Identification of the First Order Parameters of a Circulant Modulated Poisson Process[1]

Katrien De Cock and Bart De Moor[2]

Published in Proceedings of the International Conference on Telecommunications (ICT'98), Porto Carras, Greece, June 1998, pp. 420-424 (vol. II).

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/decock/reports/97-104.ps.gz*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *katrien.decock@esat.kuleuven.ac.be, bart.demoor@esat.kuleuven.ac.be*. Bart De Moor is a senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders). Katrien De Cock is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). Work supported by the **Flemish Government** (Administration of Science and Innovation (Concerted Research Action MIPS: *Model-based Information Processing Systems*, Bilateral International Collaboration: *Modelling and Identification of nonlinear systems*, IWT-Eureka SINOPSYS: *Model-based structural monitoring using in-operation system identification*), FWO-Vlaanderen: *Analysis and design of matrix algorithms for adaptive signal processing, system identification and control, based on concepts from continuous time system theory and differential geometry*, *Numerical algorithms for subspace system identification: Extension towards specific applications*, FWO-Onderzoeksgemeenschappen: *Identification and Control of Complex Systems*, *Advanced Numerical Methods for Mathematical Modelling*); **Belgian Federal Government**(Interuniversity Attraction Pole IUAP IV/02: *Modelling, Identification, Simulation and Control of Complex Systems*, Interuniversity Attraction Pole IUAP IV/24: IMechS: Intelligent Mechatronic Systems); **European Commission:** (Human Capital and Mobility: *SIMONET: System Identification and Modelling Network*, *SCIENCE-ERNSI: European Research Network for System Identification.*)

Identification of the First Order Parameters of a Circulant Modulated Poisson Process

Katrien De Cock and Bart De Moor*

K.U.Leuven, Dept. of Electrical Engineering (ESAT),
Research group SISTA,
Kardinaal Mercierlaan 94,
3001 Leuven, Belgium,
Tel. 32/16/32 19 25, Fax 32/16/32 19 70,
WWW: *http://www.esat.kuleuven.ac.be/sista.*
E-mail: *katrien.decock@esat.kuleuven.ac.be*

ABSTRACT

In this paper we discuss a new method for the identification of the first order parameters of a circulant modulated Poisson process. We compare its time complexity and accuracy with the method of Li and Hwang [8].

I. INTRODUCTION

The circulant modulated Poisson process (CMPP) is a restricted version of the Markov modulated Poisson process (MMPP), which is known to be a good model for the arrival processes in telecommunication networks [6, 7, 11]. Applications of these models for ATM networks are found in [6, 4, 10].
A real traffic stream $a(t)$, arriving at a queueing system is generally described by a train of impulses corresponding to message arrivals. In this paper, we consider the stochastic process a_k $(k = 1, 2, \ldots)$, where a_k is the number of cells that arrive during the time interval $(k-1, k\,]$.
The research of Li and Hwang [7] has demonstrated that only the first order and second order statistics of the arrival process have a significant impact on queueing performance.
The task of traffic identification is to construct stochastic models, the first and second order statistics of which match those of the data. In this paper we only deal with the identification of the first order statistics, namely the cumulative distribution function F. This first order identification forms part of a complete identification method which is described in [2], but it could also be integrated in other approaches.
We describe the structure of the CMPP in Section II. The first order statistics of the MMPP and of the CMPP are given in Section III. In Section IV we treat three first order identification methods. In Section *A* we give an overview of the method of Li and Hwang [8] and in Section *B* our new first order identification method is explained. Our approach is based on a new method for the identification of the first order parameters of the MMPP. A simplified, more efficient version of our method is given in Section *C*. A comparison between the different methods is made in Section V. Finally we end with some conclusions in Section VI.

