96-63

Yi C., De Moor B., ``Traffic identification of ATM networks with optimization algorithms", in *Proc. of the 35th IEEE Conference on Decision and Control (CDC'96)*, Kobe, Japan, Dec. 1996, pp. 277-282., Lirias number: 179579.

Traffic identification of ATM networks with optimization algorithms

Cheng YI, Bart De Moor*
ESAT - Katholieke Universiteit Leuven,
Kardinaal Mercierlaan 94,
B-3001 Leuven (Heverlee), Belgium,
tel 32/16/22 09 31, fax 32/16/22 18 55,
cheng.yi@esat.kuleuven.ac.be
bart.demoor@esat.kuleuven.ac.be

Abstract

In this research work, we cast the traffic identification problem for multimedia network (ATM e.g.) as an optimization problem and solve it by general optimization algorithms.

1 Introduction

Traffic modeling of multimedia networks consists of two steps: traffic identification and queuing analysis . This research work is mainly concerned with traffic identification.

The input (or arrival) traffic process in high speed multimedia networks is strongly correlated and is modeled by a so-called Markov Modulated Poisson Process (MMPP) [7, 6]. Let $P = [p_{ij}]$ be the one-step state transition matrix (also called stochastic matrix) of an n-state Markov Chain (MC). That is $P \in \mathbb{R}^{n\times n}$, $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$. Assume $\gamma = [\gamma_i] \in \mathbb{R}^N$, where $\gamma_i \geq 0$ is the arrival rate of the Poisson process associated with state i. Then at time k, the MMPP (P, γ) will emit $a(k)$ arrivals generated by the Poisson process with the arrival rate γ_i if the state of the MC is i at time k. See Fig. 1. Note that $a(k)$ can be measured. The aim of traffic identification is to find P and γ from arrivals $a(k)$. P and γ will then be used in queuing analysis.

In this work, we will present one step approach for the traffic identification which is based optimization algorithms.

2 The first and the second order statistics of arrivals

As already mentioned, the ultimate goal of the traffic identification is to find the parameters (P, γ) of an MMPP for analyzing queue responses to this MMPP. For the following two reasons, only the first and the second statistics of arrivals $a(k)$ are considered. First, the queuing performance is found to be much less dependent on the higher-order statistics of $a(k)$, second,

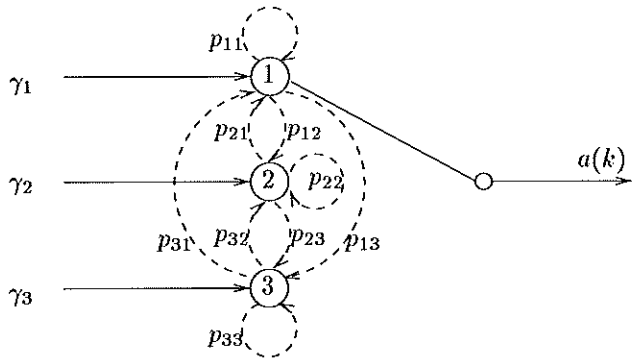

Figure 1: A 3-state MMPP.

the first and the second statistics are easy to be measured in practice.

For steady-state queuing analysis, $a(k)$ is assumed to be a stationary random process and only the steady-state statistics are considered.

2.1 Cumulative distribution function of arrival $a(k)$

The first statistics is described by Cumulative Distribution Function (CDF) $F(x)$ of the arrivals $a(k)$ defined as:

$$F(x) = \lim_{k\to\infty} \Pr\left(a(k) \leq x\right),$$

which is typically expressed as a piecewise step function.

