

A HEURISTIC APPROACH FOR PERFORMANCE ANALYSIS OF ATM SYSTEMS

Ke-Qiang Liao and Lorne G. Mason

INRS-Telecommunications

3 Place du Commerce, Ile-des-Soeurs, Verdun, P.Q., Canada H3E 1H6

Abstract

For the design and control of ATM systems, it is important to have a performance model which produces reasonably accurate results over a wide range of parameters, since one major feature of ATM is that it can support many kinds of bursty traffic which have different communication speeds and time variations. We consider in this paper a heuristic approach which is based on the fluid approximation. In order to remove the possible inaccuracy of the fluid approximation in light to moderate traffic, we propose a G/D/1 adjustment. The performance analysis predicts the buffer content and packet delay distributions which usually have stringent requirements. An infinite buffer is assumed in the analysis. However for ATM networks, where the buffer overflow probability is small, the infinite buffer model well approximates the finite buffer case. A number of numerical examples show that this heuristic approach is very accurate over a wide range of system parameters.

1. INTRODUCTION

To construct a multiservice network capable of fulfilling all a user's requirements for data, voice and video communications, fast packet (also known as cell in ATM terminologies) switching technology seems promising for implementing an integrated access and transport network since the switching mechanism is independent of service bandwidth and offers potential gains in bandwidth efficiency by statistically multiplexing bursty traffic streams. The performance analysis of integrated services ATM systems has been an intense area of research in recent years motivated by the need to explore the advantages and the limitations of the novel and largely untried switch architectures. However most existing relevant work is performed in the context of packetized voice with suppressed silent periods or of packetized voice/data where the data arrival process is assumed to be a Poisson process.

The packetized voice with suppressed silent periods constitutes an example of bursty traffic. However, packetized voice systems have several characteristics which are not shared by other systems. For example, in a typical packetized voice system, the source peak rate is low compared with line speed and the average burst length (the average number of cells generated in a burst) is short. One major feature of ATM is that it can support many kinds of bursty traffic which have different communication speeds and time variations. It is therefore important to have a performance model which can produce reasonably accurate results over a wide range of system parameters.

The performance models in [1-8] are all performed in the packetized voice context and their applicability in the cases other than packetized voice systems has not been demonstrated. The MMPP model presented by Heffes and Lucantoni [1] gives very accurate results (at least for the mean and variance of delay) in packetized voice systems. In [9] it was shown that this model can be less accurate for other cases and an alternative method for the determination of the four parameters in MMPP was proposed. The latter method produces better results when the ratio of the source peak rate to the line speed is high or the average burst length is long.

In [10], Viterbi gave an explicit formula for the average packet delay when the ratio of the source peak rate to the line speed is unity. Bruneel [11] considered discrete-time statistical multiplexors. However

in his model the packet interarrival time of an active source is random variable rather than deterministic. For the special case where we have a deterministic packet interarrival time, the ratio of source peak rate to line speed is unity or higher.

In this paper, we consider a heuristic approach which basically is a fluid approximation [12] also referred to as uniform arrival and service (UAS) model in [2] and [6]. To remove the possible inaccuracy of the fluid approximation in light to moderate traffic, we propose an adjustment. More specifically, the packet delay is divided into two parts; the first part is approximated by the fluid approximation and the second by a G/D/1 queue.

In Section 2, we give a brief description of the system we deal with in this paper. The fluid approximation or UAS model is presented in Section 3. Section 4 describes in detail the G/D/1 adjustment while Section 5 considers the calculation of the packet delay as the sum of the delay approximated by UAS model and that by G/D/1 queue. Section 6 presents numerical results which are validated against simulation results.

2. SYSTEM DESCRIPTION

We first consider the packetized voice case. A voice source is active when the talker is actually speaking. During an active period, the voice generates fixed length packets at regular intervals. When the speaker is silent, the voice source is inactive and generates no packets. While it is difficult to model the speech activity of an individual speaker with accuracy, Weinstein [13] found that the number of active voice sources can be modeled very well by a continuous-time birth-and-death process. It is then commonly assumed that the lengths of both active and inactive periods are exponentially distributed. It follows that the number of packets generated in an active periods is geometrically distributed.