II. THE CIRCULANT MODULATED POISSON PROCESS

A circulant modulated Poisson process (CMPP) of order N consists of an N-state circulant Markov chain in which each state i $(i = 1, 2, \ldots, N)$ represents a Poisson process with rate γ_i. In this paper, we only consider discrete time Markov chains.
Let S_k $(k = 1, 2, \ldots)$ be the stochastic process that denotes the evolution of the state in a Markov manner. The circulant Markov chain is characterised by a circulant stochastic transition matrix Q:

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_N \\ q_N & q_1 & \cdots & q_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ q_2 & q_3 & \cdots & q_1 \end{bmatrix},$$

where

$$\begin{aligned} Q(i,j) &= Pr\{S_{k+1} = j \mid S_k = i\} \\ q_i &\geq 0, \text{ for } i = 1, 2, \ldots, N \\ \sum_{i=1}^{N} q_i &= 1\ . \end{aligned}$$

The notation $Q(i,j)$ represents the entry of Q on the ith row and the jth column.

* K. De Cock is Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). B. De Moor is Senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders). This work was supported by the Flemish Government (BOF (GOA-MIPS), AWI (Bil. Int. Coll.), FWO (projects, grants, res. comm. (ICCoS)), IWT (IWT-VCST (CVT), ITA (ISIS), EUREKA (Sinopsys))), the Belgian Federal Government (IUAP IV-02, IUAP IMechS), the European Commission (HCM (Simonet), TMR (Alapedes), ACTS (Aspect), SCIENCE (ERNSI)), NATO (CRG) and industry (Electrabel).

If $\theta(k)$ is the column vector with i-th entry $Pr\{S_k = i\}$, then

$$\theta(k+1)^T = \theta(k)^T Q \ .$$

Throughout this paper, we will deal exclusively with the steady state case, where the state distribution θ is time independent:

$$\theta^T = \theta^T Q \ .$$

Vector θ is the left eigenvector of the circulant stochastic matrix Q corresponding to eigenvalue 1. It is equal to

$$\theta = \left[\begin{array}{cccc} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{array} \right]^T \qquad (1)$$

because circulant stochastic matrices are double stochastic matrices (see e.g. [1]).
The Poisson parameters of the CMPP are arranged in a column vector $\gamma \in \mathbb{R}^N$, where each element γ_i $(i = 1, 2, \ldots, N)$ is a nonnegative real number, representing the arrival rate of the Poisson process associated with state i.
At time step k, the CMPP (Q, γ) will emit a_k arrivals generated by the Poisson process with the arrival rate γ_i if the state of the Markov chain is i.
The CMPP has several computational advantages compared with the MMPP:

- The eigenvalue decomposition of a circulant matrix can be computed very efficiently. Moreover, the matrix which consists of the eigenvectors of a circulant matrix, is a Fourier matrix. This property is exploited while matching the autocorrelation function by using fast Fourier transforms instead of matrix-vector multiplications. The computational complexity of one function evaluation of the autocorrelation is reduced from $\mathcal{O}(N^3)$ for an MMPP to $\mathcal{O}(N\log(N))$ for a CMPP, where N is the model order. Our identification method for the autocorrelation function is beyond the scope of this paper. We refer to [2] for a complete treatment of the traffic identification problem.
- The eigenvectors of a circulant matrix only depend on the dimension of that matrix, not on its entries. Therefore we can partly uncouple the matching of the first and second order statistics, which reduces the computational complexity (see [2]).

Of course, as will be clear from Section IV, the CMPP will need a larger model order than the MMPP for modelling the same data, but the computational advantages compensate this drawback.

III. THE FIRST ORDER STATISTICS OF MMPP AND CMPP

The first order statistics of the stochastic process a_k $(k = 1, 2, \ldots)$ are described by the cumulative distribution function F, defined as

$$F(l) = Pr\{a_k \leq l\}, \ l \in \mathbb{N} \ .$$

A. The first order statistics of the MMPP

For an MMPP the cumulative distribution is equal to

$$\begin{aligned} F_M(l) &= \sum_{i=1}^{N} \theta_i e^{-\gamma_i} \sum_{j=0}^{l} \frac{\gamma_i^j}{j!} \\ &= \sum_{i=1}^{N} \theta_i F_{\gamma_i}(l), \end{aligned} \qquad (2)$$

where $F_{\gamma_i}(l)$ is the cumulative distribution function of the Poisson process with parameter γ_i. For a general MMPP the left eigenvector of the transition matrix corresponding to eigenvalue 1, θ, is not uniform, i.e. not all entries are equal, in contrast with the state distribution vector of a CMPP (1).

