MODELING OF BURSTY TRAFFIC USING HETEROGENEOUS ON-OFF SOURCE MODEL

Rajab Faraj

Computer Engineering Department, Faculty of Engineering,
Al-Fatah University
E-mail: frajab@lttnet.net

الملخص

حتى فترة قريبة لم يكن من الواضح إمكانية استخدام نماذج ماركوف لتمثيل الحركة التي على شكل سلسلة طلقات. لقد زعم أن العدد الكبير من الحالات اللازمة لتمثيل الحركة جعلت من نماذج ماركوف غير قابلة للاستعمال لكل الأغراض العملية.لهذا تم البدء في البحث عن نماذج أخرى تكون أكثر ملائمة لتمثيل الحركة التي على شكل سلسلة طلقات مثل فاركشونل قاوسين نويز، فاركشونل براونيون موشن وغيره. هذه النماذج لايصلح استخدامها في تحليل ظاهرة الاصطفاف ولكن يمكن استخدامها لغرض المحاكاة. إن أكثر النماذج استخداما لتمثيل الصوت هو نموذج أون- أوف. حيث تم استخدامه لتمثيل الحركة المرئية على أساس عدة مصادر صغيرة مثل ما تم في نموذج ماقلاريس. انيك ومترا وسوندي استخدموا مصادر أون أوف في تحليل الحركة التي على شكل سلسلة طلقات. يكون نموذج أون- أوف جذاب من الناحية التحليلية عندما يكون الانتقال من حالة أون إلى حالة أوف والعكس على شكل أسي. في هذه الورقة سوف نستخدم عدة درجات غير متجانسة من مصادر أون- أوف لتمثيل الحركة المرئية. النموذج مبني على أساس مطابقة الانحراف المزدوج للمصادر غير المتجانسة مع الحركة الحقيقية. الانحراف المزدوج للمصادر غير المتجانسة يتكون من مجموعة مختلفة من الدوال الآسية بينما في حالة التجانس يتكون من دالة آسية واحدة. النموذج المستخدم له جاذبية عالية، وذلك لأنه باستخدام عدد قليل من المصادر يمكن الحصول على نتائج مرضية للانحراف المزدوج ومعامل التشتت. كما أن العدد القليل من العوامل المتغيرة المستخدمة في التمثيل لإيجاد الانحراف المزدوج تجعل منه نموذجا بسيطا. يختلف النموذج الذي تم تطويره في هذا البحث عن نموذج اندرسون حيث يستخدم خوارزم فالدمان لتبسيط دالة الانحراف المزدوج بمجموعة من الدوال الآسية. وبذلك يكون النموذج الحالي أكثر بساطة من نموذج اندرسون. كما أن نتائج مطابقة الانحراف المزدوج للحركة الحقيقية مع الحركة التي تم توليدها باستخدام النموذج المطور جيدة.

ABSTRACT

Until recently it has not been clear whether Markov based models could be used to model bursty traffic. It has been claimed that the large number of states needed to model

the traffic makes Markov models inapplicable for all practical purposes. This has initiated the search for other models that might be more suitable for modeling bursty traffic such as Fractional Gaussian Noise (FGN), Fractional Brownian Motion (FBM), Fractional Autoregressive Integrating Moving Average (F-ARIMA). For these models, however, the analytical tools for analyzing queuing behavior do not exist. However, they may be used in simulation.

The ON-OFF source model is the most popular model for voice. It was used to model video traffic based on the minsources approach by Maglaris. Anick, Mitra and Sondhi used the ON-OFF sources to analyze bursty traffic. The ON-OFF source model is tractable for analysis when the transitions from the ON state to OFF state and from OFF state to ON state are exponentially distributed.

In this paper, we will use classes of heterogeneous ON-OFF sources to model video data. This model is based on matching the total covariance of the heterogeneous sources to the real data. The covariance of the heterogeneous sources is composed of different exponential functions, while in the homogenous case it is just one exponential. The model is very attractive, because as we will see for a small number of ON-OFF sources it is possible to get good results for the covariance and Index of Dispersion for Count (*IDC*). Moreover, the small number of parameters makes the analysis in finding the covariance and the parameters of the sources simple.

The model we developed is different from that of Andersen, et. al. We used the Feldmann algorithm for approximating a long-tail covariance function by a finite mixture of exponentials. However, Feldman, et. al., used the algorithm to fit probability distribution. Our model is simpler than Anderson's model. The matching of the covariance and for the real data to the traffic generated using the model is quite good.

KEYWORDS: Markov Chains; Probabilistic Models**;** ON-OFF source model**s;** Bursty traffic

1. THE MATHEMATICAL MODEL

In this section we consider independent classes of ON-OFF sources, let N_i (i= 1, 2, ... m) denote the number of sources in class i to model long-range dependence traffic such as video [1,3 and 11]. Within a class the sources are identical and independent. In this model, for the *ith* class, packets are generated during talk spurts which are the ON state, and no packets are generated during the OFF state. The times spent in the ON and OFF states are exponentially distributed with means $1/\beta_i$ and $1/\alpha_i$, i= 1, 2, ... m, respectively. When the source is in the ON state it generates data at rate of R_i , i= 1, 2, ... m.

The Asynchronous Transfer Mode (ATM) multiplexer consists of a server transmitting cells at a specified line rate and a buffer whose size is determined by the delay constraints on cell transmission. Cells arrive at the multiplexer from a number N_i (i= 1, 2,... m) of sources. See Figure (1).

The basic idea of the 3-class model is that there are three time frames for transitions: short term, medium term and long term, respectively. The transition rates are such that $\alpha_1 >> \alpha_2$ ....>> α_m and $\beta_1 >> \beta_2$ ....>> β_m, where for our model we have m = 3, so that the shorter the time frame, the more rapid the transition. For example, in the case of three level Markov chain the possible transitions are illustrated in Figure (2). From each state there are three possible transitions: from state 1 there is a short term transition that will take us to state 2 given by α_1 a medium term transition to state 3 given by α_2 and finally a long term transition that will take us to state 5 given by α_3. The most likely transition is to state 2. As shown in Figure (2), the three level two state models are complicated.

Using the state transition diagram, the 3 level Markov chain shown in Figure (2) can be mapped into the simple three independent ON-OFF sources model shown in Figure (3). This can be verified by finding out the infinitesimal generator matrix from Figure (2) and that from Figure (3).