For a MMPP,

$$\lim_{k\to\infty} \Pr(\gamma(k) = \gamma_i) = \pi_i,$$

where $\gamma(k)$ stands for the input rate process (without the local Poisson processes), where $\pi = [\pi_i]$ is the left eigenvector corresponding to the eigenvalue 1 of the stochastic matrix P: $\pi P = \pi$ [3]. Thus one has:

$$\begin{aligned} F(x) &= \lim_{k\to\infty} \Pr\left(a(k) \leq x\right) \\ &= \lim_{k\to\infty} \sum_{j=1}^{N} \Pr(\gamma(k) = \gamma_j)\Pr(a(k) \leq x | \gamma(k) = \gamma_j) \\ &= \sum_{j=1}^{N} \pi_j F_{\gamma_j}(x), \end{aligned}$$

*Bart De Moor is a senior research associate of the N.F.W.O. The following text presents research results obtained within the framework of the Belgian programme on interuniversity attraction poles (IUAP-17 and IUAP-50) initiated by the Belgian State - Prime Minister's Office - Science Policy Programming. The scientific responsibility is assumed by its authors. The research work was supported by the F.K.F.O. project G0292.95.

where $F_{\gamma_j}(x)$ is the cumulative distribution function for a single Poisson process:

$$F_{\gamma_j}(x) = e^{-\gamma_j} \sum_{i=0}^{\text{floor}(x)} \frac{\gamma_j^i}{i!}.$$

2.2 Autocorrelation function of arrivals $a(k)$

The second statistics of $a(k)$ is characterized by the autocorrelation function $R(n)$ of $a(k)$ and defined as:

$$R(0) \triangleq E(a(k)a(k)) = \pi(\Gamma + \Gamma^2)e,$$

$$R(n) \triangleq E(a(k)a(k+n)) = \pi\Gamma P^n \Gamma e, \quad (1)$$

where $\Gamma = \text{diag}(\gamma)$, $e = [\ 1\ 1, \ldots, 1\ ]^T$, see [8] for the proof. Now let $C = \pi\Gamma$ and $G = \Gamma e$, then

$$CP^{n-1}G = \begin{cases} R_\gamma(0) = R(0) - \pi\Gamma e, & n = 1, \\ R(n-1), & n > 1. \end{cases} \quad (2)$$

This last observation indicates that the autocorrelation function of the arrival process $a(k)$ basically consists of the Markov parameters of a deterministic linear time invariant system with state space model (P, G, C)(Note that $\pi\Gamma e = \text{mean}(a(k))$, so it can be easily calculated). This is an important observation that might lead to use stochastic subspace identification algorithms [11, 10] we have developed for linear stochastic systems. We will discuss this problem in another paper. Later on we will use this property to determine the order (the number of states) of a MMPP from the measurements by checking the rank or the singular values of the Hankel matrix formed with the $R(n)$.

3 Optimization approach to statistics matching

Suppose that the measured cumulative distribution function and autocorrelation function are $F_m(x)$ and $R_m(n)$. Now the traffic identification problem is to find a pair of (P, γ) such that the first and the second statistics $F(x)$ and $R(n)$ of the corresponding MMPP matches the measured $F_m(x)$ and $R_m(n)$.

3.1 Formulation as an optimization problem

Obviously the problem can be cast as the following optimization problem:

$$\min_{P:\ \mathbb{P},\ \gamma \geq 0} J(P, \gamma) \quad (3)$$

where $\mathbb{P}$ stands for the set of all stochastic matrices, the cost function $J(P, \gamma)$ is defined as

$$J(P, \gamma) \triangleq W_F \|(F(x) - F_m(x))\| + W_R \|R(n) - R_m(n)\| \quad (4)$$

where $\|\cdot\|$ stands for 2-norm:

$$\|(F(x) - F_m(x))\|^2 = \int_0^\infty (F(x) - F_m(x))^2 dx, \quad (5)$$

$$\|(R(n) - R_m(n))\|^2 = \sum_{i=0}^{\infty} (R(i) - R_m(i))^2, \quad (6)$$

and W_f and W_r are two weighting factors. When $W_f = 0$, only the second statistics is matched, or when $W_r = 0$, only the first statistics is matched. If one find a P and a γ such that $J(P, \gamma) = 0$, then the problem is solved.

