4

Mean End-to-End Packet Delays in Hybrid Integrated Services Networks

T.V.J. Ganesh Babu and A. Kumar
Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore 560 012, INDIA.
e-mail: anurag@ece.iisc.ernet.in

Abstract

We study a scheme for approximate calculation of mean *end-to-end* packet delay in a network that integrates circuit switched and packet switched traffic, using the movable boundary traffic integration scheme on each link. Numerical results obtained from the approximation are compared with simulation results. We analyse the discrepancies between the simulation and the approximation. The analysis reveals an important phenomenon that occurs when packets and circuits are integrated in a *network*. When a link is shared between circuit switched and packet switched traffic, using the movable boundary scheme, then the packet departure process is burstier than Poisson. In a network of such links, at light circuit switched loads, the active calls occupy multiple hops, hence bursts of packets departing from a link are likely to find the next link available. Thus at light circuit loads there is a positive correlation between packet bursts and link availability, yielding packet delays less than what would be expected on the grounds of burstiness, and causing the Poisson process based approximation to overestimate mean packet delay. At higher circuit loads the active circuits occupy fewer hops, and the correlation is diminished; the Poisson approximation then underestimates the delay of the bursty packet arrivals.

Keywords

Circuit switching, packet switching, burst reservation, queueing models, Poisson arrival approximation, correlation effects in queuing networks

1 INTRODUCTION

In this paper we study end-to-end packet delays in integrated services networks that multiplex circuit switched and packet switched traffic using the movable boundary scheme [Fischer and Harris, 1976] on each link. Although much work has appeared in the literature, on the performance of the movable

boundary scheme on a single link (e.g., [Atmaca and Pujolle, 1992], [Kraimeche and Schwartz, 1985], [Rosenberg and Le Bon, 1993]), little is known about the end-to-end performance when such links are interconnected to form networks.

The objective of this paper is to study the quality of and the errors in a natural approximation scheme for estimating the mean end-to-end packet delays. Calculations are carried out for the linear network topology (i.e., a linear array of nodes and full-duplex links) with single and multiple channels on each link. Path counting techniques, or well known approximation techniques, are used for calculating link blocking probabilities due to circuit switched traffic. The approximation uses these blocking probabilities to replace each link with an equivalent isolated link model. Mean packet delays from these link models are summed to obtain mean end-to-end packet delays.

Careful simulations, using the regenerative technique, are used to understand the accuracy of the approximation. Detailed stochastic analysis is performed to understand the discrepancies between the approximation results and the simulation results. Apart from explaining the trends in the errors in the approximation, this analysis reveals an important effect that will arise only in the network context. The movable boundary scheme, for sharing a link between circuits and packets, yields bursty packet departures. We show that the presence of multiple link circuit switched calls in the network tends to reduce packet delays below what would be expected on grounds of burstiness. As the circuit arrival rate of each source-destination pair increases, circuit switched calls tend to become shorter in length. It turns out that shorter circuit switched calls in the network result in larger packet delays.

Models of integration of circuit switched traffic and packet switched traffic on a TDM link have been analysed in [Fischer and Harris, 1976] [Sriram et al, 1983] [Li and Mark, 1985] [Yang and Mark, 1991] in the discrete time domain, and in [Atmaca and Pujolle, 1992] [Kraimeche and Schwartz, 1984] [Kraimeche and Schwartz, 1985] [Weinstein et al, 1980] [Williams and Garcia, 1984] in continuous time domain. Continuous time analysis is adopted here, i.e., if on a link there are C slots per frame, then these C servers are assumed to be operating as if they are C independent channels in parallel.

Two notable references on the end-to-end performance analysis of integrated circuit/ packet switching networks are [Li and Mark, 1988] and [Leon-Garcia and Aboul-Magd, 1986]. Our work differs significantly from these two references, in that we are concerned with the dynamics of the occupancy of network links by circuit switched traffic and the effect of this on the packet switched traffic. In particular, our work reveals the effect of the correlation between packet arrivals at a link and the availability of free channels on the link. Under lower circuit arrival rates for each source-destination pair, there is positive correlation, hence packet delays are less than one might expect based just on the burstiness of the packet arrival process. At higher circuit arrival rates for each source-destination pair, this correlation decreases and packet delays increase.

While the work reported in this paper has been motivated by channelised narrowband ISDN, very similar models will arise if in B-ISDN/ATM isochronous connections are allocated peak bandwidth, and data services are given "best-effort" service (as in the Internet). Further, in this scenario, data services

could be allowed to use bandwidth temporarily left vacant by variable-bit-rate isochronous connections in between bursts (movable boundary).