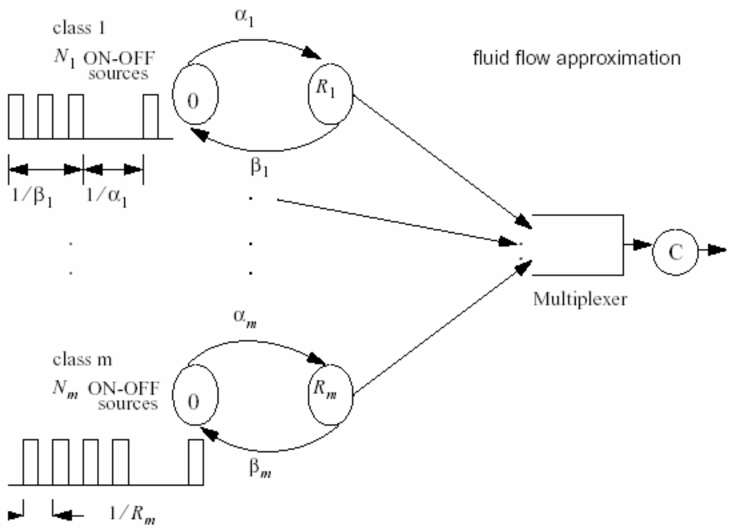

Figure 1: m-class ON-OF source mode

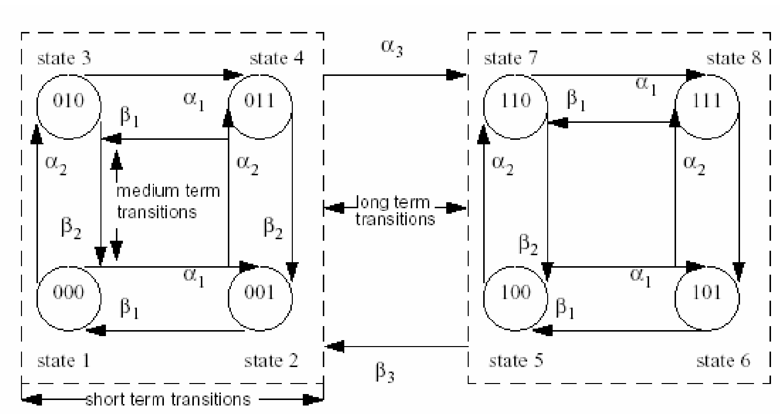

Figure 2: Three level Markov Chain

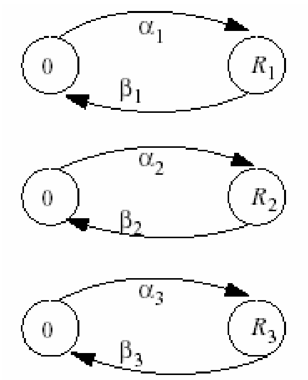

Figure 3: Three independent ON-OFF sources

Let [n_1 n_2 ... n_i...n_m ; u] be the state with n_i source in class i ON and the buffer content does not exceed u and $p_{n_1 n_2 \dots n_i \dots\dots n_m}(u)$ be its equilibrium probability. Packets are served at a rate of C packets per time unit. We utilize the fluid flow approximation [2,15], which has shown much promise in the analysis of ATM networks. Similarly to [2] we have the following:

$$\left(\sum_{i=1}^{m} R_i n_i - C\right)\frac{dp......(u)}{du} = \sum_{i=1}^{m}\left[\alpha_i\left(N_i - n_i + 1\right)\right]P_{....n_i-1..}(u) - \left\{\alpha_i\left(N_i - n_i\right)+\beta_i n_i\right\}P_{.......n_i........}(u) + \beta_i\left(n_i+1\right)P_{....n_i+1..}(u) \qquad (1)$$

We express Equation (1) in the following familiar matrix form,

$$D\frac{dp(u)}{du} = p(u)M \qquad (2)$$

where D is an (N_1+ 1)×....× (N_m+ 1) diagonal matrix, M is (N_1+ 1)×....× (N_m+ 1) infinitesimal generator matrix and $\boldsymbol{p}(u)$ is a vector equal [$p_{00......0}(u)$ ...., $p_{n_1 n_2 \dots\dots\dots n_m}(u)$].

In the next section, we introduce the covariance function of Equation (2), which will be used to derive the parameters that characterize the independent m class ON-OFF sources. That is, the parameters determination is based on second order statistics.

2. MODEL PARAMETER DETERMINATION

Our work is based on finding the total covariance of the independent m classes N_1, N_2, ...,N_m heterogeneous ON-OFF source model. Then by matching to the real data we find out the parameters that characterize the ON-OFF sources by adapting the Feldman algorithm to the fitting of the covariance [4]. We may view this algorithm as analogous to Gram-Schmidth orthogonalization over the time axis. The goal is to approximating a long-tail covariance distribution by a finite mixture of exponentials over shorter time scales. That is, we approximate a non-exponential function with a sum of exponential terms that we can easily deal with. The quality of the approximation is based on goodness of fit of the approximation by comparing the covariance function of the model with that of the data.

The covariance *CO* of the number of packets of a long-tail process as a function of the lag *k* and the Hurst parameter *H* [3, 9 and 10] behaves asymptotically as:

$$CO(k) \sim k^{2H-2} \tag{3}$$

The covariance function given by Equation (3) decays hyperbolically (obeying some power law) as the lag k increases rather than exponentially, where *k* is the lag and *H* is the Hurst parameter.