B. The first order statistics of the CMPP

For a CMPP the expression for the cumulative distribution function can be simplified because of the uniform state distribution vector θ (1):

$$F_C(l) = \frac{1}{N} \sum_{i=1}^{N} e^{-\gamma_i} \sum_{j=0}^{l} \frac{\gamma_i^j}{j!} \qquad (3)$$

$$= \frac{1}{N} \sum_{i=1}^{N} F_{\gamma_i}(l) \ . \qquad (4)$$

IV. IDENTIFICATION OF THE FIRST ORDER PARAMETERS

The purpose of traffic identification is to determine a model that matches the measured first and second order statistics of a cell arrival process a_k $(k = 1, 2, \ldots)$. The identification of the first order parameters of a CMPP boils down to the determination of N and γ so that F_C approximates the cumulative distribution function of the data, F_d, as well as possible.
The cumulative distribution function of the data sequence is a staircase function. It is computed as follows:

$$F_d(l) = \frac{1}{m} \sum_{j=0}^{l} \sum_{k=1}^{m} \delta(a_k, j) \ , \qquad (5)$$

where $\delta(i, j)$ is the Kronecker delta and m is the number of data.

A. The first order identification method of Li and Hwang [8]

Li and Hwang consider the *rate* cumulative distribution function:

$$F_r(l) = \lim_{t \to \infty} Pr\{\gamma(t) \leq l\} \ ,$$

where $\gamma(t)$ is the input rate process (see [8]). For a CMPP this is a staircase function, which jumps by $\frac{1}{N}$ at each individual value of γ in ascending order. We call the rate cumulative distribution function of a CMPP F_{r_C}.

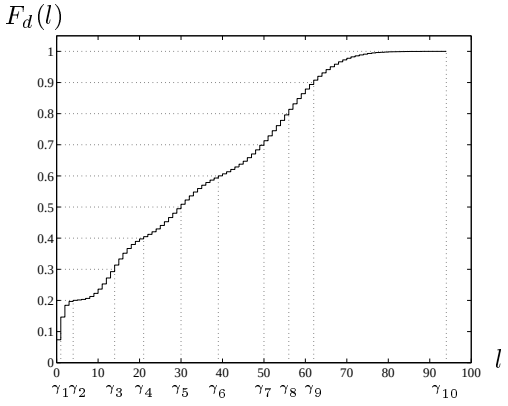

Figure 1: Partitioning $F_d(l)$ into a set of equal-probability rates.

The computed cumulative distribution function of the data, F_d (see (5)), is also treated as a *rate* cumulative distribution. In order for F_d to be matched by F_{r_C}, F_d is discretized by partitioning the range of l into a set of N equal-probability rates.
In figure 1 we show an example of Li and Hwang's method [8] where Poisson parameters γ_i of a 10th order CMPP are determined.
A disadvantage of this method is that the choice of possible γ_i's is restricted by the discretisation step of the cumulative distribution function: the fewer points in which the cumulative distribution function is computed, the less accurate the γ_i's will be.

B. The first order identification of the CMPP, based on the first order parameters of the MMPP

Our new method consists of two steps:

1 Based on the cumulative distribution function of the data a_k, we determine which Poisson processes are involved in the arrival process. The Poisson rates γ_{M_i} $(i = 1, 2, \ldots, N_M)$ and the state distribution vector θ_M of the associated MMPP are obtained by solving a nonnegative least squares problem [5]. The first order MMPP parameters determine the rate cumulative distribution function of the data. The Poisson dynamics are completely eliminated.

2 In the second step we translate the first order MMPP parameters N_M, γ_M and θ_M into the first order CMPP parameters N_C, γ_C and θ_C.

In the following sections we will explain how both steps are performed.