The outline of this paper is as follows. In Section 2 we describe the model. In Section 3 we explain how the approximate analysis obtains mean end-to-end packet delays. In Section 4 we discuss the numerical results obtained from the approximation and from simulations. In Section 5 we make an extensive study to understand the discrepancies. In Section 6 we see how the approximation technique is applied to networks with many channels per link, and discuss simulation results for a particular type of network. In Section 7 we present our conclusions and discuss possible extensions of this work.

2   THE NETWORK MODEL

We consider a communication network, consisting of N nodes (switches) and L links. Link $j, 1 \leq j \leq L$ has c_j full-duplex channels; m_j channels are used by circuit traffic and $(c_j - m_j)$ channels are dedicated to packet traffic. Packets can use free circuit channels (i.e., the movable boundary scheme), but circuits have preemptive priority over packets. Hence, as far as circuit traffic is concerned, we just have a circuit switching network with m_j channels on link j, $1 \leq j \leq L$.

External arrival processes are assumed to be Poisson. We assume that for every source-destination pair $i, 1 \leq i \leq N(N-1)$, the circuit arrival rate is λ_c and the packet arrival rate is λ_p. The holding times of all circuit switched calls is exponentially distributed with mean μ_c^{-1}. All packet service times are exponentially distributed with mean μ_p^{-1}.

We assume fixed routing for circuits and packets. Routing for circuit traffic and packet traffic is defined by the same routing matrix R, where R is an $N(N-1) \times L$ matrix whose element $r_{ij} = 1$ if the route for the source-destination pair i uses link j, and $r_{ij} = 0$ otherwise.

3   THE APPROXIMATE ANALYSIS METHOD

We calculate the probability at each link that all the circuit channels are busy for a given λ_c and μ_c. Next we calculate the composite arrival rate of packets at each link, from the routing matrix R and the packet arrival rates for all the source-destination pairs. Then we consider each link in isolation, and assume that the blocking probability of circuit channels on the link is due to an independent Poisson circuit arrival process to this link. Further, we assume that the composite arrival process of packets to the same link is Poisson with arrival rate as calculated above. We now have an analysable model that can be used to obtain the approximate mean delay for packets at that link in the network. The mean end-to-end packet delay is obtained by summing the mean packet delays over the various links in the route of a packet.

3.1 Link blocking probability calculation

Let $a = \frac{\lambda_c}{\mu_c}$. In steady state, let $\underline{X} = (X_1, X_2, \ldots, X_{N(N-1)})$ denote the random vector whose j^{th} component X_j denotes number of calls active on route j. Let

$$\pi(\underline{x}) := P(\underline{X} = \underline{x}), \quad \underline{0} \leq \underline{x}R \leq \underline{m}$$

denote the joint distribution of the random vector $\underline{X}$. Here, $\underline{0}$ is the L-vector of all 0's and $\underline{m}$ is the L-vector of m_j's. It follows from the product form solution [Burman et al, 1984] [Kelly, 1986] that

$$\pi(\underline{x}) = \frac{1}{G(a)} a^{\sum_{i=1}^{N(N-1)} x_i} \cdot \prod_{i=1}^{N(N-1)} \frac{1}{x_i!} \quad \text{for } \underline{x}, \quad \text{such that} \quad \underline{0} \leq \underline{x}R \leq \underline{m}$$

where $G(a)$ is for normalisation of $\pi(\cdot)$ to 1. Letting $M = \sum_{j=1}^{L} m_j$, it follows that there exist constants $A_K^{(N)}, 1 \leq K \leq M$, such that

$$G(a) = 1 + A_1^{(N)} \cdot a + \ldots + A_M^{(N)} \cdot a^M$$

For networks with simple topologies and only one channel per link, with uniform traffic for circuits, it is possible to calculate the blocking probability of each link directly from the product form solution [Babu, 1991]. For networks with multiple channels on each link, it becomes extremely difficult to calculate these link blocking probabilities directly from the product form solution. So in these cases we resort to an approximate iterative method to calculate the link blocking probabilities (see Section 6.1).

3.2 Model for packet delay on a single link in the network: single channel per link

In the approximate link model, the circuit arrival rate γ_c is obtained as the Poisson arrival rate that yields the blocking probability β for this link in the network, and the packet arrival rate γ_p is obtained from λ_p and the routing matrix R. Thus, γ_c is obtained by Erlang's blocking formula. Since we are interested in packet delays, when the server is working on circuits it is equivalent to the server being on "vacation". The preemptive vacation model with service order independent of service times was discussed by Gaver in [Gaver, 1962]. Denote by S the stationary sojourn time of packets in this model. For, $\frac{\gamma_c}{\mu_c}(1 - \beta) + \frac{\gamma_p}{\mu_p} < 1$, using Laplace transform technique and using regenerative analysis [Babu, 1991], we can write,

$$E(S) = \frac{1}{\mu_p} \left[\frac{(1 + \rho_c)^2 + \alpha \rho_c}{(1 - \rho_p(1 + \rho_c))(1 + \rho_c)} \right]$$

where $\rho_c = \frac{\gamma_c}{\mu_c}$, $\rho_p = \frac{\gamma_p}{\mu_p}$ and $\alpha = \frac{\mu_p}{\mu_c}$. This result can also be derived using Markov chain analysis and the z-transform technique as shown in [Schwartz, 1987].