The covariance $COV(\tau)$ of independent *m* class ON-OFF sources described by Equation (2) is simply given by:

$$COV(\tau) = \sum_{i=1}^{m} \alpha_i \beta_i N_i \frac{R_i^2}{(\alpha_i + \beta_i)^2} e^{-(\alpha_i + \beta_i)\tau} \tag{4}$$

Applying the additivity property to Equation (4), we find for the mean μ,

$$\mu = \sum_{i=1}^{m} \frac{\alpha_i N_i R_i}{(\alpha_i + \beta_i)} \tag{5}$$

and for the variance *Var*,

$$Var = \sum_{i=1}^{m} \frac{\alpha_i \beta_i N_i R_i^2}{(\alpha_i + \beta_i)^2} \tag{6}$$

Let,

$$\lambda_i = \alpha_i + \beta_i, \quad i=1,2,\ldots\ldots\ldots\ldots m \tag{7}$$

and

$$k_i = \alpha_i \beta_i N_i \frac{R_i^2}{\lambda_i^2}, \quad i=1,2,\ldots\ldots\ldots\ldots m \tag{8}$$

Substitute Equation (7) and Equation (8) into Equation (4) and assume that frames are generated at rate of *f* frames /sec:

$$COV(\tau) = k_1 \mathrm{e}^{\frac{-\lambda_1 \tau}{f}} + k_2 \mathrm{e}^{\frac{-\lambda_2 \tau}{f}} + \ldots\ldots + k_m \mathrm{e}^{\frac{-\lambda_m \tau}{f}} = \sum_{i=1}^{m} k_i \mathrm{e}^{\frac{-\lambda_i \tau}{f}} \tag{9}$$

Equation (9) is a finite mixture of exponentials that approximate the long-tail distribution function given by Equation (3). The idea is to approximate Equation (3) by Equation (9), because performance models with component long-tail distributions tend to be difficult to analyze.

As can be seen from Equation (9) we have 2*m* unknowns and therefore we need 2*m* Equations to find them. Since the covariance is composed of exponential components $\lambda_1, \lambda_2, \ldots, \lambda_m$, and *m* arguments $K_1, K_2, \ldots K_m$, we match at the quantiles: $0 < c1 < c2 < \ldots < cm$, which represent how many classes that we have. For example, for two classes we have two quantiles c_1, c_2 and for three classes we have three quantiles c_1, c_2, c_3 and so on. In order to solve 2*m* Equations to find the 2*m* unknowns, let *b* be a scaling factor such that $1 < b < (c_i + 1)/c_i$ for all i ; e.g., we could have b=4, $c_i = 10^{(i-1)} c_1$ for . $2 \leq i \leq m$. $\frac{c_2}{c_1} = \frac{c_3}{c_2} = ... \frac{c_m}{c_{m-1}}$. See Figure (4) for the three source case.

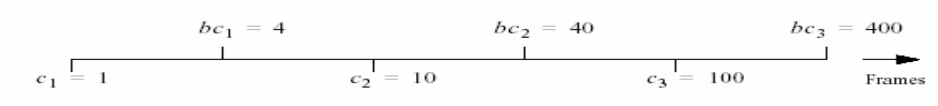

Figure 4: Illustration of how to choose the quantiles c_1, c_2, c_3 and the scaling factor b ($c_1 = 1, c_2 = 10, c_3 = 100, b = 4$)

Given the real data and using the technique in [4] we can obtain the exponential components $\lambda_1, \lambda_2, ..., \lambda_m$ and the arguments $K_1, K_2, ...K_m$ in reverse order by finding first λ_m and K_m and then λ_{m-1} and K_{m-1} so on until we find λ_1 and K_1. From Equation (7) $\lambda_1 >> \lambda_2 >> .... >> \lambda_m$. Therefore, at the quantiles c_m and bc_m, only the terms of the covariance that have argument λ_m count and those terms of covariance that have arguments $\lambda_{m-1}, \lambda_{m-2}, ..., \lambda_1$ are negligibly small. Therefore,

$$COV(c_m) = k_m e^{\frac{-\lambda_m c_m}{f}} \quad (10)$$

and

$$COV(bc_m) = k_m e^{\frac{-\lambda_m bc_m}{f}} \quad (11)$$

From Equations (10) and (11), we find the two unknowns λ_m and K_m.
Now we proceed to find the other two unknowns λ_{m-1} and K_{m-1} at the quantiles c_{m-1} and bc_{m-1}. In this case only the terms of the covariance that have argument λ_{m-1} and λ_m count and the terms of the covariance that have arguments $\lambda_{m-2}, \lambda_{m-3}, ..., \lambda_1$ are negligibly small.

$$COV(c_{m-1}) = k_m e^{\frac{-\lambda_m c_{m-1}}{f}} + k_{m-1} e^{\frac{-\lambda_{m-1} c_{m-1}}{f}} \quad (12)$$

$$COV(bc_{m-1}) = k_m e^{\frac{-\lambda_m bc_{m-1}}{f}} + k_{m-1} e^{\frac{-\lambda_{m-1} bc_{m-1}}{f}} \quad (13)$$

where λ_m and K_m are already known from Equations (10) and (11).
Given λ_m, K_m, λ_{m-1} and K_{m-1} we find the next two unknowns λ_{m-2} and K_{m-2} at the quantiles c_{m-2} and bc_{m-2} ,

$$COV(c_{m-2}) = k_m e^{\frac{-\lambda_m c_{m-2}}{f}} + k_{m-1} e^{\frac{-\lambda_{m-1} c_{m-2}}{f}} + k_{m-2} e^{\frac{-\lambda_{m-2} c_{m-2}}{f}} \quad (14)$$

$$COV(bc_{m-2}) = k_m e^{\frac{-\lambda_m bc_{m-2}}{f}} + k_{m-1} e^{\frac{-\lambda_{m-1} bc_{m-2}}{f}} + k_{m-2} e^{\frac{-\lambda_{m-2} bc_{m-2}}{f}} \quad (15)$$

and so on until we end up with the last two unknowns λ_1 and K_1.

The final step is to find the parameters that characterize the ON-OFF sources i.e, $\alpha_1, \beta_1, R_1; \alpha_2, \beta_2, R_2; .....; \alpha_m, \beta_m, R_m$. We have a system of $3m$ (where m is the number of classes and the factor 3 comes from the fact that each source has 3 parameters to be determined) parameters to be calculated, however we have in hand only *2m* known factors (λ_1, k_1; λ_2, k_2;......; λ_m, k_m). The basic property of our model is given in section 1, where the transitions rates are assumed to be such that $\alpha_1 >> \alpha_2$ …….>> α_m and β_1 >> β_2 ....>> β_m, (already from 7, we have $\lambda_1 >> \lambda_2 >>$ ….>>λ_m.). We use this assumption in such a way that we have fewer unknowns to evaluate

