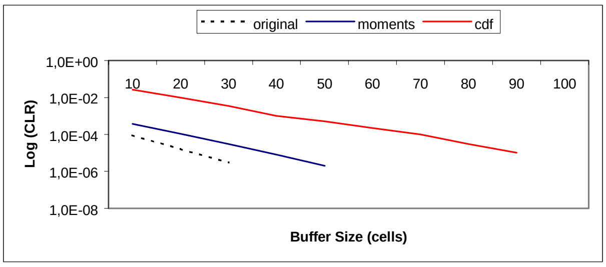

Figure 5b: CLR comparison for the IPP O and M traffic

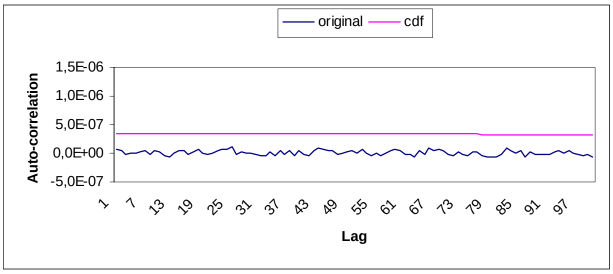

Figure 6a: ACF for the IDP O and M traffic

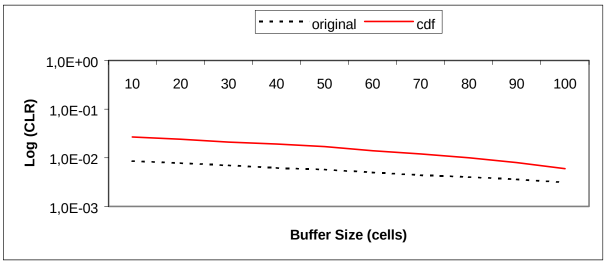

Figure 6b: CLR comparison for the IDP O and M traffic

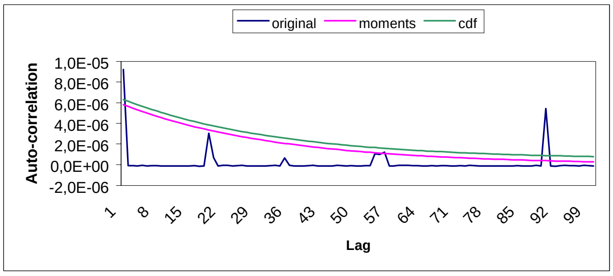

Figure 7a: ACF for the hyper On-Off O and M traffic.

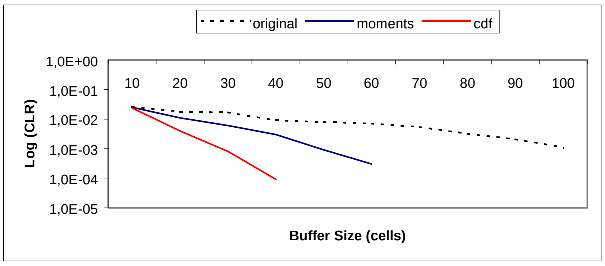

Figure 7b: CLR comparison for the hyper On-Off O and M traffic

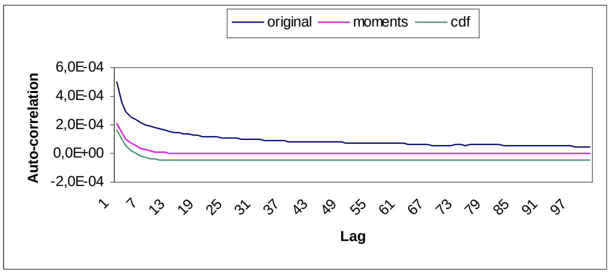

Figure 8a: ACF for the Self-similar O and M traffic.

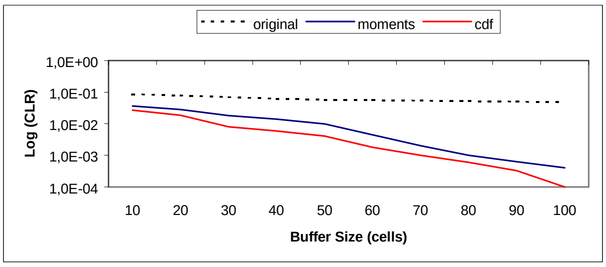

Figure 8b: CLR comparison for the Self-similar O and M traffic