3.2 Reformulation as an unconstrained optimization problem - parameterization

Let $A \in \mathbb{R}^{n \times n}$, define the following map: $A \rightarrow (P, \gamma)$:

$$P(A) = [p_{ij}],\ p_{ij} = \begin{cases} \frac{1}{1 + \sum_{k=1}^{n-1} a_{ik}^2}, \text{for } j = 1,\ \forall i, \\ \frac{a_{ij}^2}{1 + \sum_{k=1}^{n-1} a_{ik}^2}, \text{for } j \neq 1,\ \forall i. \end{cases}$$

$$\gamma(A) = [\gamma_i],\ \gamma_i = a_{in}^2, \forall i.$$

Now the constrained nonlinear optimization problem of (3) over the stochastic matrix set $\mathbb{P}$ and the positive vector set is changed to an unconstrained nonconvex optimization problem over $\mathbb{R}^{n \times n}$:

$$\min_{A \in \mathbb{R}^{n \times n}} J(P(A), \gamma(A)). \quad (7)$$

One can use any general unconstrained optimization algorithm to solve the problem.

3.3 Calculation of norms

The exact calculation of $\|F(x) - F_m(x)\|$ is time consuming. We use the following approximation:

$$\|(F(x) - F_m(x))\|^2 \approx \sum_{i=1}^{N_f} (F(x_i) - F_m(x_i))^2.$$

Note also that N_f need not to be very large. It is in fact good enough for N_f to be about 10 points with equalized interval points x_i covering all the possible range [0 max$(a(k))$]. See a numerical example later on.

The calculation of $\|R(n) - R_m(x)\|$ is straight forward. However instead calculating infinity number of terms in (6), only first N_r terms are calculated, N_r should be larger than the order of the MMPP which will discussed later on. For each correlation term,

$$R(n) = \lim_{j \rightarrow \infty} \sum_{k=1}^{j} \frac{a(k)a(k+n)}{j}$$

is approximated by

$$R(n) = \sum_{k=1}^{i} \frac{a(k)a(k+n)}{i}$$

where $i+n$ is the number of total data points. Since the steady state statistics is used, i should be large enough.

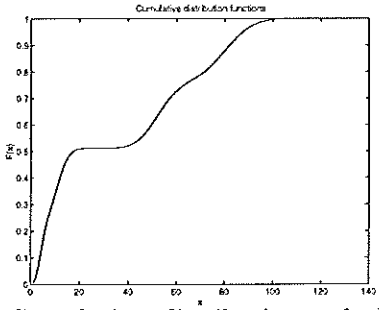

Figure 2: Cumulative distributions of the identified MMPP (——), data of arrivals (– – –) and the original MMPP (– · – ·)

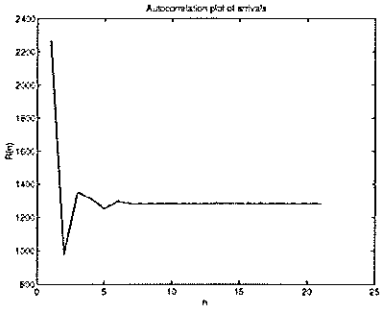

Figure 3: Autocorrelations of the identified MMPP (——), data of arrivals (– – –) and the original MMPP (– · – ·)

3.4 Example 1

We used Matlab function 'fminu' to solve the unconstrained optimization problem of (7). The algorithm in 'fminu' is the BFGS [2] quasi-Newton method using a mixed quadratic and cubic line search procedure.