Let $\alpha_i = 10^{-(i-1)}\alpha_1$, i=2,…………..m

From (8) we find Ri in terms of Ki, αi and βi,

$$R_i = \sqrt{\frac{\lambda_1^2 k_i}{\alpha_i \beta_i N_i}} \tag{16}$$

Substitute for R_i, β_i ($\beta_i = \lambda_i - \alpha_i$) in equation (5) we have, after some manipulation, the following,

$$\sqrt{\frac{\alpha_1 N_1 k_1}{(\lambda_1 - \alpha_1)}} + \sqrt{\frac{\alpha_2 N_2 k_2}{(\lambda_2 - \alpha_2)}} + ... + \sqrt{\frac{\alpha_i N_i k_i}{(\lambda_i - \alpha_i)}} + ... + \sqrt{\frac{\alpha_m N_m k_m}{(\lambda_m - \alpha_m)}} = \mu \tag{17}$$

Since $\alpha_i = 10^{-(i-1)}\alpha_1$, equation (17) can be written in the following form:

$$\sqrt{\frac{\alpha_1 N_1 k_1}{(\lambda_1 - \alpha_1)}} + \sqrt{\frac{\alpha_1 N_2 k_2}{(10\lambda_2 - \alpha_1)}} + ... + \sqrt{\frac{\alpha_1 N_i k_i}{(10^{i-1}\lambda_i - \alpha_1)}} + ... + \sqrt{\frac{\alpha_1 N_m k_m}{(10^{m-1}\lambda_m - \alpha_1)}} = \mu \tag{18}$$

The number of sources N_1, N_2, ..,N_m is given in advance. Also, we know the values of λ_1, λ_2, .., λ_m, K_1, K_2, ...K_m from matching to the data, and also we know $\mu$the estimated mean value of the real data. Therefore, the non-linear Equation (18), which is a function of only one unknown parameter α_1, can be solved numerically. Knowing the parameters $\alpha_1\beta_1$, R_1;α_2, β_2, R_2 ;.. ; α_m , B_m , R_m can be obtained very easily using Equations (7) and (8).

3. NUMERICAL RESULTS

The model can be applied to any number of classes and any number of sources per class. However, as the number of sources increases the solution of Equation (18) becomes more difficult. Because of this, we apply the model to the three classes and one source per class. Using the video data available in [7,8], we calculate the covariance function for the real data and apply the procedure presented in section 2 to find the parameters for the heterogeneous ON-OFF source model. These three heterogeneous ON-OFF source models are used to generate video traces in Optimization Network (OPNET) program. We generate almost the same number of real video data frames of approximately 50,000 frames for video-conferencing, video-phone, TV series and Movie video sequences. From the real traffic we find the covariance function, and *IDC* [5] and compare them with that of the generated traffic. Also, as a reference, we calculate covariance, *IDC* based on the Maglaris model with 20 minisources [11].

The values of the parameters depends on the quantiles c_i's, the scaling factor *b* and the number of frames over which the matching is going to be done. As the number of quantiles, the scaling factor and the number of frames over which the matching is going to be done increases the accuracy will be increased. However, this is not always possible since increasing the number of quantiles means increasing the number of classes (for the three class model, three quantiles are needed), which makes the solution of Equation (18) more difficult.

Tables (1-8) show, respectively, the estimated parameters α_i ,β_i ,R_i for two and three class single ON-OFF sources needed to model the video-conferencing, video-phone, TV series, and Movie data are available in [7]. They also show the estimated exponential components λ's and the arguments k_i's. Given the parameters α_i's, β_i's, and R_i's for ON-OFF sources for the Variable Bit Rate (VBR) video traffic traces, a replica of the traffic is generated using OPNET. From the real traffic we find out the covariance function, *IDC* and compare them with that of the generated traffic. We also plot the covariance based on Equation (9).

Table 1: Parameters for the two class single ON-OFF sources matched to the video-conferencing trace over 768 frames with *C1=1, C2=192 and b=4*

Parameter	λ	k	α	β	R
Source1	0.588	5034.2	0.348	0.24	144.43
Source2	0.046	620.78	0.0348	0.011	58.68

Table 2: Parameters for the three class single ON-OFF sources matched to the video-conferencing trace over 768 frames with *C1=1, C2=16, C3=256 and b=3*

Parameter	λ	k	α	β	R
Source1	2.156	401.69	1.79	0.36	53.81
Source2	0.39	4815.4	0.179	0.22	139.38
Source3	0.03	429.1	0.0179	0.016	41.55

Table 3: Parameters for the two class single ON-OFF sources matched to the video-phone trace over 768 frames with *C1=1, C2=192, and b=4*

Parameter	λ	k	α	β	R
Source1	0.94	9346.0	0.494	0.445	193.599
Source2	0.08	2485.7	0.0494	0.026	104.953

Table 4: Parameters for the three class single ON-OFF sources matched to the video-phone trace over 768 frames with *C1=1, C2=16, C3=256 and b=3*

Parameter	λ	k	α	β	R
Source1	2.53	1883.4	1.75	0.779	94.035
Source2	0.42	8883.4	0.175	0.24	190.832
Source3	0.05	1042.6	0.0175	0.029	66.702

Table 5: Parameters for the two class single ON-OFF sources matched to the TV series video trace over 768 frames with *C1=1, C2=192, and b=4*

Parameter	λ	k	α	β	R
Source1	3.033	746110.0	0.305	2.99	2980.9
Source2	0.031	700350.0	0.030	0.0008	5199.0

Table 6: Parameters for the three class single ON-OFF sources matched to the TV series video trace over 768 frames with *C1=1, C2=16, C3=256 and b=3*

Parameter	λ	k	α	β	R
Source1	16.47	461150.0	2.63	13.84	1853.4
Source2	0.408	505950.0	0.263	0.1451	1486.0
Source3	0.027	617570.0	0.0264	0.00097	4233.4

Table 7: Parameters for the two class single ON-OFF sources matched to the Movie video trace over 768 frames with *C1=1, C2=192, and b=4*

Parameter	λ	k	α	β	R
Source1	2.8733	817700.0	0.08099	2.7923	5463.4
Source2	0.083	830860.0	0.008099	0.000201	5937.7

Table 8: Parameters for the three class single ON-OFF sources matched to the Movie video trace over 768 frames with *C1=1, C2=16, C3=256 and b=3*

Parameter	λ	k	α	β	R
Source1	21.53	469380.0	0.60	20.93	4147.2
Source2	0.46	592920.0	0.060	0.40	2282.9
Source3	0.006	779260.0	0.0060	0.00015	5674.0

As a reference, we compare the statistical measures such as covariance, *IDC*, of the real video data and the generated video traffic based on heterogeneous ON-OFF source model with that of the generated traffic based on Maglaris model. To do that we need to find the parameters that characterize the Maglaris model. Using Equations (8), (9),(10) and (11) in the paper of Maglaris [11] we calculated the parameters of the Maglaris model for the four video traces. These are shown in Table (9) below.