4 NUMERICAL RESULTS (APPROXIMATION AND SIMULATION): LINEAR NETWORKS, SINGLE CHANNEL PER LINK

4.1 Approximate analysis

In the single channel per link case, each link has two servers one for either direction. A circuit switched call is bidirectional and hence seizes both servers of each link in its path. Packets traverse the network unidirectionally, and hence require service from only one server of a link in the direction of traversal. The busy probability due to circuits on each link of an N-node linear network, with a single channel per link, can be obtained directly from the product form solution [Babu, 1991]. In the case of linear networks, the total packet arrival rate on link j (we use the convention of labelling the links in the linear network from left to right), in both directions, based on uniform packet traffic assumption, is $\gamma_p = j(N - j)\lambda_p$. We calculate the approximate mean packet delay at each link using the isolated link model discussed in the Section 3.2. This calculated approximate mean delay will be the same for packets in both directions of a link.

   We fix the mean holding time for circuits and mean service time for packets in our entire calculations and simulations. We have chosen $\mu_c = 0.01$ and $\mu_p = 0.1$. Observe that these parameter values are quite reasonable for a voice-data integrated network that does fast circuit switching of voice bursts and packet switching of data. This combination of parameter values has led to some very interesting phenomena.

4.2 Comparison with simulation

The simulation was carried out using the SIMSCRIPT II.5 simulation language. Our main focus is on estimating the mean delay experienced by packets. We use regenerative simulation to obtain the estimates and confidence intervals [Lavenberg, 1983], [Wolff, 1989]. We use as regeneration epochs those departure epochs of packets that leave the network empty.

   The central link in a linear network with an even number of nodes, or the two central links in a linear network with an odd number of nodes, experience the maximum blocking probability and a maximum arrival rate of packets. Since a linear network is not symmetric, the mean packet delay on each link has to be estimated individually, in order to estimate mean end-to-end packet delay. Since the load on the central link/s is the maximum, however, we only maintain regeneration statistics for the central link/s. The simulation is run until an adequately narrow confidence interval is obtained for the central link/s. It is then expected that the confidence interval for the estimated mean packet delays on the other links, are at least as narrow.

   The simulation results and the corresponding values obtained from the approximate analysis, for various values of λ_c and keeping λ_p constant, are shown in Tables 1, 2 and 3 for the second link in a 3-node (2 link) linear network. In each of these tables (and in the tables to follow) two sets of simulation results

Table 1: Approximate mean packet delay and its corresponding simulation result for link 2 of a 3-node linear network with $\mu_c = 0.01$ and $\mu_p = 0.1$. Packet load on link 2 = 0.1

Circuit occupancy on Link 2	Mean delay on Link 2 With 95% confidence interval (Sampling of packet size done only once at network entry)	Mean delay on Link 2 with 95% confidence interval (Sampling of packet size done at each link)	Mean delay from approximate analysis
0.1	21.35 ± 0.19	21.17 ± 0.19	23.75
0.2	32.69 ± 0.28	32.48 ± 0.28	37.14
0.3	45.49 ± 0.36	45.24 ± 0.36	51.67
0.4	60.99 ± 0.46	60.84 ± 0.46	68.00
0.5	81.66 ± 0.59	81.02 ± 0.59	87.50
0.6	109.81 ± 1.57	108.87 ± 1.59	113.33
0.7	159.64 ± 2.49	158.79 ± 2.54	155.00
0.8	279.01 ± 5.19	279.58 ± 5.51	260.09

are shown. In one case the packet length is sampled from the packet length distribution when the packet enters the network, and the packet retains this length as it traverses the links of the network. In the other set of simulations the packet length is sampled each time it is queued at a new link. Thus the latter results incorporate Kleinrock's independence assumption, and are included to show that the errors in the approximation under study are *not* due to the independence assumption, and that in any case there is enough mixing in these example networks that Kleinrock's independence assumption is quite good.

In Tables 1, 2 and 3, the packet load on link 2 is 0.1, 0.3, and 0.6, respectively. As packet load increases, the maximum circuit occupancy must, of course, decrease. We make the following observations from these tables:
(i) The maximum error is less than 10%.
(ii) For a particular packet load, as the circuit occupancy increases from a low value to a high value, the approximation overestimates at low circuit occupancies and underestimates at high circuit occupancies.
(iii) The circuit load at which the crossover occurs decreases as packet load on link 2 increases (≈ 0.65 when packet load is 0.1, ≈ 0.55 when the packet load is 0.3, and ≈ 0.325 when the packet load is 0.6).