Here we take the same system model where each source has alternatively active periods (burst) and inactive periods. An active source generates packets with peak rate P and the packet interarrival time is a constant and equals P^{-1} . The lengths of both bursts and silent periods are assumed to be exponentially distributed. The number of packets generated in a burst is then geometrically distributed and we denote the mean by B which is called average burst length. The average silent period length is denoted by S which is measured in the same units as the average burst length. Without loss of generality, we assume that the output link speed is unity. As a result, the source peak rate P is also the ratio of source peak rate to link speed which is a very important parameter for resource management in an ATM network [14]. Packets have constant length, the transmission link acts as a single server and waiting packets are served in FIFO order. The service time is then deterministic. If the total number of sources is N , the system utilization rate ρ is given by

$$\rho = NP \frac{B}{B+S} \quad (1)$$

3. UNIFORM ARRIVAL AND SERVICE MODEL

In the UAS model, described in [12], each active source generates information to the buffer at a uniform rate of 1 unit of information per time unit and the server removes the information from the buffer at a uniform rate not to exceed C units of information per time unit. Here the server's capacity is C . When i sources are active simultaneously, as long as the buffer is not empty, the instantaneous rate of change is $i - C$. If the buffer empties while the number of active sources is less than C , the

Research supported by Bell-Northern Research, Canada.

buffer remains empty until the number of active sources again exceeds C . For reasons of analytical tractability, the capacity C can not be integer valued in the elegant algorithm developed in Anick, Mitra and Sondhi [12].

In order to fit our system to the model described in [12], it is necessary to make some parameter transformations. First, the server's capacity in our system is equivalent to P^{-1} . Second, in [12], the unit of time is taken to be the average duration of an active period. The unit of information is taken to be the amount of information that would be generated by a source during an active period of average length; a unit of information would therefore be equivalent to the average burst length B packets.

We now turn to the mathematical formulation of the system. Let $P_i(t, x)$, $0 \leq i \leq N$, $t \geq 0$, $x \geq 0$, denote the probability that at time t , i sources are active and the buffer content does not exceed x . We can first write a set of simple partial differential equations for $P_i(t, x)$ and then obtain a set of differential equations for $F_i(x) = \lim_{t \rightarrow \infty} P_i(t, x)$. In fact, we obtain, for $i \in [0, N]$,

$$(i - C) \frac{dF_i}{dx} = (N - i + 1) \lambda F_{i-1} - \{(N - i) \lambda + i\} F_i + (i + 1) F_{i+1} \quad (2)$$

where $\lambda = BS^{-1}$. The equations defining the functions F_i can be expressed in the following vector-matrix form (for details, see [12])

$$\frac{d}{dx} \mathbf{F}(x) = \mathbf{A} \mathbf{F}(x) \quad (3)$$

The solution to the differential equations (2) can be written as

$$\mathbf{F}(x) = \mathbf{F}(\infty) + \sum_{i=0}^{N-[C]-1} e^{z_i x} \alpha_i \Phi_i \quad (4)$$

where z_i are stable or negative eigenvalues of the matrix \mathbf{A} , Φ_i are the associated eigenvectors, α_i is determined by the boundary conditions and $[C]$ is the greatest integer less or equal to C . $F_i(\infty)$ is the probability that i out of N sources are on simultaneously and is given by the binomial distribution

$$F_i(\infty) = \frac{1}{(1 + \lambda)^N} \binom{N}{i} \lambda^i \quad (5)$$

Anick, Mitra and Sondhi [12] study in detail the system and they derive simple analytic formulae to compute the complementary buffer occupancy distribution, $G(x)$. The buffer occupancy corresponds, with a change of scale, to the virtual waiting time. In this paper, we are interested in packet delay rather than buffer occupancy. Taking into account the change of scale, we have

$$Pr(\text{delay} \leq x) = \frac{S + B}{NB} \sum_{i=0}^N i F_i(B^{-1}x) \quad (6)$$

where the time unit of delay is the packet service time.