Table 9: Maglaris model parameters for VBR traces, video-conferencing, video-phone, TV series, and Movie. Each video source is characterized by 20 minisources.

Parameter	α	β	R
Video-conferencing	0.05	0.3	39
Video-phone	0.04	0.3	75.81
TV series	0.07	0.06	520.41
Movie	0.05	0.05	559.59

3.1 Covariance and *IDC*

In this section we consider the matching of the covariance and the *IDC* of the generated traffic to the real data using heterogeneous sources. As a reference, we have generated traffic based on the Maglaris model of 20 minisources [11]. We compare the accuracy of our model with that based on the Maglaris model. Also, we consider the effect of increasing the number of ON-OFF heterogeneous sources to the accuracy of the matching. Given the exponential components λ_i's and the arguments k_i's of the heterogeneous ON-OFF source model, we see how well the covariance represented by Equation (9) matches the covariance of the real data and that based on the Maglaris model.

The covariances of the real video-conferencing and the generated 3 and 2 class single ON-OFF sources are shown in Figure (5). We also plot the covariance based on Magalris model of 20 minisources. In Figure (6), we plot the covariance of real video-conferencing and that based on the formula given by Equation (9) of our work and that given by Maglaris of one exponential term, which is given by Equation (5) in [11]. We do the same for video-phone data which, are shown in Figure (7) and Figure (8), respectively. As expected, the accuracy increases as the number of classes increases. Moreover, in comparison with the Maglaris model, the matching based on the 3 class single ON-OFF source is shown to be better. For both video teleconferencing data, the match to the traces based on the generated traffic and that using formula (9) is quite good over a large number of lags.

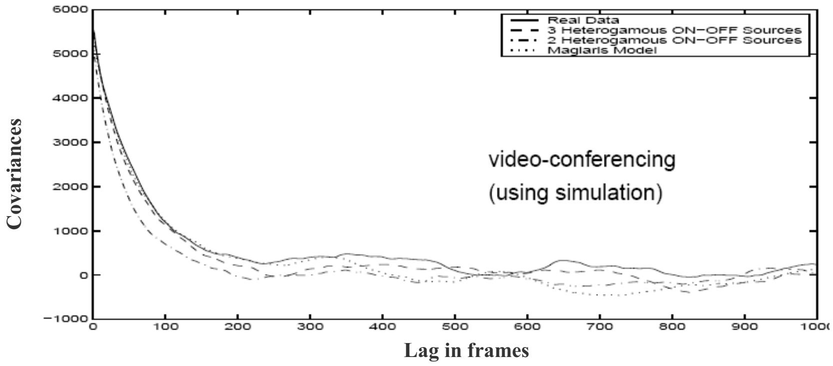

Figure 5: Covariances functions of real video-conferencing data compared with that of Maglaris and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source.

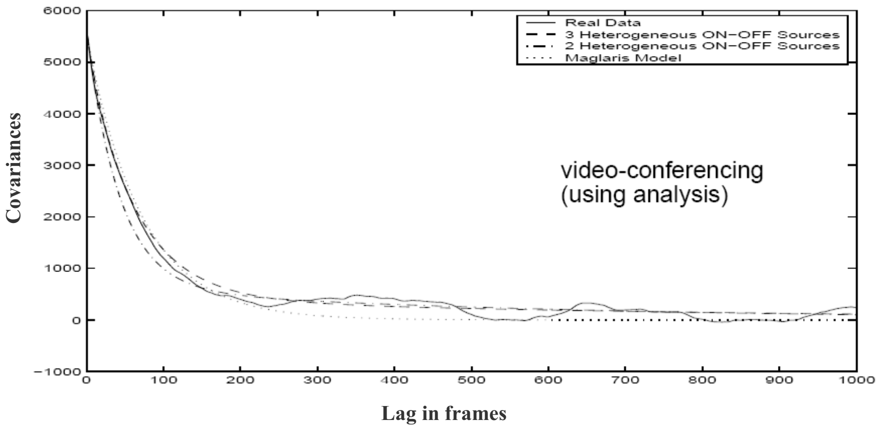

Figure 6: Covariance functions of real video-conferencing data, and using formula (9) for 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source

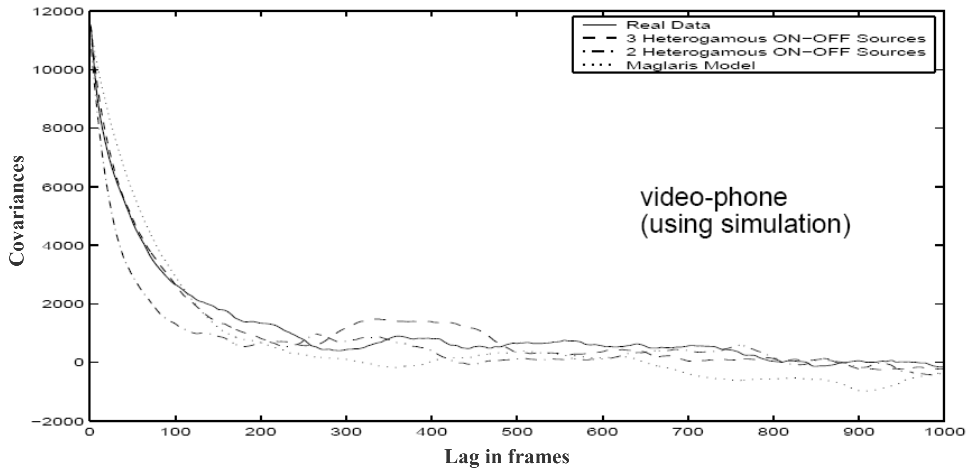

Figure 7: Covariances of real video-phone data compared with that of Maglaris, and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source.

