INFORMATION MEASURE BASED STOCHASTIC SYSTEM IDENTIFICATION OF ATM NETWORK TRAFFIC

Baibing Li and *Bart De Moor*

ESAT-SISTA, Dept. of Electrical Engineering
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, 3001 Leuven, Belgium

ABSTRACT

For ATM network traffic, a new approach based on the Kullback-Leibler information measure is proposed for stochastic system identification of packet traffic. This approach, equivalent to the maximum marginal likelihood estimate, can overcome the over-modeling problem in [1] such that much more parsimonious model order N can be obtained, and then can lead significant reduction in the latter queueing analysis involving in $O(N^3)$ computational complexity. A practical case study is provided for a set of Internet traffic data.

1. INTRODUCTION

The purpose of this paper is to propose a new stochastic identification approach based on the Kullback-Leibler information measure for ATM network traffic.

When modeling network traffic in the classical queueing theories, packet arrivals are often assumed to be Poisson processes. A number of studies have shown, however, that for both local-area and wide-area network traffic, the distribution of packet inter-arrivals clearly differs from exponential (see, for instance, [6]). One of the major reasons for this problem is that a process of ATM network traffic is highly correlated for high speed networks rather than independent (uncorrelated) in classical Poisson processes.

To overcome the difficulty of the classical queueing analysis, Markov chain has been proposed as a statistical model to fit the correlation nature of the input process [3][4]. In such models, a typical ATM network traffic stream is modeled by Markov modulated Poisson process (MMPP), where the underlying Markov chain is used to reflect the time correlation of the input process.

Li & Hwang [3][4][5] investigated identifications of the MMPP models based on frequency domain. Yi and De Moor [7] reconsidered the problem in time domain and converted the identification problem into a nonlinear optimization problem. Recently, noticing that large computational efforts were involved in Yi's approach, De Cock & De Moor [1] decomposed the nonlinear optimization problem into two sub-problems such that the identification of model order and the identification of Markov transition matrix can be made separately. This can significantly simplify the large scale optimization problems.

However, De Cock's approach will cause serious over-modeling problems, i.e., an over-estimated model order N. For the identification stage, to estimate $O(N^2)$ parameters in Markov transition matrix, $O(N^3)$ operations are needed per function evaluation in the optimization algorithm [1]. For the latter queueing analysis stage, the computational complexity of the existing queueing analysis techniques is at least $O(N^3)$ [5]. Since the complexity of many these latter analyses closely depends on model order N, it is very important to obtain an appropriate model order for improving computational efficiency.

In this paper, the identification problem of the MMPP models is re-formulated based on the Kullback-Leibler information measure. Then an information measure based identification of ATM network traffic is proposed to overcome the over-modeling problem in De Cock's approach.

2. PROBLEM FORMULATION

The packet traffic analysis of the arrival process in one node of the network can normally be partitioned into different stages: traffic measurement, identification, queueing analysis and connection admission control. In this paper, the identification problem will be concentrated on.

An MMPP model of order N consists of an N-state Markov chain in which each state i (i=1,...,N) represents a Poisson process with rate λ_i, that is, an MMPP is a Poisson process for which the rate is modulated according to a Markov chain.

Let $P=[p_{ij}]_{N\times N}$ denote the markov transition probabilities. Under the steady situation, consider its eigenvector

$$q^T=q^TP$$
$$\sum_{i=1}^{N} q_i = 1 \ \text{ and } q_i\geq 0, \qquad j=1,\dots,N.$$

Denote two vectors $\lambda=[\lambda_1,\dots,\lambda_N]^T$ and $q=[q_1,\dots,q_N]^T$

For the MMPP model, the one dimensional probability mass function and the cumulative distribution function of a stochastic process a_k are given by (see, e.g. [7])

$$Pr\{a_k=x\} = f(x;\lambda,q)= \sum_{j=1}^{N} q_j P(x,\lambda_j)$$

and

$$Pr\{a_k\leq x\}=F(x,\lambda,q)= \sum_{j=1}^{N} q_j G(x,\lambda_j)$$

where $P(x;\lambda_j)$ and $G(x,\lambda_j)$ are the probability mass function and the cumulative distribution function of Poisson with parameter λ_j. Hence, for the MMPP model, the one-dimensional probability function of a stochastic process a_k is actually a mixture of several Poisson distributions with q_j as its weight.