The arrival data $a(k)$ is generated by an MMPP with

$$P = \begin{bmatrix} 0.3514 & 0.2521 & 0.3648 & 0.0317 \\ 0.2045 & 0.1508 & 0.4315 & 0.2132 \\ 0.1831 & 0.0228 & 0.0079 & 0.7862 \\ 0.2746 & 0.5123 & 0.2131 & 0.0001 \end{bmatrix},$$

and

$$\gamma = \begin{bmatrix} 11 & 80 & 52 & 4 \end{bmatrix}.$$

see [9] for the algorithm for MMPP simulation. The number of data points is 100,000. $W_f = 10$ and $W_r = 1$.

The optimization algorithm with random initial values converges to the following point:

$$P_m = \begin{bmatrix} 0.0001 & 0.7482 & 0.1224 & 0.1293 \\ 0.2720 & 0.0790 & 0.4970 & 0.1519 \\ 0.4717 & 0.0088 & 0.1118 & 0.4078 \\ 0.1790 & 0.2331 & 0.2691 & 0.3189 \end{bmatrix},$$

and

$$\gamma_m = \begin{bmatrix} 80.0184 & 3.9748 & 52.0406 & 10.9785 \end{bmatrix}.$$

The corresponding cost is 4.11. Note that though P and P_m are not the same, but the eigenvalues of P:

$$1.0000, \quad -0.3019 \pm 0.3330i, \quad 0.1140$$

and the eigenvalues of P_m

$$1.0000, \quad -0.3068 \pm 0.3280i, \quad 0.1233$$

are almost same.

Fig. 2 and 3 shows the cumulative distributions and the autocorrelations of the identified MMPP (P_m, γ_m), data of arrivals $a(k)$ and the original MMPP (P, γ) respectively. The differences between them are very small. This means that both simulation algorithm and identification algorithm work quite well.

4 MMPPs with double stochastic matrices

One problem related to above optimization based MMPP identification algorithm for MMPPs with stochastic matrices is high cost of computation. This is basically because of two reasons: eigenvalue decomposition to fine π (the left eigenvector of the stochastic matrix P) and matrix multiplication for P^k. In this section, a double stochastic Markov jumping matrix, instead of a general row stochastic Markov jumping matrix, is used such that the eigenvalue decomposition will be eliminated.

4.1 Doubly stochastic matrix

A special subset of stochastic matrices is the so-called doubly stochastic matrices. A stochastic matrix P with the property that P^T is also stochastic is said to be doubly stochastic, that is, P is both row and column stochastic and thus all row and column sums are equal to 1. Further, a nonnegative matrix P is doubly stochastic matrix if and only if both $Ae = e$ and $e^T A = e^T$, which means that the steady state probability of each state is uniform for the Markov Chain with doubly stochastic jumping matrix. The later is crucial since it shows that e^T is the left eigenvector of any doubly stochastic matrix associated with the eigenvalue 1, that is

$$\pi = e^T/n. \tag{8}$$

This leads to elimination of eigenvalue decomposition during optimization procedure.

4.2 Parameterization

Birkhoff's theorem [3] gives a way for parameterization of all doubly stochastic matrices:

Theorem 1 (Birkhoff) *A matrix $P \in \mathbb{R}^{n\times n}$ is a doubly stochastic if and only if for some $N \leq \infty$ there are permutation matrices $P_1, \ldots, P_N \in \mathbb{R}^{ntimesn}$ and positive scalars $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ such that $\alpha_1 +, \ldots, +\alpha_N = 1$, $P = \alpha_1 P_1 +, \ldots, +\alpha_N P_N$.*

The proof [3] relies on the facts that the set of doubly stochastic matrices is a compact convex set in $\mathbb{R}^{n\times n}$ and that every point in a compact convex set is a convex combination of the extreme points of the set. The extreme points of the set of doubly stochastic matrices are precisely the permutation matrices. In [3], it is shown that $N \leq n^2 - 2n + 2$.