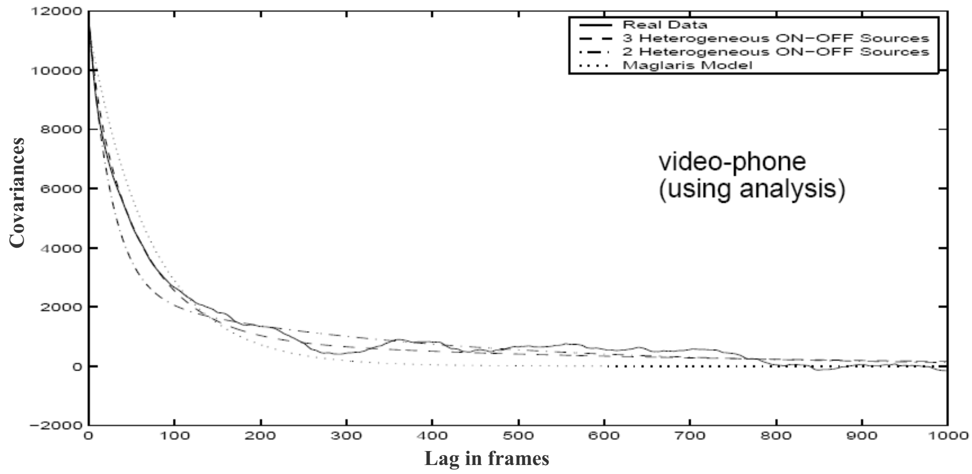

Figure 8: Comparison of the covariances of real video-phone data,and using formula (9) for 3 and 2 class heterogeneous source model,each class has 1 ON-OFF source.

The covariances for the highly correlated traffic real TV series and Movie and that of the 3 and 2 class single ON-OFF sources are shown in Figure (9) and Figure (11), respectively. Also in the figures we plot the covariance based on Maglaris model. The results shows that our model has better matching of the covariances than that based on Maglaris model. However, the matching for this kind of entertainment traffic is less accurate than that for the teleconferencing traffic shown in Figure (5) and Figure (7). This due to the high value of *H*.

In Figure (10) and Figure (12), we plot the covariance of real sequences TV series and Movie, respectively, and that based on the formula given by Equation (9) and that given by Maglaris of one exponential term, which is given by Equation (5) of [11]. The matching for both sequences and that based on formula (9) is reasonable. The matching of the Maglaris covariance given by Equation (5) in [11] and the two real sequences TV series and Movie is reasonable when the lag is not large. As the lag increases the deviation of the covariance of the real data from that of the covariances given by formula (9) and that given by Maglaris of one exponential term given by Equation (5) in [11] becomes clear.

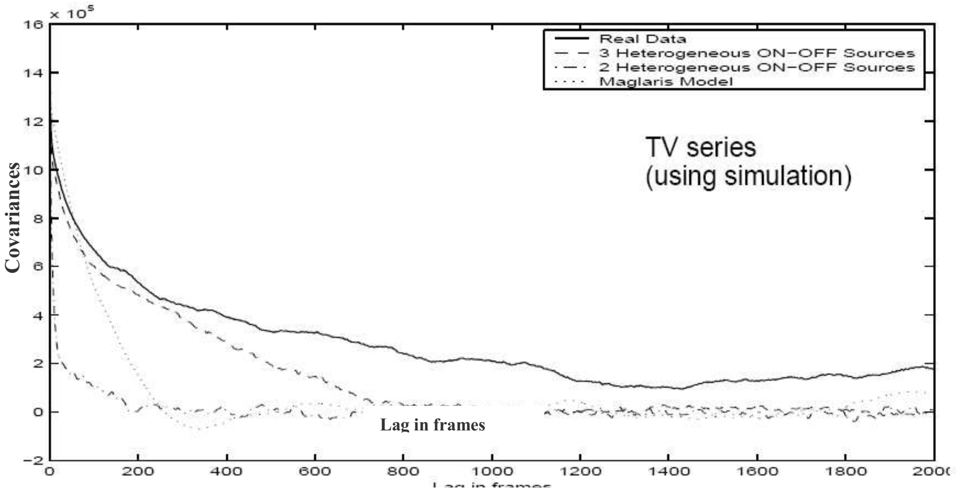

Figure 9: Covariances of real TV series data compared with that of Maglaris, and generated 3 class heterogeneous source model, each class has 1 ON-OFF source

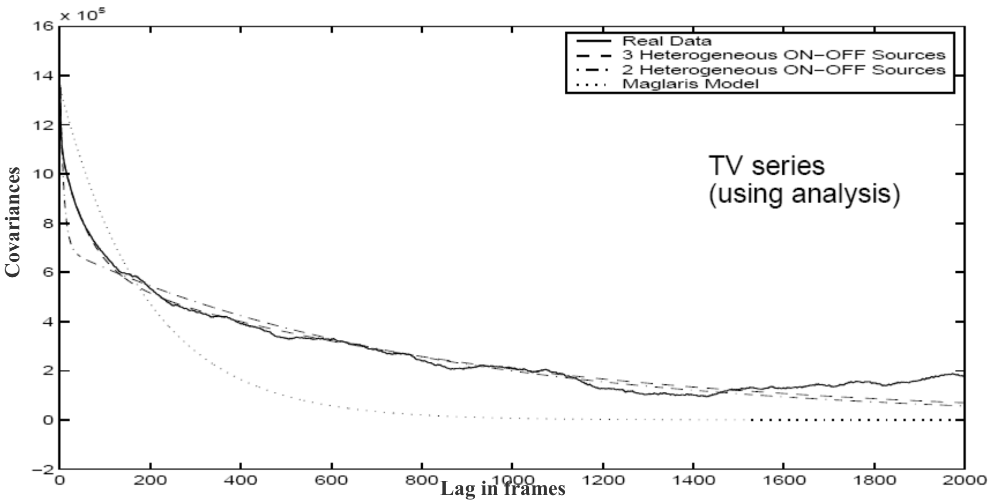

Figure 10: Comparison of the covariances of real TV series data, and using formula (9) for 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source.