The autocorrelation function $R(n;\lambda,q,P)=E\{a_k \ a_{k+n}\}$ of the stochastic process a_k is given by (see [7] for the details)

$$R(n;\lambda,q,P)=q^T\Lambda P^n\Lambda 1 \qquad n=1,2,\dots$$

where $\Lambda = \mathrm{diag}(\lambda_1,...,\lambda_N)$ and $1=[1,...,1]^T$.

Finally, the problem of system identification for the MMPP model is to determine an appropriate MMPP model order N, and to estimate parameters P, q and λ from arrival observation a_k (k=0,1,...).

3. INFORMATION MEASURE BASED IDENTIFICATION

3.1 A Brief Summary of Previous Work

According to Li and Hwang [3][4], only the first and the second statistics of arrivals a_k have a significant impact on queueing performance. Noticing these features, Yi & De Moor [7] converted the identification problem of the MMPP model, by matching the first two order statistics in time domain, into following nonlinear optimization:

$$\min_{P,\lambda,q} W_F\left\|F(x;\lambda,q)-\hat{F}(x)\right\|_2 + W_R\left\|R(n;\lambda,q,P)-\hat{R}(n)\right\|_2$$

where $\hat{F}$ is an estimate of the cumulative distribution function, $\hat{R}(n)$ is an estimate of the autocorrelation function $R(n;\lambda,q,P)$ of the stochastic process a_k.

One of the problems in the above approach is that users must a prior provide suitable weights W_F and W_R . The choice of weights are quite important for the estimates of parameters. However, it is not easy to determine in practice. Another problem is that the identification of model N must be made together with the estimation of $O(N^2)$ parameters, which leads the computation quite laborious.

De Cock & De Moor [1] considered a separation strategy for the identification of the MMPP model in time domain. The above optimization problem is re-considered as two separate sub-problems:

$$\min_{\lambda,q} \|\hat{F}(x)-F(x,\lambda,q)\|_2 \tag{1a}$$

$$\min_{P} \|\hat{R}(n)-R(n;\lambda,q,P)\|_2 \tag{1b}$$

In this way, the weights W_F and W_R are no longer needed. Moreover, the model order N is determined in Eq. (1a) which involves in only $O(N)$ parameters, λ and q. Hence, the complexity of the problem is largely reduced.

For a set of observations $X=\{x_k \mid k=1,...,K\}$, with its values in the set $\{0,1,...,M\}$, where $M=Max\{x_k \mid k=1,...,K\}$, it should be noted that the possible values of the parameters $\lambda_j \in (0, M]$. To solve the sub-problem (1a), De Cock & De Moor discretized the interval $(0, M]$ with a step h and let $\tilde{\lambda}_j=\tilde{\lambda}_1+(j-1)h$. Then, for given fixed $\tilde{\lambda}_j$, the sub-problem (1a) becomes

$$\min_{q} \|\hat{F}-Aq\|_2 \tag{2}$$

$$\text{subject to } q_i \geq 0, \quad j=1,...,N.$$

where $A=[A_{ij}]_{M\times N}$, $A_{ij}=G(x_i, \tilde{\lambda}_j)$, $y=[y_1,...,y_M]^T$, $y_i=F(x_i, \tilde{\lambda})$

After solving the optimization problem (2) and obtaining the optimal solution q^*, the model order is then determined as the number of non-zero components of the optimal solution q^*.

Obviously, to accurately approximate the unknown parameters λ_j, the step h must be taken as sufficiently small. This may produce a large scale problem for (2) and cause a serious over-modeling which can cause difficulty in the estimation of $P=[p_{ij}]_{N\times N}$ as mentioned before.