A formula is presented in [12] to calculate the moments of the buffer content. In this formula, we need to calculate only eigenvalues of the matrix \mathbf{A} and the associated eigenvectors need not to be computed explicitly. Therefore if we are interested only in the average packet delay, we can use this formula and the Little's theorem without computing explicitly the eigenvectors. However in this paper we are also interested in the packet delay probability distribution function. If we first calculate the $F_i(x)$ and use the formula (6) to get the packet delay distribution function, we need to compute explicitly the eigenvectors Φ_i which is the most time consuming part. In the following we present a method to avoid the calculation of the eigenvectors.

Given an eigenvalue z , we can obtain three quantities r_1 , r_2 and k (for details, see [12]). The i th component of the associated eigenvector is written as

$$\phi_i = (-1)^{N-i} \sum_{j=0}^k \binom{k}{j} \binom{N-k}{i-j} r_1^{k-j} r_2^{N-k-i+j} \quad 0 \leq i \leq N \quad (7)$$

Let Θ denote the row vector $[0, 1, 2, \dots, N]$. Therefore we have

$$\begin{aligned} \Theta \Phi &= \sum_{i=0}^N i \phi_i \\ &= \sum_{i=0}^N i (-1)^{N-i} \sum_{j=0}^k \binom{k}{j} \binom{N-k}{i-j} r_1^{k-j} r_2^{N-k-i+j} \end{aligned} \quad (8)$$

After some algebraic manipulations, we obtain

$$\Theta \Phi = (1 - r_1)^k (1 - r_2)^{N-k} \left(\frac{k}{1 - r_1} + \frac{N - k}{1 - r_2} \right) \quad (9)$$

The packet delay distribution function can be expressed as

$$Pr(\text{delay} \leq x) = 1 + \frac{S + B}{NB} \sum_{i=0}^{N-[C]-1} e^{z_i B^{-1}x} \alpha_i (\Theta \Phi_i) \quad (10)$$

We observe that the n th moment of the packet delay

$$E\{x^n\} = \int_0^\infty x^n dPr(\text{delay} \leq x) \quad (11)$$

We then obtain

$$E\{x^n\} = \frac{n!}{(-1)^{n+1}} \sum_{i=0}^{N-[C]-1} \frac{\alpha_i (\Theta \Phi_i)}{z_i^n} \quad (12)$$

The formula (9) improves considerably the computational efficiency.

Maglaris et al. [15] used a fluid approximation to obtain the queue length distribution in a packet video multiplexor. The UAS model was used in [6-8] to study packetized voice systems with a finite buffer size. In [16], Roberts and Simonian used a two-state fluid approximation to analyse packet voice and video multiplexors. For the packetized voice system, the two-state fluid approximation, as the UAS model, accurately estimates the mean delay in heavy traffic but seriously underestimates it for low loads.

4. G/D/1 ADJUSTMENT

As indicated in [2] and [6], the UAS model lacks the concept of packetization. It follows that this model ignores the "high frequency" variations in buffer content present in the real system. From a multi-layer concept, Filipiak [17] concluded that the fluid approximation can only characterize the delay in the burst layer but does not capture the small time scale fluctuations in the packet layer. He also proposed a multi-layer analysis where the input process is modeled by a multi-layer Markov Chain. The discrete-time queueing model in the burst level is basically a fluid model.

It is well known that the fluid approximation generally does not produce accurate results in light to moderate traffic. Here we use the word "generally" since the fluid approximation gives very accurate results when the peak rate is high even in the light traffic case. This can be explained by the fact that the UAS model ignores "high frequency" variations in buffer content, since in the real system, information does not enter the transmission buffer, and therefore can not be transmitted, until a particular source completes generation of one packet. It follows then that the higher is the source peak rate (the shorter is the packet interarrival time), the slighter is the effect of packetization on the accuracy of UAS model. The effect will completely disappear when the source peak rate is unity or higher.