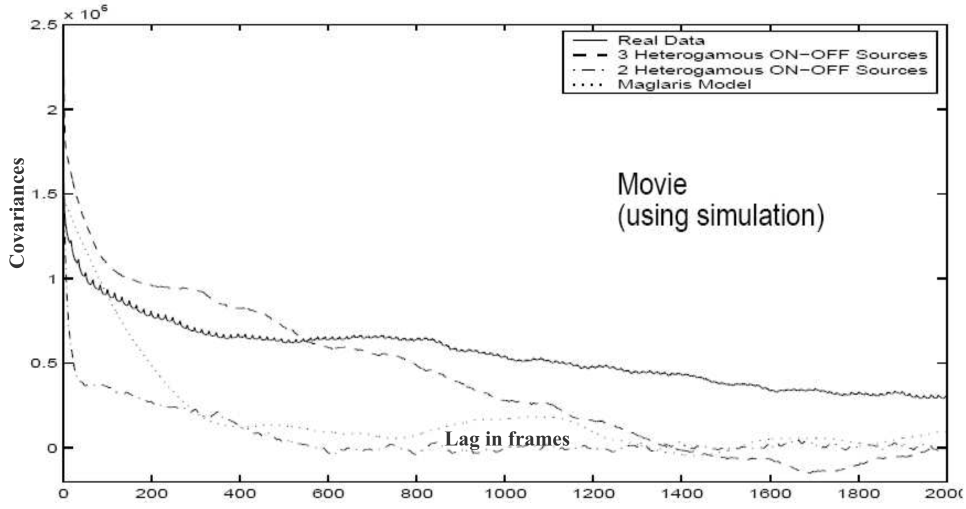

Figure 11: Covariances of real Movie data compared with that of Maglaris, and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source

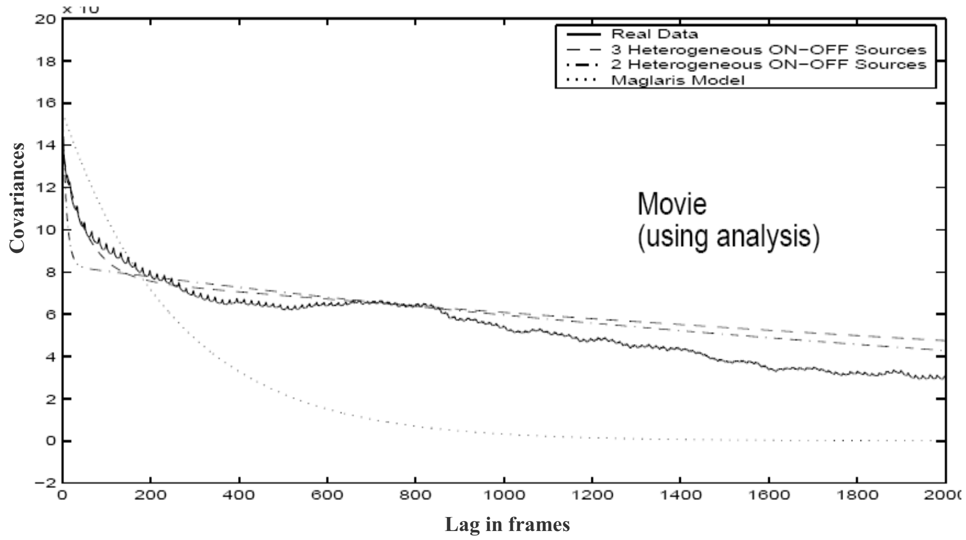

Figure 12: Comparison of the covariances of real Movie data, and using formula (9) for 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source.

The other part of matching the heterogeneous sources to the real data is the *IDC*. As shown in Figure (13) and Figure (14), respectively, the *IDC* of the synthetic video-conferencing and video-phone traffic matches the *IDC* of the real data. As for the covariance, increasing the number of heterogeneous sources from 2 to 3 will increase the accuracy of the matching. We also show the *IDC* based on the Maglaris model.

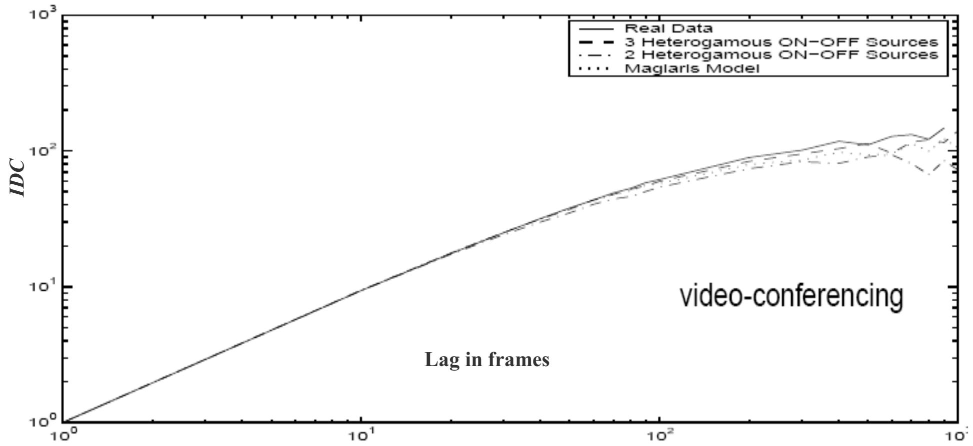

Figure 13: ***IDC*** **of real video-conferencing data compared with that of Maglaris and generated 3 and 2class heterogeneous source model, each class has 1 ON-OFF source.**

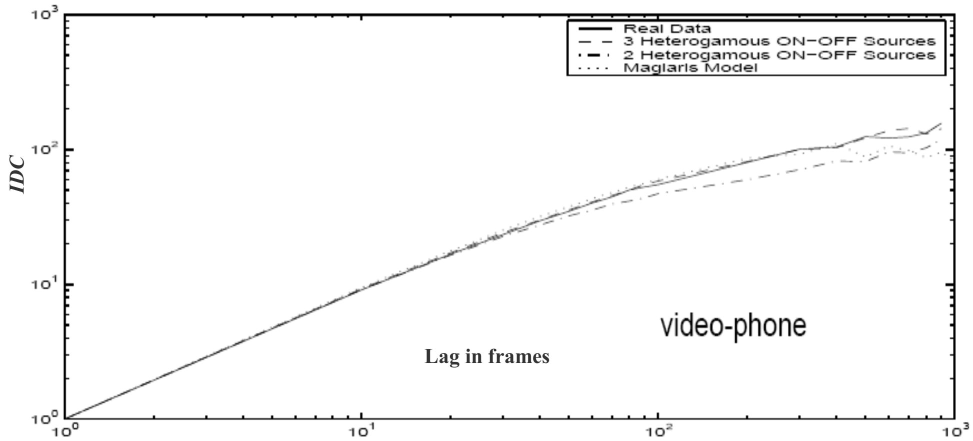

Figure 14: ***IDC*** **of real video-phone data compared with that of Maglaris, and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source**

The *IDC* for the entertainment video data, TV series and Movie, are shown in Figure (15) and Figure (16), respectively. The figures show the results for real data, three heterogeneous ON-OFF source model, two heterogeneous ON-OFF source model and that of Maglaris model. As the number of heterogeneous ON-OFF sources increases, the accuracy of the matching becomes more accurate. Moreover, the matching for the three heterogeneous ON-OFF source model performs better than that of the Maglaris model. The worst performance is of the 2-class heterogeneous source model. The worst model is of the 2-class heterogeneous source model.