3.2 A New Identification Approach

Consider a stochastic process a_k having the MMPP model. Under steady situation, a_k has a probability mass function

$$f(x;\lambda,q)=\Pr\{a_k=x\}=\sum_{j=1}^{N} q_j P(x,\lambda_j)$$

For a set of observations $X=\{x_k \mid k=1,...,K\}$, based on the Kullback-Leibler information measure, the criterion for identification can be chosen as minimizing the distance between the observed probability mass function o_k and the theoretical one $f(x_k;\lambda,q)$ (k=1,...,K), i.e.

$$\min_{\lambda,q} d_{KL}(o_j, f(x_k;\lambda,q)) = \min_{\lambda,q} \sum_{k=1}^{K} o_j \log[o_k / f(x_k;\lambda,q)]$$

By some algebra, the problem can be rewritten as

$$\max_{\lambda,q} (1/K) \sum_{k=1}^{K} \log[f(x_k;\lambda,q)] + \text{constant} \tag{3}$$

$$\text{subject to } \sum_{i=1}^{N} q_i = 1 \text{ and } q_i \geq 0,\ \lambda_i \geq 0,\ j=1,...,N.$$

Therefore, by solving the above problem (3), the model order N, and the parameters λ_j and q_j (j=1,...,N) can be determined. Then, using the approach by Yi & De Moor [7] and De Cock & De Moor [1], the Markov probability transition matrix $P=[p_{ij}]_{N\times N}$ can be estimated by solving problem (1b).

REMARKS: (i) Choice of distance measure is quite important in many situations. In this paper, the Kullback-Leobler information measure is adopted. The information measure and its related maximun entropy criterion are widely used in many science and engineering fields, for instance, in the estimation of car traffic on road networks [8]. Parameter estimation based on them are usually related to the principle of most informative [2].

(ii) As the Eq. (3) indicating, one of the advantages of choosing the Kullback-Leibler information measure is that its close relation with likelihood function. Obviously, from (3), minimizing the Kullback-Leibler information distance is equivalent to maximizing a marginal likelihood function.

(iii) From (3), the identification for the MMPP model can be implemented as two steps: first, to determine model order N and $2N$-1 parameters λ_j and q_j (j=1,...,N). Then based on (1b) to estimate $N(N$-1) parameters in the Markov probability transition matrix $P=[p_{ij}]_{N\times N}$ In this paper, only the first step identification is considered. See[1] and [7] for details of estimation of transition matrix P.

To convert the constrained nonlinear optimization problem (3) into an unconstrained one, following transforms are introduced:

For parameters q, let

$$q_j=u_j^{\,2}/\,[1+\sum_{j=1}^{N-1}u_j^2\,] \quad \text{and} \quad q_N=1\,/\,[1+\sum_{j=1}^{N-1}u_j^2\,]\ (j=1,\ldots,N\text{-}1)$$

For parameter λ, let

$$\lambda_j=v_j^{\,2} \qquad\qquad (j=1,\ldots,N)$$

Let $z=[u^T, v^T]^T$. Then, the problem (3) can be re-written as

$$\underset{z}{Min}\ J(z;N) = \underset{z}{Min}\ (1/K)\ \sum_{k=1}^{K}\log[h(x_k\,;z)\,] \qquad (3)'$$

where $h(x;\, z)=f(x;\, \lambda(v),\, q(u))$.

Particularly, the first order condition gives

$$\lambda_i = \sum_{k=1}^{K}\pi_{ik}x_k$$

where

$$\pi_{ik} = \vartheta_{ik}\,/\sum_{k=1}^{K}\vartheta_{ik} \ \text{ and } \ \vartheta_{ik} = q_i P(x_k,\lambda_i)\,/\sum_{j=1}^{N}q_j P(x_k,\lambda_j)$$

π_{ik} and ϑ_{ik} satisfy normalization conditions

$$\sum_{k=1}^{K}\pi_{ik} = 1\,, \qquad \pi_{ik}\geq 0 \quad \text{for } i=1,\ldots,N \text{ and } k=1,\ldots,K$$

$$\sum_{i=1}^{N}\vartheta_{ik} = 1\,, \qquad \vartheta_{ik}\geq 0 \quad \text{for } i=1,\ldots,N \text{ and } k=1,\ldots,K$$

Hence, the estimate of λ_i is a weighed average of the observations x_k with weight π_{ik} which is a natural extension of the classical estimate of Poisson distribution $\lambda=\sum_{k=1}^{K}\pi_k x_k$ with the constant weights $\pi_k=1/K$ for all k.

4. A PRACTICAL EXAMPLE

The data, termed as lbl-pkt-4, are observations of one hour of Internet traffic between the Lawrence Berkeley laboratory and the rest of the world, made by Paxson [6].