Tables 4 and 5 show simulation results along with the corresponding results from the approximate analysis for fixed packet load, and various circuit loads for a 6-node linear network. The delays are shown at each link for packets that are entering the link from the left.

From Tables 4 and 5 we make the following observations:
(i) For each link, and each level of packet load, the approximation overestimates at light circuit loads and underestimates at circuit loads that drive the link close to saturation.

Table 2: Approximate mean packet delay and its corresponding simulation result for link 2 of a 3-node linear network with $\mu_c = 0.01$ and $\mu_p = 0.1$. Packet load on link 2 = 0.3

Circuit occupancy on Link 2	Mean delay on Link 2 With 95% confidence interval (Sampling of packet size done only once at network entry)	Mean delay on Link 2 with 95% confidence interval (Sampling of packet size done at each link)	Mean delay from approximate analysis
0.1	28.94 ± 0.25	28.67 ± 0.77	31.667
0.2	46.94 ± 1.25	46.52 ± 1.23	51.99
0.3	69.54 ± 1.83	69.41 ± 1.83	77.50
0.4	107.12 ± 3.08	106.77 ± 3.01	113.33
0.5	174.32 ± 5.57	173.96 ± 5.42	175.00
0.6	360.06 ± 8.13	364.09 ± 7.49	340.00

Table 3: Approximate mean packet delay and its corresponding simulation result for link 2 of a 3-node linear network with $\mu_c = 0.01$ and $\mu_p = 0.1$. Packet load on link 2 = 0.6

Circuit occupancy on Link 2	Mean delay on Link 2 With 95% confidence interval (Sampling of packet size done only once at network entry)	Mean delay on Link 2 with 95% confidence interval (Sampling of packet size done at each link)	Mean delay from approximate analysis
0.1	58.02 ± 2.03	57.95 ± 1.97	63.34
0.15	82.22 ± 1.42	83.82 ± 1.45	90.99
0.2	116.42 ± 4.45	117.47 ± 4.42	129.99
0.25	176.61 ± 4.29	181.64 ± 4.36	191.67
0.3	295.99 ± 9.01	302.94 ± 8.62	310.04
0.35	671.82 ± 11.38	676.21 ± 12.45	656.477

Table 4: Mean packet delays (moving from left to right) for various links of a 6-node linear network with $\mu_c = 0.01$ and $\mu_p = 0.1$. For link 3 the 95% confidence interval obtained from regenerative simulation is also shown. Packet load is as follows: link 1: 0.11, link 2: 0.18, link 3: 0.2, link 4: 0.18, link 5: 0.11

l	Circuit occupancy on link l.	Mean packet delay (simulation) (Sampling of packet size done only once at network entry.) at link l	Mean packet delay (simulation) (Sampling of packet size done at each link.) at link l	Approximate mean packet delay at link l
1	0.058	18.366	18.34	18.55
1	0.116	26.00	26.03	26.17
1	0.181	35.40	35.37	35.06
1	0.251	45.56	45.51	45.03
1	0.335	58.66	58.64	58.22
1	0.437	78.72	78.69	76.32
1	0.588	118.49	118.90	113.35
2	0.091	23.107	22.82	24.96
2	0.180	33.42	33.34	38.50
2	0.276	47.96	47.71	54.62
2	0.369	66.28	65.83	73.19
2	0.470	91.45	91.21	98.69
2	0.570	138.06	137.03	136.01
2	0.687	256.76	256.19	230.39
3	0.102	23.36 ± 0.62	22.56 ± 0.60	27.39
3	0.201	34.10 ± 0.60	33.19 ± 0.59	43.35
3	0.304	49.32 ± 0.73	48.42 ± 0.71	62.69
3	0.404	69.64 ± 0.91	68.74 ± 0.89	85.59
3	0.506	106.75 ± 1.16	105.26 ± 1.13	118.31
3	0.600	174.56 ± 1.96	173.64 ± 2.04	168.36
3	0.701	375.93 ± 5.38	377.09 ± 4.11	307.49
4	0.091	20.48	19.15	24.96
4	0.180	28.55	27.42	38.50
4	0.276	40.42	39.05	54.62
4	0.369	55.55	54.16	73.19
4	0.470	84.53	83.35	98.69
4	0.570	135.89	134.39	136.01
4	0.687	282.45	283.63	230.39
5	0.058	15.59	14.57	18.55
5	0.116	19.84	18.73	26.17
5	0.181	25.33	24.09	35.06
5	0.251	32.97	31.48	45.03
5	0.335	45.60	44.18	58.62
5	0.437	67.48	65.83	76.32
5	0.588	119.56	117.92	113.35

Table 5: Mean packet delays (moving from left to right) for various links of a 6-node linear network with $\mu_c = 0.01$ and $\mu_p = 0.1$. For link 3 the 95% confidence interval obtained from regenerative simulation is also shown. Packet load is as follows: link 1: 0.223, link 2: 0.356, link 3: 0.400, link 4: 0.356, link 5: 0.223.