Now we can map $a \in \mathbb{R}^{N-1}$ to $\alpha \in \mathbb{R}^N$ with $\alpha_i \geq$ and $\alpha_1 +, \ldots, +\alpha_N = 1$::

$$\alpha_i = \begin{cases} \frac{1}{1+\sum_{k=1}^{N-1} a_k^2}, \text{for } i = 1, \\ \frac{a_k}{1+\sum_{k=1}^{N-1} a_k^2}, \text{for } i \neq 1. \end{cases}$$

However both Birkhoff's theorem and its proof do not show which permutation matrices should be used for the convex combination. Thus the only possibility to parameterize all set of doubly stochastic matrices is to make a convex combination of all possible extreme points or permutation matrices in $\mathbb{R}^{n \times n}$. In this case, $N = n!$. Obviously the main drawback of this method is over-parameterization, the dimension of the parameters over which the cost function of (4) is optimized increases in a combinatorial explosion way with the size n of the states of MMPP. So the method can only be used for small n.

5 MMPPs with circulant stochastic matrices

In the previous section, one sees that though the eigenvalue decomposition problem is circumvented by use of doubly stochastic matrices, the number of the parameters increases in a combinatorial explosion way with the number of states of a MMPP, which indeed increases the computational costs.
In this section, a special subset of doubly stochastic matrices, called circulant stochastic matrices, will be considered. By using a circulant matrix for a MMPP, one will not only circumvented the eigenvalue decomposition but also matrix multiplication in the autocorrelation function calculation (1) in optimization algorithms.

5.1 Circulant stochastic matrix

A circulant stochastic matrix is defined as:

$$P = \sum_{k=1}^{n} \alpha_k C^k,$$

where C is the so called basic circulant permutation matrix:

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

and α_i are positive scalars and $\sum_{i=1}^{n} \alpha_i = 1$.
Obviously, since a circulant stochastic matrix is also a doubly stochastic matrix, $\pi = e^T/n$. Another very nice property of circulant stochastic matrices is that any eigenvector of a circulant matrix is independent from parameters α_i and any eigenvalue is just a linear combination of parameters α_i [6]:

$$P = V \Lambda V', \tag{9}$$

where V is a unitary matrix with:

$$[v_{ij}] = \frac{1}{n} e^{\frac{2\pi ij}{n}\sqrt{-1}},$$

and

$$\Lambda = \text{diag}(\sqrt{n}\alpha V),$$

with $\alpha \triangleq [\alpha_1, \ldots, \alpha_n]$.

5.2 Parameterization

The parameterization is the same as for doubly stochastic matrix by mapping $a \in \mathbb{R}^{n-1}$ to $\alpha \in \mathbb{R}^n$ with $\alpha_i \geq 0$ and $\alpha_1 +, \ldots, +\alpha_n = 1$:

$$\alpha_i = \begin{cases} \frac{1}{1+\sum_{k=1}^{n-1} a_k^2}, \text{for } i = 1, \\ \frac{a_k}{1+\sum_{k=1}^{n-1} a_k^2}, \text{for } i \neq 1. \end{cases}$$

Note that the number of the undetermined parameters for a circulant matrix is equal to $n-1$, while it is $n(n-1)$ for a stochastic matrix and $n!-1$ for a doubly stochastic matrix.
Now the correlation function can be rewritten as:

$$R(n) = \pi \Gamma P^n \Gamma e = \gamma V \Lambda^n V' \gamma'/n.$$

Note that V is constant and Λ is diagonal. Thus both eigenvalue decomposition and matrix $(n \times n)$ multiplication are avoided in calculation of the cost function.