To cope with the inaccuracy of the UAS model in light to moderate traffic caused by the packetization effect, we propose a G/D/1 adjustment. Basically, we divide the packet delay into two parts; one part is approximated by fluid approximation while the other is approximated by G/D/1 queue. Based on several comparisons between analytical and simulation results, we find that when the system is heavily loaded or the source peak rate is high, the packet delay can be very well estimated by the fluid approximation while in the alternative case, the M/D/1 approximation produces very good results. Here we use a two moment approach for the G/D/1 where the mean packet interarrival time is equal to that of the superposition process. As far as the coefficient of variation c_a is concerned, it should be a function of the system utilization rate ρ and the source peak rate P . Based on the above-mentioned remarks, $c_a(\rho, P)$ should have the following properties:

- 1) $\lim_{\rho \rightarrow 1} c_a(\rho, P) = 0$
- 2) $\lim_{P \rightarrow 1} c_a(\rho, P) = 0$
- 3) $c_a(\rho, P)$ is a decreasing function of ρ and P
- 4) $\lim_{P \rightarrow 0, \rho \rightarrow 0} c_a(\rho, P) = 1$

One possibility for the choice of the function $c_a(\rho, P)$ is

$$c_a(\rho, P) = (1 - \rho^m)(1 - P^n) \quad n, m > 0 \quad (13)$$

which has obviously the above-listed properties. A comparison between the simulation and the approximate results for different values of m and n indicates that the set $m = 6$ and $n = 1$ gives the best results.

For the analysis of the G/D/1 queue, we will calculate the probability of waiting, the mean and the variance of delay by using different approximations. For the probability of delay and the average delay, we use the Kraemer and Langenbach-Belz approximations [18]. Let σ and EW denote the probability of delay $Pr(W > 0)$ and the average delay respectively. Noting that in our case the coefficient of variation of service time equals 0 and the coefficient of variation of interarrival time $c_a(\rho, P) \leq 1$, we have then

$$\sigma = \rho + (c_a^2 - 1)\rho(1 - \rho) \frac{1 + c_a^2}{1 - \rho + 4\rho^2 c_a^2} \quad (14)$$

and

$$EW = \frac{\rho c_a^2 h}{2(1 - \rho)} \exp\left[-\frac{2(1 - \rho)}{3\rho} \left(\frac{1 - c_a^2}{c_a}\right)^2\right] \quad (15)$$

where h is the packet service time.

Here we can see that the average delay of the G/D/1 queue is proportional to the packet service time while the delay estimated by the fluid approximation is independent of the packet service time distribution. If all the characteristics of sources are held fixed but the packet size is decreased (the mean number of packets generated during an active period is increased), the delay estimated by the G/D/1 queue is scaled down. Decreasing the packet size increases the difference between the burst time scale and the packet time scale (multiple time scales). When the mean number of packets generated during an active period is very large, this delay can be ignored. This presence of multiple time scales provides the justification for fluid models in several cases [19,20].

In principle, it is difficult or impossible to find the second moment of the delay distribution, since generally its determination involves the third moments. In order to get an approximation for the variance of the delay, we use an argument by Whitt [21]. He observed that the conditional delay given that the server is busy in a GI/G/1 queue (rather than the total delay) depends more on the service-time distribution than on the interarrival time distribution. Let D denote the conditional delay given that the server is busy. Evidently, we have $ED = EW/\sigma$. As for the second moment of D , if we denote the squared coefficient of variation of D by c_D^2 , the M/D/1 formula for c_D^2 is used as an approximation for G/D/1 systems. The M/D/1 formula for c_D^2 is

$$c_D^2 = \frac{2\rho + 1}{3} \quad (16)$$

From D we then obtain second-moment characteristics for W

$$c_W^2 = \frac{E(W^2)}{(EW)^2} - 1 = \frac{c_D^2 + 1 - \sigma}{\sigma} \quad (17)$$

and

$$Var(W) = (EW)^2 c_W^2 \quad (18)$$

We now proceed with the approximation of the probability distribution for W . The distribution has an atom at zero as given in (14). The density is chosen so that the first two moments fit those of W and D which have already been determined. Here we consider only the case where $c_D^2 < 1$, since from (16) we can see that c_D^2 can not be greater than 1.