B1. Identification of the first order parameters of an MMPP

The cumulative distribution function of the data, F_d, must be approximated by the cumulative distribution function of an MMPP. This implies that F_d must be reconstructed as a nonnegative linear combination of cumulative distributions of Poisson processes (see (2)). Equation (2) can be rewritten as

$$F_d = D\theta_M \ , \tag{6}$$

where $F_d \in \mathbb{R}^{M+1}$ represents the computed cumulative distribution function of the data (M is the maximum number of arrivals of the given set of data). The columns of the matrix $D \in \mathbb{R}^{(M+1)\times N_M}$ are the cumulative Poisson distribution functions $F_{\gamma_{M_i}}$ $(i = 1, 2, \ldots, N_M)$.
Assume the states (i.e., the γ_{M_i}'s) that form part of the MMPP which models a given data set, are given. The matrix D is then fixed and θ_M can be determined by solving the system of equations (6).
Without knowledge of the states or the number of states, the same approach yields N, γ_M and θ_M if in (6) D is replaced by an enlarged version $\overline{D}$. This matrix $\overline{D}$ has much more columns than D and the same number of rows. Each column $\overline{D_j}$ of $\overline{D}$ represents a *possible* state or *possible* Poisson cumulative distribution. In fact, the domain of possible γ_{M_i}'s, e.g., $[0.001, M]$, is discretized with a step h in order to get a broad choice of states: $\gamma_{M_j} = \gamma_{M_1} + (j-1)h$. The first Poisson parameter, γ_{M_1}, has to be chosen small enough in order to incorporate possible zero-valued a_k's. The computed cumulative distribution function F_d must be reconstructed as a nonnegative linear combination of the columns of $\overline{D}$:

$$\min_x \|F_d - \overline{D}x\|_2, \text{ subject to } x_i \geq 0 \ . \tag{7}$$

Let x^* be the optimal value of x. The number of non-zero components of x^* gives the MMPP model order N_M. The indices of the non-zero components of x^* give the γ_{M_i}'s. The θ_{M_i}'s are given by the values of the non-zero components of x^*.
The matrix $\overline{D}$ can be constructed before one starts identifying. If the matrix $\overline{D}$ is formed over a large range of γ_{M_i}'s, $\overline{D}$ can be adjusted by the program to a given data set. This adjustment is done by taking a submatrix of $\overline{D}$ and is explained in figure 2. All the columns which represent cumulative Poisson distributions with parameter γ larger than the maximum value of the data set are eliminated. The rows

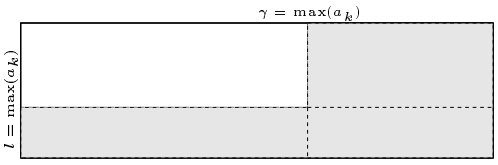

Figure 2: A representation of the adaptation of the original matrix $\overline{D}$ (the large rectangle) to the given set of data. The grey parts are left out because they do not contain valuable information. All the columns which represent cumulative Poisson distributions with parameter γ larger than the maximum value of the data set are eliminated. The rows which correspond to samples of the cumulative distribution $F_m(l)$, where $l > \max(a_k)$ are also left out. The nonnegative least squares problem (7) is solved for the resulting smaller matrix.

which correspond to samples of the cumulative distribution $F_m(l)$, where $l > \max(a_k)$ are also left out (see figure 2).

B2. Translation of the MMPP parameters into CMPP parameters

To conform the results of (7) to the state distribution vector θ_C of a CMPP, $\left[\begin{array}{cccc} \frac{1}{N_C} & \frac{1}{N_C} & \cdots & \frac{1}{N_C} \end{array}\right]^T$, a translation of N_M, γ_M and θ_M into N_C, γ_C and θ_C is necessary. In [2] a theoretical way to translate the parameters is described. A more practical translation is also mentioned in [2]. It is this method that we discuss here.

We assume that the order of the CMPP, N_C, is given. Values that are very suitable for N_C are powers of 2. This is because we use `fft`'s (fast Fourier transforms) to obtain the transition matrix Q which matches the autocorrelation (see [2]). If N_C is a power of 2 we use a fast radix-2 fast Fourier transform algorithm. Li and Hwang [8] found that taking $N_C = 100$ is sufficient in most cases of their study. Therefore we take $N_C = 128 = 2^7$.