6 Orders of MMPPs

The problem is how to determine the order or the number of states of a MMPP with a certain model from the given data of arrivals: what is the smallest order of a MMPP with a certain model which matches the statistics of the given data of the arrivals? Suppose that only the second statistics of the data is to be matched. Then this problem looks like the realization problem in system theory: given Markov parameters, determine the order of the system matrix. However it is in fact much more difficult than the realization problem for linear time invariant systems [5] because of the constrains imposed on P (stochastic, doubly stochastic or circulant) and $C(=\pi\Gamma)$ and $G(=\Gamma e)$. It is even more difficult than the nonnegative realization problem [1] , which remains to be solved, since C and G are not only nonnegative here.
However we can use the realization theory of linear systems to get a lower bound of the order for given data. Let H be a Hankel matrix formed by autocorrelation functions:

$$H = \begin{bmatrix} R_\gamma(0) & R(1) & R(2) & \cdots & R(i) \\ R(1) & R(2) & R(3) & \cdots & R(i+1) \\ R(2) & R(3) & R(4) & & R(i+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(i) & R(i+1) & R(i+2) & \cdots & R(2i) \end{bmatrix}$$

From system realization theory, the minimal order n_m of the corresponding system is equal to the rank of the Hankel matrix H. i should be large enough such that H is rank deficient.
By choosing n_m as the first try, one then could increase the order if the cost function does not converge to zero or small enough.
Fig. 4, shows the singular values for the Hankel matrices of the identified MMPP (P_m, γ_m), data of arrivals $a(k)$ and the original MMPP (P, γ) for Examples 1 (The largest singular values are not plotted).
One could see that there is a clear cut between the nth singular value and the $(n+1)$th singular value for the example. That is because we deliberately chose each model (P, γ) such that the ratio of the first singular value and the nth singular value is small, which means, in system theory, that the corresponding linear system (P, C, G) is both controllable and observable,

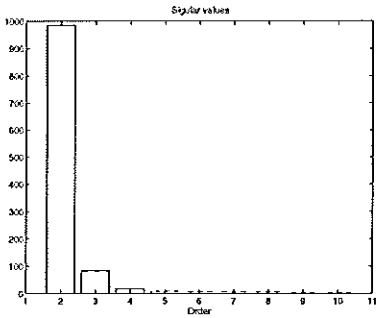

Figure 4: Singular values of the identified MMPP (———), data of arrivals (– – –) and the original MMPP (–·–·–) for Example 1

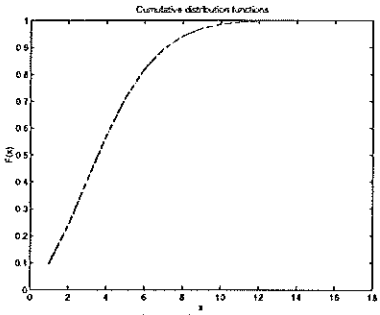

Figure 5: Cumulative distributions of the data of arrivals (– – –) and the original MMPP (– · – ·)

thus the system is of minimum order and can be identified. That is the exact reason why all eigenvalues of the Markov jumping matrices can be recovered.

In previous sections, we discussed three models: MMPPs with a (row) stochastic matrix, with a doubly stochastic matrix and with a circulant stochastic matrix. The most general model is with a stochastic matrix, the model with a circulant stochastic matrix is most restricted one, and the model with a doubly stochastic matrix is in between. Thus one could expect that the orders of models are increasing from the order of the model with a stochastic matrix, a doubly stochastic matrix, to a circulant stochastic matrix. Thus if one expects a low order MMPP, then use the model with a stochastic matrix, but the computation cost can be high, or if one expects fast computing, then use the model with a circulant stochastic matrix, but the order of the MMPP can be high.

7 Comparison with the frequency domain approach

7.1 Comparison

The traffic identification problem was tackled by San Qi Li [6, 4] based on the so-called frequency domain approach. The approach uses only a circulant stochastic matrix. The main differences between our method and Li's method are summarized as follows.