Case 1: $0.5 < c_D^2 < 1$. Let the distribution of D be the convolution of two exponential distributions with parameters γ_1 and γ_2 ($\gamma_1 > \gamma_2$), i.e., let D have density

$$f_D(x) = \frac{\gamma_1 \gamma_2}{\gamma_1 - \gamma_2} (e^{-\gamma_2 x} - e^{-\gamma_1 x}) \quad (19)$$

with

$$\gamma_2^{-1} = \frac{ED + \sqrt{2Var(D) - (ED)^2}}{2} \quad (20)$$

and

$$\gamma_1^{-1} = ED - \gamma_2^{-1} \quad (21)$$

The corresponding probability distribution is

$$P(D \leq x) = 1 - (\gamma_1 e^{-\gamma_2 x} - \gamma_2 e^{-\gamma_1 x}) / (\gamma_1 - \gamma_2) \quad (22)$$

Case 2: $c_D^2 \leq 0.5$. Let D have a gamma distribution with parameters n and γ . Its density is

$$f_D(x) = \frac{(\gamma x)^{(n-1)}}{\Gamma(n)} \gamma e^{-\gamma x} \quad (23)$$

with

$$n = [c_D^{-1}] \quad (24)$$

and

$$\gamma = n(ED)^{-1} \quad (25)$$

where the symbol $[\cdot]$ denotes the largest integer part of its argument. Here in order to facilitate the numerical calculation, we take an integer value for n and the coefficient of variation is only approximately matched.

5. PACKET DELAY DISTRIBUTION

As we mentioned in the preceding section, in order to cope with the possible inaccuracy of the fluid approximation in light to moderate traffic, we propose a G/D/1 adjustment. Therefore the packet delay is divided into two parts: one part is approximated by the fluid approximation and the other by a G/D/1 queue. We use a two moment approach for the analysis of the G/D/1 queue whose arrival process depends on several system characteristics. Based on a comparison between simulation and analytical results, we find that the coefficient of variation of the interarrival time should be a function of the source peak rate and the system utilization rate to well approximate the packet delay.

Let W_1 denote the part of packet delay approximated by UAS model and W_2 the other part. W_1 and W_2 are assumed to be two mutually independent random variables whose approximate probability distributions are given in Section 3 and in Section 4, respectively. Denoting the packet delay by W , we then have $W = W_1 + W_2$ and its probability distribution is approximated by the convolution of the approximate probability distributions of W_1 and W_2 . The calculation is straightforward but tedious. Its detail is not reported here.

While the calculation of the packet delay probability distribution is complicated, its mean and variance are easy to obtain. In fact, we have the simple formulae:

$$\begin{aligned} E(W) &= E(W_1) + E(W_2) \\ Var(W) &= Var(W_1) + Var(W_2) \end{aligned} \quad (26)$$

6. NUMERICAL RESULTS AND DISCUSSIONS

For the analysis in all the numerical examples, we always use the set $\{n = 6, m = 1\}$ in Equation (13) which is shown by a comparison between simulation and analytical results to be the best to give reasonably accurate results over a wide range of system parameters. In addition the unit of time of the delay is the packet service time. In the following we use the number of sources N , the source peak rate P , the average burst length B and the utilization rate ρ to describe a system. The average inactive period length can be easily obtained from these parameters. We have

$$S = B \left(\frac{NP}{\rho} - 1 \right) \quad (27)$$

We have programmed the MMPP model proposed by Heffes and Lucantoni [1] (but the resulting MMPP/D/1 queue is solved by a discrete-time approach described in [9] rather than the matrix-geometric method they used in [1]). For the purpose of comparison to the results of [1], [3], [4] and [16], we first consider the packetized voice system with line speed 1.536 Mbits/s, voice packet length 64 bytes, mean active period 353 ms (22 packets) and mean silent period 650 ms. In Table 1, results of six approximations are presented: 1), GI/D/1 queue by Sriram and Whitt [3]; 2), 2-state MMPP/D/1 by Heffes and Lucantoni [1]; 3), N-IPP/D/1 by Ide [4]; 4), 2-state fluid approximation by Roberts and Simonian [16]; 5), fluid approximation [12]; 6), the heuristic approach presented in this paper. The simulation results are cited from [2]. From this table, we can see that only the heuristic approach produces results which are all in the confidence intervals and that the heuristic approach significantly improves the fluid approximation results in light to moderate traffic ($\rho \leq 0.878$).

