PACKET TRAFFIC CHARACTERIZATION. ARRIVAL LAWS AND WAITING TIMES

Bernardo FONTANA and Antonio GUERRERO

ALCATEL Standard Eléctrica, S.A. Research Center
Madrid, Spain

Packet traffic flows are analyzed by means of two analytical models based on traffic patterns observed in existing packet-switched computer networks. Bulk packet arrivals, and correlations between consecutive arrivals of individual packets, are considered; criteria for deriving model parameters from traffic measurements are established; and the derivation of waiting times is described. A computer program has been developed in order to provide numerical results from these models. Applications to practical cases are discussed.

1. INTRODUCTION

Accurate predictions of packet arrivals in a communication system or network are essential to efficient resource management. Unfortunately the well-known Poisson process used to derive many analytical queueing models, mainly applying to telephone traffic flows, is not valid for packet traffic flows, as shown by observations of several computer networks [1,2,3]. The analysis of the dispersion factor clearly demonstrates this.

This paper derives arrival law models that better characterize the packet traffic flows that we might find in the ISDN, or the future IBCN probably based on asynchronous time division, where complex mixes of quite different packet services may coexist. To accomplish this task we first look at the statistical behavior of packet flows as determined by measurements.

One of the bases of this paper is the measurement analysis of the interarrival times of packets in a computer network presented in [3], where it is shown that (1) the overall coefficient of variation of these interarrival times is far from unity, (2) there are several types of interarrival times and (3) the interarrival times are correlated. To explain these facts, a packet-train model has been developed [3]. This model considers that packets flowing between two nodes are grouped in 'trains'. A train is formed of cars (packets), with the intercar time shorter than a specified maximum allowed intercar gap (MAIG); if no packet is seen during an MAIG, the previous train is assumed to be terminated and the next packet is assumed to be the header of a new train.

When looking for arrival processes suited to packet traffic flows, we should disregard the renewal processes since they cannot include correlation.

Batch Poisson processes [4,5] may account for the dispersion factor and approximate the typical grouping effect of the packet flows; but they cannot explain the correlation mentioned above. Nevertheless, due to their simplicity
they may be conveniently applied to cases in which this correlation does not seem relevant.

For the analysis of correlated flows, the superposition of Poisson processes has been applied. An excellent survey of the applicability of this type of process is given in [6]. Several analytical models have been developed [7,8,9] to derive the laws that govern the number of arrivals in an interval, stationary state probabilities and waiting times. The authors of this paper have paid special attention to [9], where the correlation is introduced by means of a Markov modulated Poisson process (MMPP) which approximates the superposition of packetized voice and data flows. The parameters that define this model are obtained from the moments of the number of arrivals in an interval. The model uses the powerful methodology developed in [7,8], which also proved very useful to these authors.

The paper focuses its attention on two models:

1. A batch Poisson arrival process with general batch size distribution applying to an m negative-exponential servers case. In the authors' opinion this model (BM/H/m) can approximate the traffic behavior of packet flows in CCITT No. 7 networks, where several signalling links may share a given traffic flow that originates at many independent sources, and convey packets of different lengths that do not seem strongly correlated. The parameters of the arrival process are obtained by fitting the moments of the number of arrivals to the statistical values found by measurement. Then the steady-state probability generating function (p.g.f.) and Laplace-Stieltjes Transform (LST) for the waiting time distribution are obtained.

2. A correlated arrival process based on the packet-train model in [3]. This process is formed by the superposition of M interdependent Poisson processes, called arrival modes in the model, in such a way that upon an arrival of a given mode l, the next arrival mode m, is scheduled according to a transition probability aLm. In this way the correlation between successive arrivals shown in [3] is modelled. The characterization of the arrival process can be made by matching the moments of the number of arrivals in an interval and/or the moments and correlations at lag k of the time intervals between arrivals, so that a better understanding of the arrival process can be obtained. This correlated arrival process, called the multi-mode arrival process (MM) in this paper, is then applied to the single server queue with general service time distribution (MM/G/1) in a first-in-first-out (FIFO) service discipline. The analysis is presented using matrix analytical procedures, from which we obtain the p.g.f. for the queue length at departure and at random instants, and the LST of the waiting time distribution.

Finally, some numerical results for the waiting time distributions given by these models are presented. The LST is numerically inverted using the Piessens method [10].

In the first model (BM/H/m), the geometric distribution of batch sizes has been selected so that the parameters of the arrival process are found through the first two moments of the number of arrivals. The results are compared with those provided by the M/H/m model.