- The first statistics is not directly matched in Li's method. Instead, the Poisson parameters are matched, and the Poisson parameters are from discretization of the cumulative distribution function which leads to very inaccurate results when the order is lower since the order of a MMPP, in this case, is forced to be equal to the number of points where the cumulative distributed function is matched, as Li indicated that 100 states are needed only for the first statistic match. In our method, the first statistics is directly matched by optimizing the cost function, the order of the MMPP is independent of the number of the match points (which is set to be larger then the order), and thus it works for a MMPP with any order.
- Li's method starts from the power spectrum of the data of the arrivals, and thus the second statistics of the data is in fact not directly matched either. Since Li's method works only with circulant stochastic matrices, the order should be high. By contrast, our method starts from the data directly, and the second statistics is also directly matched by optimizing the cost function. Our method also works for any stochastic matrices. The users can choose the models with stochastic, doubly stochastic and circulant stochastic matrices (Note that even for the case of circulant matrices, our method is different from Li's.)
- Computational cost is of course one of the main concerns. Li's method consists of several steps. Our method uses only one step: nonconvex optimization for both first and second statistics matching without pre signal processing. The exact cost comparison is difficult since in both methods, nonconvex optimization procedure is involved. But we expect that our method with circulant statistic matrices is comparable to Li's methods

7.2 Example 2

To complete this section, we take an example from [4]. Consider the integration of three types of 3-state MMPPs. The transition matrix of each type is designed by $P_a = P$, $P_a = P^5$ and $P_c = P^{20}$, where P is a 3×3 Markov transition matrix:

$$P = \begin{bmatrix} 0.1597 & 0.0296 & 0.8107 \\ 0.8104 & 0.0649 & 0.1247 \\ 0.1176 & 0.8753 & 0.0071 \end{bmatrix}$$

The corresponding input rate vectors (of the Poisson processes) are

$$\gamma_a = [0,\ 0.7,\ 0.3],$$
$$\gamma_b = [0,\ 1.2,\ 2.3],$$
$$\gamma_c = [0,\ 1.5,\ 0.9].$$

A superimposed MMPP is assumed to consist of one type-a, two type-b and one type-c elements, then this superimposed MMPP have

$$P = P_a \otimes P_b \otimes P_b \otimes P_c,$$
$$\gamma = \gamma_a \oplus \gamma_b \oplus \gamma_b \oplus \gamma_c.$$

where $\otimes$ stands for Kronecker product and $\oplus$ for Kronecker sum. The number of states of this MMPP is $3^4 = 81$. 100,000 points of arrivals were generated by this MMPP. Fig. 5 and 6 show the cumulative distribution functions and the autocorrelation functions of the model $(P,\ \gamma)$ and the data respectively. Now we use a 3-state MMPP with a stochastic matrix and a 7-state MMPP with a circulant stochastic matrix to match the statistics of the data. $N_f = 10$ for the first statistics matching and $N_r = 20$ for the second statistics matching. The weighting

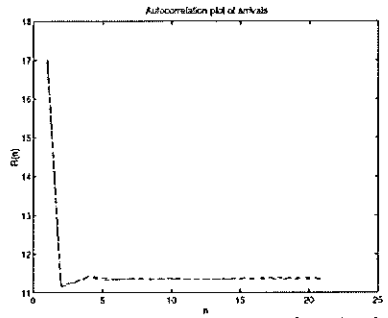

Figure 6: Autocorrelations of data of arrivals (– – –) and the original MMPP (– · – ·)

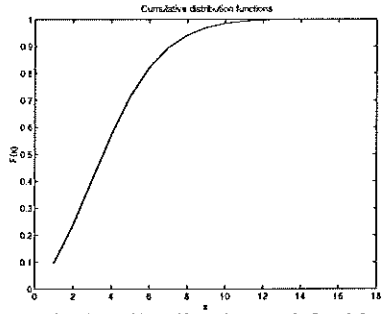

Figure 7: Cumulative distributions of the identified 3-state MMPP with a stochastic matrix (——), data of arrivals (– – –) and the identified 7-state MMPP with a circulant stochastic matrix (– · – ·)

factors are $W_f = 10$ and $W_r = 1$ respectively. The initial values for the algorithm are randomly generated. The cost functions of the two MMPPs converge to 0.0699 and 0.0683 respectively. Fig. 7 and 8 show the cumulative distribution functions and the autocorrelation functions of the two identified MMPP and the data.