In Table 2, a number of examples are presented with different source peak rates, average burst lengths and utilization rates for the average

No. of voice sources N	simulation (95% C.INT.)	approximations [ms]					
		1	2	3	4	5	6
20	0.03 (± .001)	0.04	0.03	0.03	0.00	0.00	0.03
40	0.07 (± .001)	0.08	0.07	0.07	0.00	0.00	0.07
60	0.13 (± .001)	0.16	0.14	0.14	0.00	0.00	0.13
80	0.22 (± .007)	0.31	0.27	0.26	0.00	0.00	0.22
90	0.31 (± 0.02)	0.45	0.38	0.37	0.00	0.00	0.29
100	0.45 (± 0.14)	0.72	0.57	0.57	0.04	0.02	0.40
110	0.89 (± 0.14)	1.36	1.04	1.14	0.50	0.36	0.84
120	4.07 (± 0.4)	3.75	3.57	4.15	3.6	3.49	4.10
125	10.4 (± 1.3)	8.5	11.33	10.25	9.3	10.14	10.83
130	31.9 (± 4.8)	28.4	37.25	31.41	28.9	32.60	33.36
132	52.1 (± 7.5)	55.3	63.98	55.73	52.0	57.90	58.70
134	109.6 (± 21.4)	129.3	131.38	120.58	115.	124.65	125.68