For the application of the second model (MM/G/1), the case of two arrival modes is considered. The parameters that define the arrival process in this case are determined by matching the average interval times (overall and for each mode) and the auto correlation function between any two consecutive interval times, to the values provided by the measurements in [3]. Comparisons with the M/G/1 are presented.
2. MODEL I. BATCH (BM) ARRIVAL MODEL

In this section we describe the assumptions of the model and derive the corresponding arrival law and its associated moments. Then the steady state probabilities and the waiting time distribution at a queuing system with m negative exponential servers (BM/M/m) are computed.

2.1. Model Assumptions

The arrival process here considered is a batch Poisson process where:

- The interarrival times are identically and independently negative exponentially distributed with mean $\lambda^{-1}$. Upon arrival a batch of size $i$, with probability $g_i$, is generated. The p.g.f. for the batch size is $Z(\alpha)$.

- There are $m$ fully accessible servers in front of an unlimited queue. The service times of each server are identically and independently negative exponentially distributed with rate $\mu$.

- The service discipline is first-in-first-out (FIFO).

2.2. BM Arrival Process

The number of arrivals in any interval of length $t$ has a p.g.f. $G(\alpha,t)$ given by:

$$G(\alpha,t) = \exp\{-\lambda (1 - Z(\alpha)) t\} \quad (2.1)$$

The mean and variance of this process are:

$$N(t) = \lambda m z t \quad \text{and} \quad V(t) = N(t)[m_z + v_z/m_z] \quad (2.2)$$

$m_z$ and $v_z$ being the mean and variance of the batch size.

Knowing the moments of the arrival process, the parameter $\lambda$ and the p.g.f. $Z(\alpha)$ are immediately determined.

2.3. The BM/M/m System: Waiting Times

In order to derive the waiting time distribution, we first need the steady-state probabilities $\{p_i, i \geq 0\}$ of the system. From the birth-death equations applied to a batch process, the p.g.f is obtained as:

$$P(\alpha) = \frac{\mu (1 - \alpha)^{m-1} \sum_{i=0}^{m-1} (m - i) p_i \alpha^i}{m \mu (1 - \alpha) - \lambda \alpha [1 - Z(\alpha)]} \quad (2.3)$$

where the $p_i$ appearing in (2.3) can be obtained from the system birth-death equations in a recurrent manner.

The LST for the waiting time distribution is:
\[ W(s) = z^{-m+1} \left[ P(z) - \sum_{i=0}^{m-1} p_i z^i \right] + \sum_{i=0}^{m-1} p_i \]

where \( z = m \mu (s + m \mu)^{-1} \)

3. MODEL II. MULTI-MODE (MM) ARRIVAL MODEL

The MM arrival process herein developed attempts to incorporate the correlation between successive arrivals shown by the measurements. To accomplish this we will (1) establish the main assumptions of the model, (2) derive the arrival law and its moments and correlation functions, and (3) analyze the MM/G/1 queue.

3.1. Description of the Model

Let us consider a single server queue model with the following assumptions:

- An arrival process formed by the superposition of \( M \) interdependent negative exponential interarrival modes, each one with average value \( \lambda_m^{-1} \), \((m = 1, \ldots, M)\). The transition between any two modes \( 1, m \), is: immediately after an arrival of mode 1 occurs, the next arrival mode \( m \) is determined with transition probability \( \alpha_{1m} \) \((\sum_m \alpha_{1m} = 1)\). This arrival process is intended to introduce the correlation between successive arrivals mentioned in the introduction.

- The service times are identically and independently distributed positive random variables with distribution function \( S(x) \), whose \( i \)-th moment about the origin is \( h_i^{(1)} \).

- The service discipline is first-in-first-out (FIFO).

3.2. MM Arrival Process

In order to characterize the arrival process, we consider a \( M \)-state (mode) imbedded Markov chain at instants \( t_0, t_1, \ldots \), when an arrival has just occurred and the next arrival is scheduled.

Let \( U^1 \) denote the stationary probability associated with mode 1 at points \( t_0, t_1, \ldots \) If \( U^1 \) exists, the following equations must be verified:

\[ U^T = U^T A, \quad U^T e = 1 \]  

where \( U \) is a \( M \)-column vector with components \( U^1 \), \( A \) is the \( M \times M \) matrix of the \( \alpha_{1m} \) coefficients, and \( e \) is a \( M \)-column vector in which all components are 1.