l	Circuit occupancy on link l.	Mean packet delay (simulation) (Sampling of packet size done only once at network entry.) at link l	Mean packet delay (simulation) (Sampling of packet size done at each link.) at link l	Approximate mean packet delay at link l
1	0.0577	21.01	20.98	21.45
1	0.1160	30.45	30.36	30.63
1	0.1813	42.02	41.87	41.68
1	0.2507	55.60	55.66	54.65
1	0.3354	73.96	73.98	73.03
1	0.3788	87.65	87.80	84.09
2	0.0912	31.00	30.58	33.09
2	0.1804	47.55	47.26	53.45
2	0.2755	74.24	73.68	81.29
2	0.3692	115.64	115.35	121.12
2	0.4703	201.93	201.01	200.93
2	0.5158	293.14	291.56	272.92
3	0.1022	33.21 ± 0.875	32.47 ± 0.938	38.55
3	0.2009	52.77 ± 1.052	52.04 ± 1.063	65.38
3	0.3046	88.72 ± 1.737	88.37 ± 1.699	105.73
3	0.4039	155.33 ± 3.240	158.22 ± 3.480	174.22
3	0.5062	394.68 ± 10.31	399.73 ± 9.82	375.08
3	0.5500	789.35 ± 23.74	804.67 ± 23.49	702.25
4	0.0912	27.24	25.86	33.09
4	0.1804	39.97	39.36	53.45
4	0.2755	61.13	61.63	81.29
4	0.3692	96.26	99.55	121.12
4	0.4703	188.02	198.98	200.93
4	0.5158	280.503	299.63	272.92
5	0.0577	18.69	17.17	21.45
5	0.1160	23.29	22.16	30.63
5	0.1813	30.28	29.38	41.68
5	0.2507	40.42	40.04	54.65
5	0.3354	59.14	59.71	73.03
5	0.3788	72.18	73.10	84.09

$$\gamma_c, \frac{1}{\mu_c}$$
$$\gamma_p, \frac{1}{\mu_p}$$

Server 1   Server 2
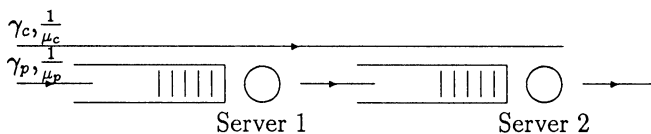
Figure 1: A two link integrated services model with two link circuits (Model-I)

$$\gamma_c, \frac{1}{\mu_c} \qquad \gamma_c, \frac{1}{\mu_c}$$
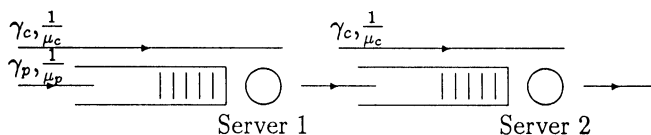$$\gamma_p, \frac{1}{\mu_p}$$

Server 1   Server 2

Figure 2: A two link integrated services model with single link circuits (Model-II)

(ii) On each link, the circuit load at which crossover occurs decreases as the packet load increases.

These two observations are the same as for Tables 1, 2 and 3. Further we observe:

(iii) There is load symmetry between links 2 and 4. For a given packet load, however, underestimation starts at a higher circuit load on link 4 than on link 2. As we will argue later (Section 5), this has to do with the fact that, as compared with link 2, more of the arrivals to link 4 are bursty departures from the previous link, rather than "fresh" Poisson arrivals. A similar observation holds for links 1 and 5.

We explain these observations in the next section.

5 A DETAILED STUDY OF THE APPROXIMATION

We begin by considering the two simple models shown in Figure 1 and Figure 2.

In both models, there are two single server queues in tandem. Packets entering the queue of Server 1, after finishing their service will enter the queue of Server 2. In the first model (Model I), circuits arriving will seize both servers simultaneously. In the second model (Model II), the servers will be seized by independent circuit arrivals and no arriving circuit can seize both servers. In other words, in the first model we allow only double link circuits, and in the second model we allow only single link circuits. The approximation we are studying in this paper will yield the same approximate model for link 2 in both models, i.e., the model (Model III) shown in Figure 3.
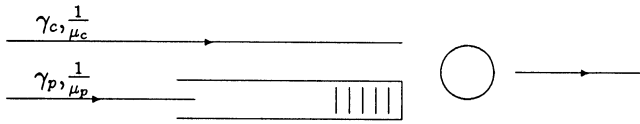
Figure 3: Approximate model for the second link of both models in Figures 1 and 2. (Model-III)

Table 6: Comparison of simulation estimates of mean delay experienced by packets at Server 2 of model-I and model-II with the results obtained using the approximate model for link 2 (model-III) with $\mu_c = 0.01$ and $\mu_p = 0.1$.