To avoid too many zeros in the data sequence, the raw data are binned into 1-second bins (which is termed as "pktdata") for analysis according to the standard method [6]. The histograms of the "pktdata" are given by Fig. 1.

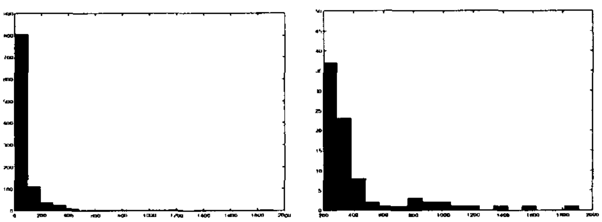

Figure 1. The histograms for all data of the pktdata (left) and for those data greater than 200 (right)

It can be seen from Fig. 1 that the histogram has a relative heavy tail which means that there are noticeable probabilities for the random variable taking larger values which can not be neglected. Thus, most of distributions with light tails are not appropriate for the data.

4.1 The Fitness by the Proposed Method

To illustrate the determination of model order N, the initial value of N is taken as 1. Then N is increased gradually. Fig. 2 gives the performance characterized by $J(z(N)^*;N)$, where $z(N)^*$ is the optimal solution of the optimization problem (3)′.

It can seen from the Fig. 2 that as the model order N increases from 1, the fitness becomes better and better. Particularly, at first, the improvement is significant, while after N reaches around 14, the improvement becomes very limited. This suggests that the model order should be taken as a value around 14. Hence, the model order N is taken as 14 in this case study.

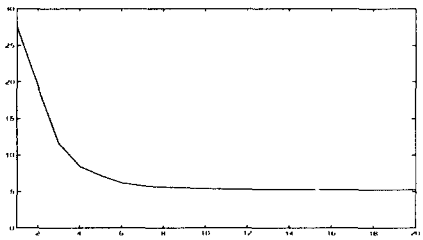

Figure 2. The cost function against model order

Fig. 3 (left) gives the plot of the observed and fitted frequencies for the MMPP model with N=14, and their difference by using the information measure based identification proposed in this paper. It can be seen that the data set "pktdata" is fitted quite well.

In practice, the step of increment for N can be taken larger than 1 to speed up the search of model order N. After locating the approximate value of model order N, further search can be made around that approximate value of N.

4.2 A Comparison with the Previous Method

In this sub-section, a comparison is given for the results obtained by using the approach proposed in this paper and by using De Cock's approach.

According to [1], a partition for the parameter λ is firstly needed. For the data set "datapkt", $M=Max\{x_k \mid k=1,\ldots,K\}=1910$ and K=1000. Take h=10. Then λ is discretized on the interval $(0,\, M]$ with a step h such that $\tilde{\lambda}_j = \tilde{\lambda}_1 + (j-1)h$ for $j=1,\ldots,N$, where N=191. Then, from (2), the problem is

$$\underset{q}{Min}\ \|\hat{F} - Aq\|_2$$

$$\text{subject to } q_j\geq 0, \qquad j=1,\ldots,N.$$

with $A=[A_{ij}]_{1910\times 191}$, $A_{ij}= G(x_i,\, \tilde{\lambda}_j)$, $y=[y_1,\ldots,y_{1910}]^T$, $y_i=F(x_i,\, \tilde{\lambda})$

The above problem is solved and the optimal solution is denoted as q^*. Take all of the elements of q^* which is not less than 0.0001 (other elements are regarded as "zero") to constitute a new vector q_{new}^*. The corresponding elements of λ is denoted as λ_{new}. For the data set "datapkt", q_{new}^* is a vector with dimension 37. This means that the estimated model order $N^*=37$. The estimated parameter vector λ is λ_{new}.

The fitted frequencies by using De Cock's approach and its difference with observed frequencies are given by Fig. 3 (right). It can be seen that the fitness is satisfactory. Of course, to approximate the parameters λ with more sufficient precision, a much smaller step h is needed, say h=0.5 or 0.1. However, this will lead a very large scale for the optimization problem (2), and the estimated model order N^* will increase to some extent.