Because we know γ_M and θ_M, the rate cumulative distribution function F_r can be computed: F_r is a staircase function which jumps by θ_{M_i} at each value γ_{M_i} in ascending order. Instead of discretising this rate cumulative distribution function as is done by Li and Hwang [8], we discretise the jumps of the rate cumulative distribution function. In other words, we discretise the rate probability distribution function instead of the rate cumulative distribution function. Each value of the state distribution vector θ_M is divided in pieces of size $\frac{1}{N_C}$. This means that each MMPP state is copied several times. The number of copies of MMPP state i is called t_i ($i = 1, 2, \ldots, N_M$). To obtain the t_i's a rounding off is executed:

$$\begin{aligned} t_i &= \texttt{floor}\left(\frac{\theta_{M_i}}{\frac{1}{N_C}}\right), \text{ for } i = 1, 2, \ldots, N_M \\ &= \texttt{floor}(\theta_{M_i} N_C), \text{ for } i = 1, 2, \ldots, N_M \ , \end{aligned}$$

where $\texttt{floor}(\theta_{M_i} N_C)$ is the largest integer which is less than or equal to $\theta_{M_i} N_C$. If $\sum_{i=1}^{N_M} t_i < N_C$ (by discretisation errors) we add 'units' $\frac{1}{N_C}$ to the states for which the error was the largest. The obtained γ_C is then equal to:

$$\left[\underbrace{\gamma_{M_1}\ \gamma_{M_1}\ \cdots\ \gamma_{M_1}}_{t_1 \text{ times}}\ \underbrace{\gamma_{M_2}\ \cdots\ \gamma_{M_2}}_{t_2 \text{ times}}\ \cdots\ \underbrace{\gamma_{M_{N_M}}\ \cdots\ \gamma_{M_{N_M}}}_{t_{N_M} \text{ times}}\right]^T .$$

C. Simplification of our identification method

Our approach is more time consuming (see Section V) than the method of Li and Hwang [8] because of the first step (see Section *B1*). This step can be omitted by treating the probability distribution function (histogram) of the data a_k as a rate probability distribution function and discretising this without eliminating the Poisson dynamics.

V. COMPARISON

In this section we compare the accuracy and the time complexity of the method of Li and Hwang [8] with our new method and its simplified version.

A. Accuracy and time complexity

As a measure for the accuracy of the identification method we take the 2-norm of the difference between the cumulative distribution function of the data (see (5)) and the cumulative distribution function computed with the identified CMPP parameters (see (3)):

$$\text{error} = \|F_d - F_C\|_2 \ . \tag{8}$$

The time complexity is measured by the CPU time needed to identify the first order parameters of the CMPP. In figure 3 the identification error (left picture) and the CPU time in seconds (right picture) of three different methods are compared. Our new method described in Section IV is drawn in solid, the method of Li and Hwang [8] is represented by the dash-dotted line and the simplified version of our method (see Section IV) is given by the dotted line. The identification error and the CPU time are given in function of the imposed model order N_C. Figure 3 represents the average of 11 identification procedures in which simulated data are used.

The complete method produces the most accurate results, but it is clearly more time consuming than the other methods (see figure 3). The simplified version of our method is more efficient from model order $N_C \approx 50$ on than the method of Li and Hwang [8] (see right picture of figure 3), and its error is bounded from above by the error of Li and Hwang's method (left picture of figure 3).

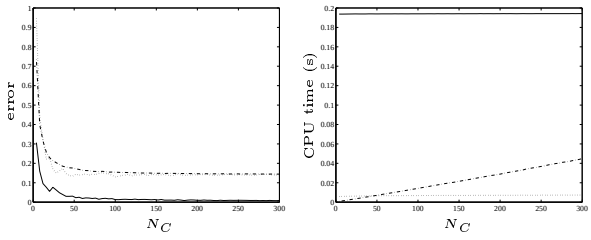

Figure 3: The error (left) and the CPU time in seconds (right) of different identification methods in function of the imposed model order N_C. The identification was done starting from simulation data. The solid line shows the results for our new method if both steps are executed. The dash-dotted line represents the identification error and the CPU time for the method of Li and Hwang [8]. The dotted line is obtained by omitting step 1 of our method.