γ_c	γ_p	Delay on Link 2(simulation) (model-I)	Delay on Link 2(simulation) (model-II)	Mean Delay on Link 2(approximation) (model-III)
0.001	0.01	12.71± 0.21	22.78 ± 0.716	22.57
0.002	0.02	17.58± 0.29	39.63 ± 0.97	37.72
0.003	0.03	25.37± 0.85	66.13 ± 3.73	59.14
0.004	0.04	43.41± 1.6	116.1 ± 4.82	96.75
0.005	0.05	92.94± 5.515	237.61± 3.84	193.33
0.006	0.05	127.68±10.65	330.73 ± 6.50	267.5
0.007	0.05	174.05±21.46	491.79± 13.22	387.84
0.008	0.05	281.93±39.35	803.45 ± 39.49	624.44
0.009	0.05	507.58± 87.1	1629.28± 107.15	1327.37

In Table 6, we compare the simulation estimates of mean delay experienced by packets at Server 2 of model-I and model-II with the results obtained by analysing, model-III.

Let S_I and S_{II} be the random variables representing the stationary sojourn times of packets at Server 2 of model-I and model-II respectively. Let S_{III} be the random variable representing the stationary sojourn time of packets at the single server of model-III. Then these results suggest that

$$E(S_I) \leq E(S_{III}) \leq E(S_{II})$$

Note that the first link has the same stochastic behaviour in model-I and model-II. The observed inequality thus suggests that if there are *multiple link circuits* then the approximation tends to *overestimate* the packet delays, whereas if there are only *single link circuits* then the approximation tends to *underestimate* the packet delays at the second link of the 2 link model.

We now present some arguments and results that explain the observations.

Observe that the packet departure process from node 1, is the same in models I and II. Yet in model-I the mean packet delay at the second server is less than in model-II. It can be shown ([Babu, 1991]) that the interdeparture times of packets from the first link have a coefficient of variation greater than 1, and that the successive interdeparture times are correlated. Thus one would expect that the Poisson process approximation in model-III should yield smaller packet delays. Yet our observation is that the approximation yields larger packet delays than model-I.

Note, however, that the bursts in the packet departure process from node 1 occur because packets accumulate while a circuit holds the server, and are released in a burst when the circuit terminates. In model-I, this burst of packets will find the second server free of any circuit, whereas in model-II there is a probability that this burst of packets will again encounter a server occupied by a circuit.

Thus in model-I, there is a correlation between server availability and bursts in the arrival stream, which helps in reducing delay even below that in model-III. On the other hand the combined effect of a bursty packet arrival process and uncorrelated circuit arrivals makes the delay at node 2 in model - II exceed that in model - III.

It is interesting to note the similarity between our observations and arguments and those in [Sriram and Lucantoni, 1989]. In this paper the effect of bit dropping congestion control on the performance of a packet voice multiplexer is studied. It is observed that bit dropping during overload periods reduces the effect of correlation between packet interarrival times, thus causing a bursty arrival process to yield delays a Poisson process would yield.

It is, in fact, possible to create a mixture of 1 link and 2 link circuits with total circuit arrival rate γ_c at Server 2, such that the mean packet delay at the second node is close to that obtained via model-III.

The above arguments suggest that if the network is more likely to have multiple link circuits then the approximation will tend to overestimate the packet delays. Whereas if single link or "short" circuits are more likely then the approximation will tend to underestimate the packet delays.

It seems intuitively clear that, as circuit load increases, the length of active circuits will tend to become shorter. If this can indeed be established, then it would explain the fact that the approximation overestimates packet delays at light circuit loads and underestimates at high circuit loads. In fact, we have proved the following two theorems [Babu, 1991].

Theorem 1 *The mean number of active circuits in the network is an increasing function of the offered circuit switched load per route.*

Theorem 2 *If all single link circuits are routes, then the average length of the active circuits approaches 1 as the load goes to infinity.*

It may also be conjectured that the average length of active circuits is a decreasing function of the offered circuit load. In [Babu, 1991] we present numerical results to support this conjecture.

From the above analysis, we come to the conclusion that the length of active circuits tends to become shorter as circuit load increases. At low circuit loads, the isolated single link model overestimates the mean delay experienced by packets at downstream links, even though the packet traffic is bursty. This is because, the correlation between channel activity on neighbouring links actually overcomes the effect of burstiness and improves the queueing performance, since a burst of packets can traverse several links in the network as soon as a multiple link circuit releases its set of servers. But as circuit load increases, the effect of single link circuits becomes predominant. This means that the internode circuit correlation decreases. This combined with the fact that packet traffic is bursty, results in an underestimation by the approximation.