The interval length between two successive arrivals has a density function \( L(t) \), whose LST is given by:

\[ L(s) = U^T \Lambda^{-1} \left( sI + \Lambda^{-1} \right)^{-1} e \]  

\( \Lambda \) being the \( M \times M \) diagonal matrix with the elements \( \lambda_1 \) along the diagonal.

The average length, \( L \), and variance, \( \text{Var} \), are then:
The stationary probabilities of being in mode $I$ at a random instant, $\pi^I$, is given, applying the key renewal theorem [11], by:

$$\pi^T = \frac{1}{L} U^T \Lambda^{-1}$$  \hspace{1cm} (3.4)

Now we define the probabilities of having $k$ arrivals at an interval $t$ of time between any two points $t_i$ and $t_{i+k}$ of the Markov chain.

Let $Q^{lm}_k(t)$ denote the conditioned stationary probability function that when the mode is $l$ at time $t_i$, there are $k$ arrivals during the interval $t$, finishing in mode $m$ at $t_{i+k}$. The p.g.f. in $k$ and LST in $t$ of the $Q^{lm}_k(t)$ is given by the following matrix expression [12]:

$$Q(\alpha,s) = [I - \alpha \Lambda (sI + \Lambda)^{-1} A]^{-1}$$  \hspace{1cm} (3.5)

Based on the above expression, the conditioned probability function $R^{lm}_k(t)$ that if the system is in mode $l$ at a given point $\tau$, there are $k$ arrivals during an interval of arbitrary length $t$, starting at $\tau$, and the mode at time $\tau + t$ is $m$, can be deduced. The p.g.f. of that function in matrix form is [12]:

$$R(\alpha,t) = \exp\{-\Lambda (I - \alpha A)t\}$$  \hspace{1cm} (3.6)

Thus, the p.g.f. of the number of arrivals during an arbitrary interval of time, $t$, is given by:

$$G(\alpha,t) = \pi^T R(\alpha,t) e = \pi^T \exp\{-\Lambda (I - \alpha A)t\} e$$  \hspace{1cm} (3.7)

which reduces to the Poisson process when all $\lambda_1 = \lambda$

The difference between this function and the one given in [9] is essentially due to the fact that in our model the change of mode is only possible upon arrivals, while in [9] the modes are governed by an infinitesimal generator $R$ which is not related to the occurrence of arrivals.

This arrival process is also characterized by the moments of the distributions of the time intervals between successive arrivals of the same mode (derivation of the LST for them is detailed in [12]), and the correlation functions for the time intervals between arrivals, at lag $k$, of any mode, which is given by [12]:

$$C_k = U^T \Lambda^{-1} A^k \Lambda^{-1} e - L^2$$  \hspace{1cm} (3.8)

and the Auto-Correlation Function at lag $k$, $(ACF(k))$, is obtained as the ratio of $C_k$ to the variance (3.3) of the arrival process.

### 3.3. The MM/G/1 System: Queue Length and Waiting Times

In order to obtain the waiting time distribution as seen by an external observer, we need first to derive the state probabilities at departure instants, then the state probabilities at random instants; then we can determine the waiting time.

Consider the imbedded Markov chain at instants $t_0$, $t_1$, ..., when customers leave the system. Let $D_1^n$ denote the stationary probability associated with the queue length at instants $t_0$, $t_1$, ..., $l$ being the arrival mode at those instants. The p.g.f. vector expression for this probability is [12]:

$$\text{4.2A.4.5}$$
\[ D^T(\alpha) = D_0^T (\alpha A - I) P(\alpha) \tilde{\phi}(\sigma(\alpha)) [\alpha I - \tilde{\phi}(\sigma(\alpha))]^{-1} P^{-1}(\alpha) \]  

(3.9)

\( P(\alpha) \) being the matrix of the eigenvectors resulting from the canonical decomposition of the \( M \times M \) matrix \( \Lambda (I - \alpha A) \) with eigenvalues \( \sigma_j(\alpha) \), and \( \tilde{\phi}(\sigma(\alpha)) \) being a \( M \times M \) diagonal matrix in which the non-null elements are the LST of the service time distribution \( S(x) \) at points \( \sigma_j(\alpha) \).