B. A real traffic sequence

In this section we compare the performance of our method, the method of Li and Hwang [8] and the simplified version of our method when dealing with real data. The data are measurements of one hour of internet traffic between the Lawrence Berkeley laboratory and the rest of the world[1]. The traces were made by Vern Paxson [9] and they are available from the Internet Traffic Archive of the Lawrence Berkeley National Laboratory [12].
The results are shown in figure 4. The differences between the different methods is now smaller than for the simulation data. The CPU time needed to perform step 1 and step 2 from Section IV is of course still much larger than the CPU time needed for the other methods.

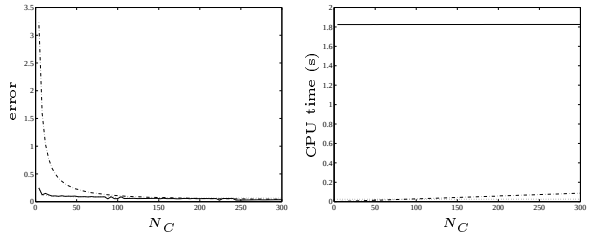

Figure 4: The error (left) and the CPU time in seconds (right) of different identification methods in function of the imposed model order N_C. The data we used are real internet data. The solid line shows the results for our new method if both steps are executed. The dash-dotted line represents the identification error of the method of Li and Hwang [8]. The dotted line is obtained by omitting step 1 of our method.

VI. CONCLUSIONS

In this paper we have explained a new method to determine the first order parameters of a circulant modulated Poisson process (CMPP). This method consists of two steps. The first step is the identification of the first order parameters of a general Markov modulated Poisson process (MMPP). In the second step the MMPP parameters are translated into CMPP parameters.
This new method is more accurate than the method of Li and Hwang [8] but it is more time consuming. However, by simplifying our method, we obtain an efficient and accurate alternative for Li and Hwang's approach.

[1]Our specific dataset is available from DAISY (Database for the Identification of Systems) at URL *http://www.esat.kuleuven.ac.be/sista/daisy/* [3].

REFERENCES

[1] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, Philadelphia: SIAM, Classics in Applied Mathematics series, vol. 9.

[2] K. De Cock, T. Van Gestel and B. De Moor. "Stochastic System Identification for ATM Network Traffic Models: a Time Domain Approach", Internal report ESAT-SISTA/97-90[2], 1997.

[3] B. De Moor, P. De Gersem, B. De Schutter and W. Favoreel. "DAISY: A Database for Identification of Systems", Journal A (Benelux publication of the Belgian Federation of Automatic Control), Special Issue on CACSD, vol. 38, no. 3, pp. 4-5, September 1997.

[4] W.C. Lau and S.Q. Li. "Statistical Multiplexing and Buffer Sharing in Multimedia High-speed Networks: A Frequency Domain Perspective", IEEE Globecom 1995 Conference.

[5] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*, Philadelphia: SIAM, Classics in Applied Mathematics series, vol. 15.

[6] S.Q. Li and C.L. Hwang. "Queue Response to Input Correlation Functions: Discrete Spectral Analysis", IEEE/ACM Transactions on Networking, vol. 1, 1993, pp. 522-533.

[7] S.Q. Li and C.L. Hwang. "Queue Response to Input Correlation Functions: Continuous Spectral Analysis", IEEE/ACM Transactions on Networking, vol. 1, 1993, pp. 678-692.

[8] S.Q. Li and C.L. Hwang. "On the Convergence of Traffic Measurement and Queueing Analysis: A Statistical-Matching and Queueing (SMAQ) Tool", IEEE/ACM Transactions on Networking, vol. 5, 1997, pp. 95-110.

[9] V. Paxson and S. Floyd. "Wide-Area Traffic: The Failure of Poisson Modeling", IEEE/ACM Transactions on Networking, vol. 3, 1995, pp. 226-244.

[10] K. Spaey and C. Blondia. "Circulant Matching Method for Multiplexing ATM Traffic Applied to Video Sources", accepted for publication in Proceedings of IFIP TC 6 Conference on Performance of Information and Communication Systems '98.

[11] T. Rydén. "Parameter Estimation for Markov Modulated Poisson Processes", Stochastic Models, no. 10, 1994, pp. 795-829.

[12] Internet address: *http://ita.ee.lbl.gov/html/contrib/LBL-PKT.html*, lbl-pkt-4.

[2]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/decock/reports/97-90.ps.gz*