At low circuit loads, the overestimation at downstream links of the linear network is higher. This is because at downstream links there is more internodal bursty packet traffic than "fresh" Poisson packet traffic, and bursts in the servers' availability correlate with bursts of arrivals. If the arrivals are uncorrelated Poisson arrivals then the approximate model underestimates the packet delays. This explains the observation (iii) following Tables 4 and 5. At circuit loads that drive the links close to saturation, the packet departure process is very bursty. If the packet load is large then this situation would correspond to a relatively low circuit load; thus there will be a high probability of multiple link circuits, and in spite of the highly bursty packet traffic the approximation can underestimate.

6   NETWORKS WITH MULTIPLE CHANNELS PER LINK

In this section we apply the approximation to linear networks with multiple channels per link.

Calculation of link blocking probabilities directly from the product form distribution becomes extremely difficult as the size of the network increases and as the link capacity vector becomes large. Hence we resort to the Erlang fixed-point approximation procedure (outlined in [Kelly, 1986]).

We approximate each link of the network by an isolated link with the same number of circuit channels and packet channels and a Poisson circuit arrival process and a Poisson packet arrival process. Assuming that the link blocking probability calculated by the fixed point method [Kelly, 1986] is due to a Poisson circuit arrival process, we calculate the equivalent Poisson circuit arrival rate, using Erlang's formula. Further, given the packet arrival rate for each source-destination pair, and the routing matrix, we find the packet arrival rate at each link. This yields an approximate model for each link, whose analysis we have adopted from [Kraimeche and Schwartz, 1985].

We have done extensive simulation studies on a 6-node linear network in which each link has 4 full duplex channels, 3 circuit channels and 1 packet channel. We draw our conclusions by comparing delays experienced by packets at link 3 and their corresponding approximate estimates. We do this because link 3 experiences maximum blocking for circuits and maximum packet arrival rate.

As before it is found that the simulation that samples packet size at each

link (i.e., simulates Kleinrock's Independence Assumption) yields values that are quite close to those obtained when the packet size is sampled upon entry into the network. Hence, as before, the effect of the assumption of independent packet sizes is insignificant. The approximate mean packet delays are calculated in two ways. One uses the approximate link blocking probability obtained from the Erlang fixed-point method, and the other uses the "exact" link blocking probability obtained from the simulation. The latter values were obtained to study the behaviour of the basic idea of the approximation, i.e., replacing a link in the network with an "equivalent" isolated link. These latter results are the ones that are plotted in Figures 4, 5, 6 and 7.

In Figure 4 the packet load on link 3 is 0.7 which is less than the number of channels dedicated to packet service (i.e., 1 channel). This is referred to as the "underload" situation [Schwartz, 1987]. The packet queue is stable without help from the circuit channels, and the departure process from a link is close to Poisson. In fact, for light circuit loads the network of packet queues is close to a product form network, and hence the approximation is very accurate. As circuit load increases, however, the interaction between the circuits and packets at a link increases. While the packet departure process is close to Poisson, correlation in channel availability due to multiple link circuits reduces packet delay below that estimated by the approximation. Hence (see Figure 4) the approximation overestimates for high circuit loads. In Figures 5, 6 and 7 the packet load on link 3 is in the "overload" region. Packets queues now need circuit channels for stability and the packet departure process becomes burstier than Poisson as circuit load increases. The packet load goes deeper into the overload region from Figure 5 to 6 to 7; hence the packet departure process burstiness correspondingly increases. In order to keep the packet queues *stable*, however, the circuit load cannot be taken too high, hence, even at the maximum circuit load shown in these figures there will be multiple link circuits with a high probability. Thus the approximation overestimates in almost all cases except when the packet queues are close to instability and the packet departures are very bursty (see Figures 6 and 7).

When the approximation uses link blocking probability obtained from the Erlang fixed point method, errors in the estimates of link blocking probability seem to dominate the final result. Either the approximate mean packet delay is close to the simulation estimates, or is a considerable under estimation.

7 CONCLUSION

Our study, that began as an attempt to evaluate a simple approximation for packet delays, ended up revealing some very interesting characteristics of packet delays when packets are integrated with circuits in a *network* context. In the movable boundary scheme for sharing channels on a link between circuits and packets, it can be shown that the packet departure process is bursty. The coefficient of variation of packet interdeparture times is > 1 and the interdeparture times are correlated. Yet it is found that in some cases the Poisson assumption at a downstream link yields an *overestimate* of packet delays.