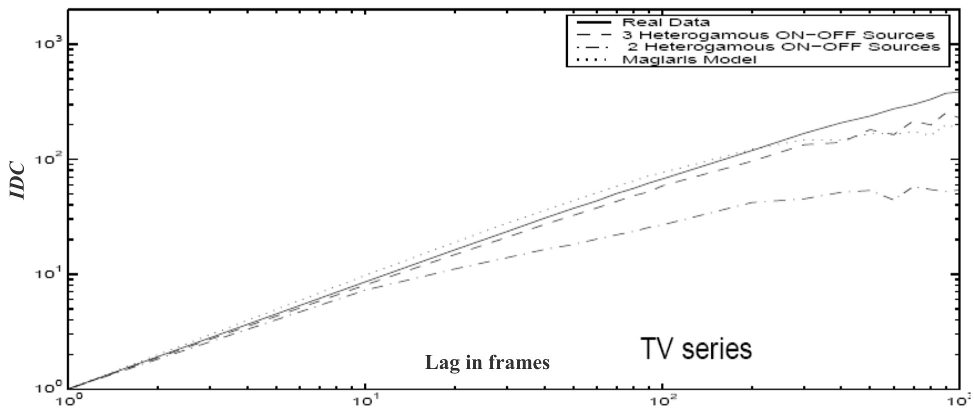

Figure 15: ***IDC*** **of real TV series data compared with that of Maglaris, and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source**

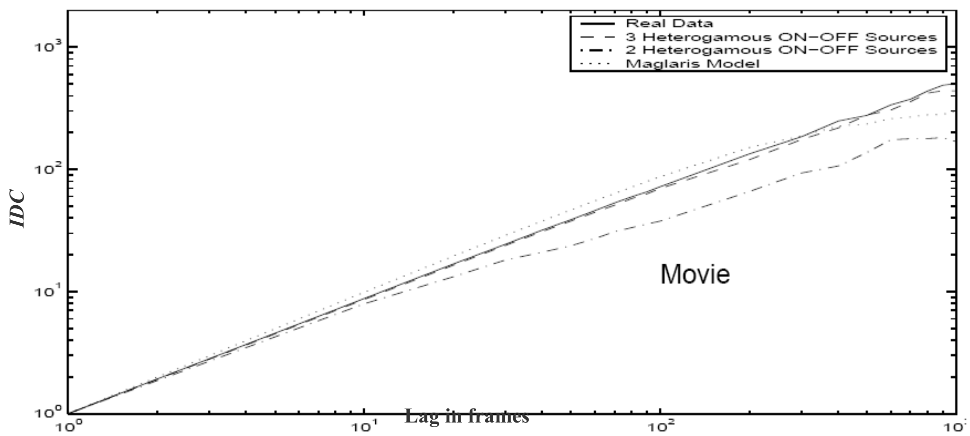

Figure 16: ***IDC*** **of real Movie data compared with that of Maglaris, and generated 3 and 2 class heterogeneous source model, each class has 1 ON-OFF source**

5. CONCLUSION

We have proposed a model for characterizing correlated cell arrival of real bursty video data. Based on a second order statistical analysis, we have used heterogeneous ON-OFF source model to characterize the traffic. The model consists of *m* classes of ON-OFF sources. Although the ON-OFF periods are exponentially distributed and the number of sources is small, we have a good matching for the covariance and the *IDC*. It is clear from the results we obtained, as the number of classes and the number of ON-OFF sources per class increases, the accuracy of the model will be increased especially for highly correlated traffic. However, increasing the number of classes and number of sources per class will result in analytical and computational complexities.

Using just the 3-heterogeneous ON-OFF source model gives good results for matching the traffic characteristics indices and at the same time the analysis is simple.

6. REFERENCES

[1] Andersen A.T. and Nielsen B., “An application of superpositions of two state Markovian sources to the modelling of self-similar behavior,” IEEE INFOCOM97, April1997, Kobe, Japan.

[2] Anick D., Mitra D and Sondhi M. M., “Stochastic theory of a data handling system with multiple sources,” Bell Syest. Tech. J. 61, 1982

[3] Cox D. R., “Long-range dependence: A review”, in statistics: An appraisal, Proceedings 50th Anniversary Conference, Iowa State Statistical Library, H. A. David and H. T. David, editors, Iowa State University Press, pp. 55- 74, 1984.

[4] Feldmann A. and Whitt W., “Fitting mixtures of exponentials to longtail distributions to analyze network performance models,” IEEE INFOCOM’97, April 1997, Kobe, Japan.

[5] Gusella R., “Characterizing the variability of arrivals process with indexes of dispersion,” IEEE JSAC, 9(2),pp. 203-211, Apr. 1991.

[6] Heffes H. and Lucantoni D. M., “A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance”, IEEE JSAC, vol. SAC-4, no 6, pp. 856-868, Sep. 1986.

[7] Heyman D. and Lakshman T. V., “Source models for broadcast video traffic,” IEEE/ACM Trans. Networking, vol. 4, no. 1, Feb 1996

[8] Heyman D. and Lakshman T. V., “What are the implications of long-range dependence for VBR-video traffic Engineering ?,’’ IEEE/ACM Trans. Networking, vol. 4, no. 3, June 1996.

[9] Hosking J., “Modeling persistence in hydrological time series using fractional differencing”, Water Resources Research, 20 (12), 1984.

[10] Hurst H. E., “Long term storage capacity of reservoirs”, Trans. Amer. Soc. Civil Engineers, pp. 770-779, 1951.

[11] Maglaris B., Anastassiou D., Sen P., Karlsson G., Robbins J., “Performance Models of statistical multiplexing in packet video communication,” IEEE Trans.