The function needed to determine the waiting time is the joint density function \( P^T(x) \), when at a random instant the mode is 1, there are \( n \) customers in queue (not including the customer in service), and the time needed to finish the current service is \( x \). Its p.g.f. and LST vector expression is [12]:

\[ P^T(\alpha; s) = P_0^T \Lambda (\alpha A - I) [s I + \Lambda (\alpha A - I)]^{-1} P(\alpha) \cdot \]

\[ \cdot [\alpha I - \tilde{\phi}(\sigma(\alpha))]^{-1} [\tilde{\phi}(\sigma(\alpha)) - S(s) I] P^{-1}(\alpha) \]  

(3.10)

where \( S(s) \) is the LST of the service distribution function. \( P^T_0 \) is the steady state probability of finding the system empty, given by:

\[ P^T_0 = (1/L_D) D^T_0 \Lambda^{-1} \]  

(3.11)

where \( L_D \) is the average time interval between departures, which is given by:

\[ L_D = h + D^T_0 \Lambda^{-1} e \]  

(3.12)

Assuming that the traffic in mode \( i \) as seen by an external observer is the product of \( \pi_i \) times the total traffic in the system, and then \( P(1; 0) = \pi \rho \), the indeterminate vector \( P_0 \) can be evaluated.

The waiting time experienced by an external observer that sees the system at random instant (virtual waiting time) has the following density function:

\[ w(s) = [P^T_0 + P^T(S(s); s)] e \]  

(3.13)

For \( \lambda_1 = \lambda \), the above expression becomes the waiting time of the \( M/G/1 \).

4. NUMERICAL RESULTS

This section presents some results obtained from the application of the above two models. In both cases the variation of the mean virtual waiting time as a function of the server load has been plotted and the resulting curves are compared with those corresponding to the classical Poisson arrival process.

In Fig.1 the case BM/M/3 has been considered. The parameter governing the set of curves is the mean value \( (s) \) of the batch size which is geometrically distributed.

Fig.2 presents two sets of curves (solid and dashed lines) for the MM/G/1 model, where the number of arrival modes is 2 (the packet-train model of [3]) and the service time is constant. The parameter of the curves is the autocorrelation function (ACF) previously defined. In the first set of curves (solid lines), the average interarrival times are \( t_1 = 100 \) ms. for mode 1 and \( t_2 = 2000 \) ms. for mode 2. For the second set (dashed lines), \( t_1 = 51.1 \) ms. and \( t_2 = 23773 \) ms., which exactly correspond to the case analyzed in [3]. It should be pointed out that in this case the average number of cars per train predicted by the analytical model [12] when ACF = 0.2 matches the measured result quite well (17.45 vs. 17.4).

4.2A.4.6
Both figures show a remarkable increase in the average waiting times as compared with those provided by the Poisson models (dot-dash lines). In the MM model (Fig. 2) we have found that the parameter ACF and the ratio $t_2/t_1$ (which is closely related to the average number of cars per train) are the most significant regarding the mean waiting time. In particular it is shown that the lower the ratio $t_2/t_1$, the closer MM/C/1 approaches the M/C/1 model for a given ACF value. It also appears that the average waiting times increase as the ACF increases.

From the above applications it can be stated that the effects of correlation and burstiness cannot be neglected when studying the performance of a system dealing with packet flows. Further investigation of the accuracy of the models, considering a wide mix of packet services and therefore a number of arrival modes greater than 2, is required.

5. CONCLUSIONS

This paper deals with two type of arrival processes intended to characterize packet traffic flows in present and coming communication networks in which the Poisson input process fails to represent the packet behavior resulting from the integration of several services, as it is shown by measurements. Two analytical queueing models, the BM/H/m and the MM/C/1 have been consequently developed that account for the grouping of packet arrivals (the first one) and the correlation between consecutive packet arrivals (the second one). The main results in the paper can be summarized as follows:

4.2A.4.7
1. Characterization of the arrival processes: average and variance values of the number of arrivals in an interval (first and second model) and moments and correlation functions for the interarrival times (second model);

2. Criteria to estimate the parameters of the analytical models from system measurements;

3. Stationary queue length probabilities at departure and random instants;

4. Distribution of the waiting times.

A computer program has been developed that provides numerical results for the above items in order to permit the proper dimensioning of system resources and optimization of system behavior.

Numerical results have been obtained for each model that demonstrate their applicability in practical cases, and comparison with classical models has been made as an example of the possible impact of factors such as grouping and correlation on system performance.

NOTES AND REFERENCES

NOTE: Because of lack of space, most formulas have been omitted from our presentation. They are contained in a report referred to in the text and available from ALCATEL Standard Electrica, S.A., Madrid.