Table 1 Comparison of approximations of the average packet delay

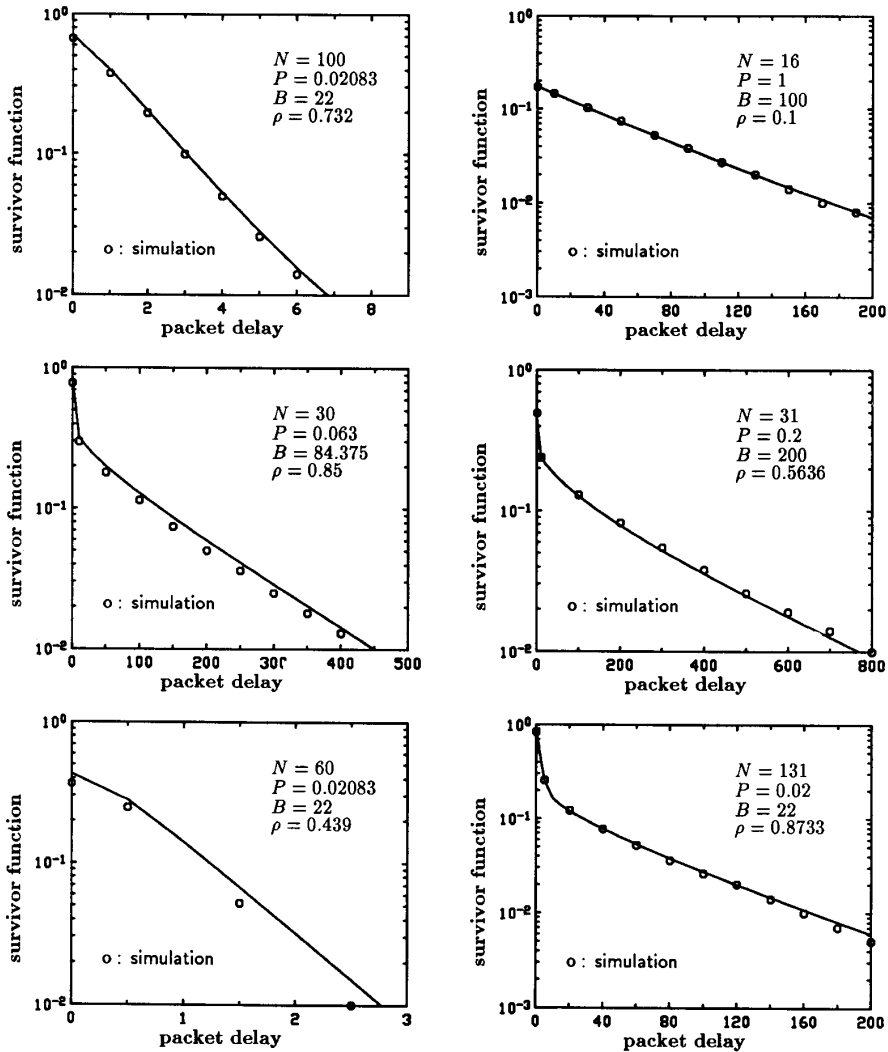


Fig. 1 Packet delay survivor function

N	P	B	ρ	simul.(95% C. INT.)	approx.
14	0.250	100	0.5675	23.63 (\pm 5.81)	25.06
140	0.025	100	0.5675	0.59 (\pm 0.03)	0.62
16	0.100	100	0.9000	88.50 (\pm 7.46)	84.70
16	0.400	100	0.9000	547.70 (\pm 70.04)	594.65
16	0.700	100	0.9000	663.46 (\pm 172.86)	735.19
16	1.000	100	0.9000	801.32 (\pm 114.31)	787.52
16	0.700	100	0.1000	5.91 (\pm 1.06)	6.53
16	1.000	100	0.1000	9.82 (\pm 1.89)	10.32
42	0.042	22	0.8750	14.71 (\pm 2.25)	10.43
128	0.020	22	0.8587	9.16 (\pm 1.96)	7.78
131	0.020	22	0.8733	10.82 (\pm 1.31)	11.26
21	0.267	200	0.5087	49.00 (\pm 5.80)	46.45
22	0.200	200	0.4000	10.68 (\pm 3.25)	7.47
31	0.200	200	0.5636	48.30 (\pm 14.66)	49.46
2	0.889	200	0.1616	16.50 (\pm 2.68)	15.57
3	0.778	200	0.2121	22.87 (\pm 3.16)	24.44
4	0.694	200	0.2525	34.17 (\pm 5.20)	29.88
32	0.100	500	0.1000	0.15 (\pm 0.17)	0.05

Table 2 Comparison of simulation and analytical results for average packet delay

packet delay. The agreement between simulation results and the heuristic approach is excellent over a wide range of system parameters: ρ from 0.1 to 0.9, P from 0.02 to 1.0 and B from 22 to 500. All the analytical results except one (the 9th example) are in the confidence intervals. This shows that the heuristic approach is robust over a wide range of system parameters.

As we mentioned in Section 4, the fluid approximation generally does not give accurate results in light to moderate traffic. But if the source peak rate is high compared with the line speed, the fluid approximation produces very accurate results even in light traffic case. Here we present two examples. The parameters of the first example are $N = 14$, $P = 0.25$, $B = 100$ and $\rho = 0.5675$. The simulation estimation of the average packet delay is 23.63 with a 95% confidence interval [17.82, 29.44] and the fluid approximation gives an estimation of 24.61. In the second example, we have $N = 16$, $P = 1.0$, $B = 100$ and $\rho = 0.1$. The simulation and analytical results are 9.82 (± 1.96) and 10.32, respectively.

Study of packet loss is of great importance to the design of ATM systems. Here we have an infinite buffer system. However, if the packet loss probability is very low, an infinite buffer system can constitute an accurate approximation for a finite buffer system. Now we are speaking of a packet loss probability between 10^{-6} to 10^{-9} in an ATM network [22]. In that range, a truncated packet delay distribution in an infinite buffer system can be used as an accurate approximation for packet delay distribution in a finite buffer system. Fig. 1 show the complementary cumulative distribution functions (or survivor functions) for different system parameters. It is seen that the heuristic approach gives also a good approximation for the packet delay probability distribution function. It was shown [2] that generally, the fluid approximation overestimates the probability that the queue is empty. But the heuristic approach produces reasonably good results for the delay probability. The case where the heuristic approach gives the worst results is with system parameters: $N = 60$, $P = 0.02083$, $B = 22$ and $\rho = 0.439$. In this case the delay distribution is essentially approximated by a $G/D/1$ queue and the average delay estimated by the fluid approximation is zero.

7. CONCLUSIONS

We have presented a heuristic approach for the performance analysis of ATM systems. This approach is basically a fluid approximation and a $G/D/1$ queue adjustment which removes the possible inaccuracy of the fluid approximation in light to moderate traffic. The performance analysis predicts the buffer content and packet delay distributions. Comparisons with simulation show the approach to be accurate over a wide range of system parameters.