In fact, we tested that any MMPP with a circulant statistic matrix with states less than 7 will have the optimal cost value larger than 0.0699 which is the cost value of the 3-state MMPP with a stochastic matrix. This results show that to get the same cost level, less states of a MMPP with a stochastic matrix is needed than the states of a MMPP with a circulant stochastic matrix. In both cases, it takes less than 2 minus to finish all the calculations in SUN ULTRA-1 station. For comparison, in [4], a 100-state MMPP with a circulant statistic matrix was used, while here only a 7-state MMPP of the same type is used, or even 3-state MMPP with a stochastic matrix is used.

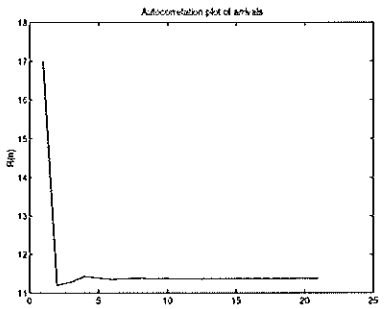

Figure 8: Autocorrelations of the identified 3-state MMPP with a stochastic matrix (——), data of arrivals (– – –) and the identified 7-state MMPP with a circulant stochastic matrix (– · – ·)

8 Conclusions

In this research work, we cast the traffic identification problem for multimedia network (ATM e.g.) as an optimization problem and solve it by general optimization algorithms. Three types of MMPPs are discussed, namely MMPPs with a stochastic matrix, a doubly stochastic matrix and a circulant stochastic matrix. Numerical examples show that the method works well. To get more practical algorithms, further research work is needed, in fact, it is already undergoing, to find more fast algorithms by use of the gradient information of the cost functions. Another point which is worth to explore is to use the so-called orthostochastic matrices [3] instead of doubly stochastic matrices. The orthostochastic matrices have no combinatorial exploration problem for parameterization and have, indeed, very nice structure.

References

[1] Brian D.O. Anderson, Manfred Deistler, and Lorenzo Farina. Nonnegative realization of a linear systems with nonnegative impulse response. Private communication, 1995.

[2] Andrew Grace. *Optimization Toolbox for Use with MATLAB.* The Math Works, Natick, Mass. 01760, 1992.

[3] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, New York, 1985.

[4] Chia Lin Hwang and San Qi Li. On the convergence of traffic measurement and queueing analysis: a statistical-match queueing (smaq) tool. *IEEE/ATM Trans. Networking,* 1(6):522–533, 1993.

[5] T. Kailath. *Linear Systems.* Prentice-Hall, 1980.

[6] San Qi Li and Chia Lin Hwang. Queue response to input correlation functions: continuous spectral analysis. *IEEE/ATM Trans. Networking,* 1(6):678–692, 1993.

[7] San Qi Li and Chia Lin Hwang. Queue response to input correlation functions: discrete spectral analysis. *IEEE/ATM Trans. Networking,* 1(6):522–533, 1993.

[8] Piet Van Mieghem. The mapping of traffic parameters into a mmp(n). Technical report, Aicatel Corporate Reserch Center, Anterwerp, Belgium, 1995.

[9] Bart De Moor. Identification of markov modulated poission processes. Technical report, SISTA, Elec. Eng. Dept., K.U. Leuven, Belgium, 1995.

[10] Peter Van Overschee. *Subspace Identification: Theory - Implementation - Application.* PhD thesis, Katholieke Universiteit Leuven, Leuven, B3001, Belgium, February 1993.

[11] Peter Van Overschee and Bart De Moor. Subspace algorithms for the stochastic identification problem. *Automatica,* 29(3):649–660, 1993.