It should be noted that, however, even if the current estimate of the parameter vector λ is accepted in precision, the resulted increment in model order is significant compared with the estimate of model order N=14 given in Section 4.1. For latter analysis which involves in the complexity of $O(N^3)$, the magnitude in computation will be reduced from 50,000 (37^3) to 2,700 (14^3) if the new approach proposed in this paper is adopted. This strongly suggests that it should be important to overcome over-modeling problem in such kind of studies.

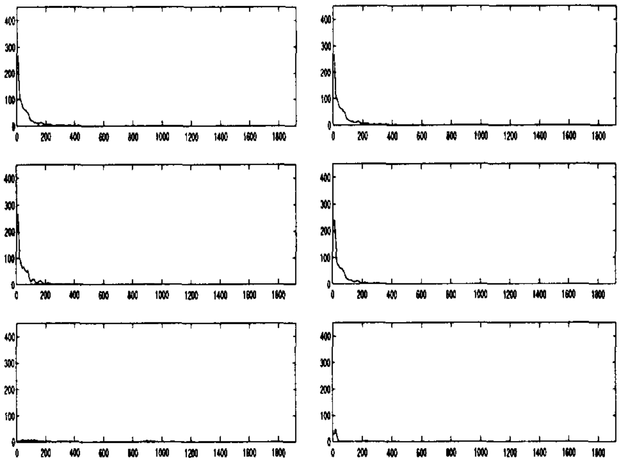

Figure 3. The comparisons of fitness. Top: the observed frequencies; Left middle: fitted frequencies for the MMPP(N=14) model identified by the approach proposed in this paper, Left bottom: the difference with the observed frequencies; Right middle: fitted frequencies for the MMPP(N=37) model identified by De Cock's approach; Right bottom: the difference with the observed frequencies.

5. CONCLUSIONS

A new information measure based approach for stochastic system identification of ATM network traffic is proposed in this paper. This study is concluded as follows:

First, similar to the method presented by [1], this new approach possesses an advantage over [7] due to its decomposed feature. The explorations for model orders are no longer based on overall models with $O(N^2)$ parameters but on a reduced one with only $O(N)$ parameters. Computation for such explorations is then substantially reduced.

Second, much more parsimonious models can be obtained by this new approach to overcome the over-modeling problem in the De Cock's approach. This means that not only much computation can be reduced during the successive identification stage of Markov transition matrix with $O(N^2)$ parameters, but considerable efforts in computation can be reduced in the latter stage of queueing analysis for which computational complexity is at least $O(N^3)$ [5].

ACKNOWLEDGEMENTS

The second author is a senior Research Associate with the FWO-Flanders. This work was supported by Concerted Research Action GOA-MIPS, the FWO project G.0256.97: Numerical Algorithms for Subspace System Identification, the FWO Research Communities: ICCOS and Advanced Numerical Methods for Mathematical Modeling, and the Belgian Government: Interuniversity Attraction Poles (IUAP P4-02 and P4-24).

6. REFERENCES

[1] De Cock, K. and B. De Moor. (1998) "Stochastic system identification for ATM network traffic models: a time domain approach," *Proceedings of EUSIPCO-98*. (available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/decock/reports/98-22.ps.gz*).

[2] Kullback S. *Information Theory and Statistics*, John Wiley & Sons, 1959.

[3] Li S.-Q. and Hwang C.-L "Queueing response to input correlation functions: discrete spectral analysis". *IEEE/ACM Trans. on Networking*, 1: 522-533, 1993.

[4] Li S.-Q. and Hwang C.-L "Queueing response to input correlation functions: continuous spectral analysis". *IEEE/ACM Trans. on Networking*, 1: 678-692, 1993.

[5] Li S.-Q. and Hwang C.-L "On the convergence of traffic measurement and queueing analysis: a statistical-matching and queueing (SMAQ) tool". *IEEE/ACM Trans. on Networking*, 5:95-110, 1997.

[6] Paxson, V. and Floyd S. "Wide area traffic: the failure of Poisson modeling". *IEEE/ACM Trans. on Networking*, 3: 226-244, 1995.

[7] Yi, C. and De Moor B "Traffic identification of ATM networks with optimization algorithms," *Proceedings of the 35th IEEE Conference on Decision and Control*, Dec,1996, pages 277-282.

[8] Van Zuylen H. J. and Willumsen L. G. "The most likely trip matrix estimated from traffic counts". *Transportation Research*, 14B: 281-293, 1980.