References

- [1] H. Heffes and D.M. Lucantoni, 'A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance', IEEE J. Selec. Areas in Commun., SAC-4, 6, pp. 856-868, Sept. 1986.
- [2] J.N. Daigle and J.D. Langford, 'Models for analysis of packet voice communication systems', IEEE J. Selec. Areas in Commun., SAC-4, 6, pp. 847-855, Sept. 1986.
- [3] K. Sriram and W. Whitt, 'Characterizing superposition arrival processes in packet multiplexers for voice and data', IEEE J. Selec. Areas in Commun., SAC-4, 6, pp. 833-846, Sept. 1986.
- [4] I. Ide, 'Superposition of interrupted Poisson processes and its application to packetized voice multiplexers', 12th Int. Teletraffic Cong., Turin, paper 3.1B.2, 1988.
- [5] S.-Q. Li and J.W. Mark, 'Performance of voice/data integration on a TDM system', IEEE Trans. on Communications, Vol. COM-33, No. 12, pp. 1265-1273, Dec. 1985.
- [6] R.C.F. Tucker, 'Accurate method for analysis of a packet-speech multiplexer with limited delay', IEEE Trans. on Communications, Vol. COM-36, No. 4, pp. 479-483, April 1988.
- [7] N. Yin, S.-Q. Li and T. Stern, 'Congestion control for packet voice by selective packet discarding', IEEE GLOBECOM'87, pp. 1782-1786, Tokyo, Nov. 1987.
- [8] S.Q. Li, 'Study of packet loss in a packet switching voice system', ICC'88, pp. 1519-1526, Philadelphia, June 1988.
- [9] K.-Q. Liao and L.G. Mason, 'A discrete-time single server queue with a two-level modulated input and its applications', IEEE GLOBECOM'89, paper 26.1, Dallas, Texas, Nov. 1989.
- [10] A.M. Viterbi, 'Approximate analysis of time-synchronous packet networks', IEEE J. Selec. Areas in Commun., SAC-4, 6, pp. 879-890, Sept. 1986.
- [11] H. Bruneel, 'Queueing behavior of statistical multiplexers with correlated inputs', IEEE Trans. on Communications, Vol. 36, No. 12, pp. 1339-1341, Dec. 1988.
- [12] D. Anick, D. Mitra and M.M. Sondhi, 'Stochastic theory of a data-handling system with multiple source', B.S.T.J., pp. 1871-1894, Oct. 1982.
- [13] C.J. Weinstein, 'Fractional speech loss and talker activity model for TASI and for packet-switched speech', IEEE Trans. on Communications, Vol. COM-26, pp. 1253-1257, Sept. 1978.
- [14] K. Noguchi, T. Okada and H. Ohnishi, 'Resource management in an ATM network', 2nd IEEE ComSoc Int. Multimedia Communications Workshop, Montebello, Canada, April 1989.
- [15] B. Maglaris et al., 'Performance models of statistical multiplexing in packet video communications', IEEE Trans. on Communications, Vol. COM-36, No. 7, pp. 834-843, July 1988.
- [16] J. Roberts and A. Simonian, 'Some queueing models for an ATM multiservice network', Doc. Cost-224, Issue 1 (Private Communication).
- [17] J. Filipiak, 'Multi-layer analysis of packet delay and blocking in statistical multiplexing', submitted for publication.
- [18] W. Kraemer and M. Langenbach-Belz, 'Approximate formulae for the delay in the queueing system $G/G/1$ ', 8th Int. Teletraffic Cong., Melbourne, pp. 235-1/8, 1976.
- [19] D.P. Gaver and J.P. Lehoczky, 'Channels that cooperatively service a data stream and voice messages', IEEE Trans. on Communications, Vol. 30, No. 5, pp. 1153-1162, May 1982.
- [20] D. Mitra, 'Stochastic theory of a fluid model of multiple producers and consumers coupled by a buffer', Adv. Appl. Prob., Vol. 20, pp. 646-676, 1988.
- [21] W. Whitt, 'The queueing network analyzer', Bell Syst. Tech. J., Part 1, Vol. 62, No. 9, pp. 2779-2815, Nov. 1983.
- [22] G.M. Woodruff, G.H. Rogers and P.S. Richards, 'A congestion control framework for high-speed integrated packetized transport', IEEE GLOBECOM'88, paper 7.1, Hollywood, Florida, Nov. 1988.