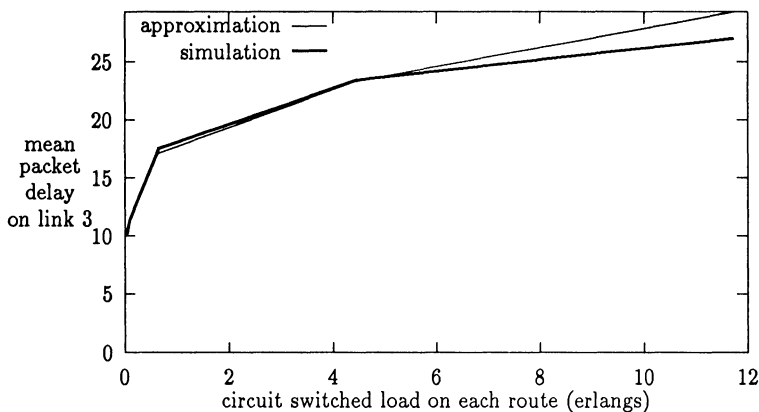This is explained by observing that in these cases the active circuits span

Figure 4: Mean packet delay on link 3 vs. circuit switched load on each route, in a 6-node linear network with 3 channels for circuits and 1 channel for packets on each link. Estimates from approximation are based on link blocking probability obtained from simulation. Packet load on link 3 = 0.7

multiple links, consequently when a circuit terminates it releases several channels simultaneously. Packets waiting for use of the freed circuit channels are now released in a burst but can flow through several links with low delay. In these cases the Poisson arrivals assumption overestimates packet delays. On the other hand, when active circuits are short then packets departing one link may again find channels on the next link occupied by circuits. Clearly this phenomenon is more significant at loads for which packets require the use of circuit channels (see Section 6). We also show that the average number of active circuits decreases as circuit load increases, and at heavy circuit loads the active circuit load approaches 1 (Section 5, Theorems 1 and 2). Thus the correlation effect is pronounced at low circuit loads leading to less than expected packet delays, and the correlation effect decreases with high circuit loads.

As mentioned in the Introduction, models similar to those discussed here will also arise in the B-ISDN/ATM context, if peak rate allocation is done for isochronous connections. Hence these models will continue to be of interest. Future work would include establishing the truth, or otherwise, of the conjecture in Section 5. Further, clearly the net arrival process of circuit switched calls to a link is not Poisson. The approximation may be improved by modelling these arrivals with a Markov modulated Poisson process.
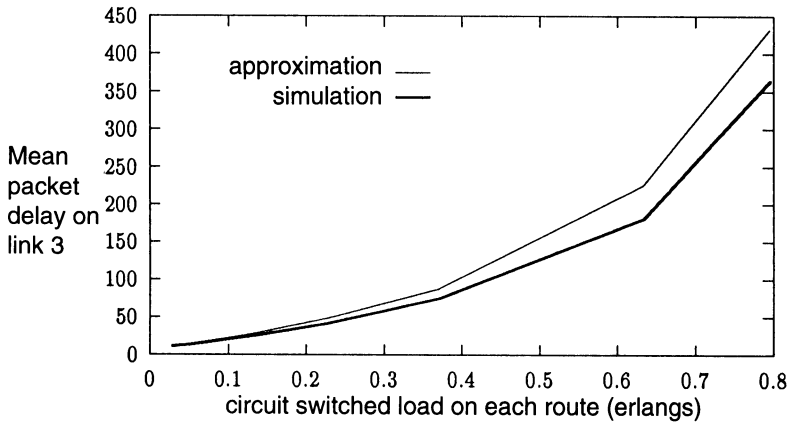
Figure 5: Mean packet delay on link 3 vs. circuit switched load on each route, in a 6-node linear network with 3 channels for circuits and 1 channel for packets on each link. Estimates from approximation are based on link blocking probability obtained from simulation. Packet load on link 3 = 1.5
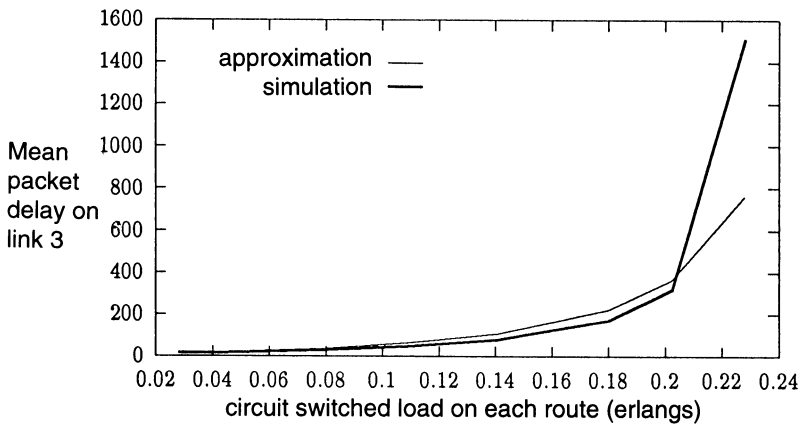

Figure 6: Mean packet delay on link 3 vs. circuit switched load on each route, in a 6-node linear network with 3 channels for circuits and 1 channel for packets on each link. Estimates from approximation are based on link blocking probability obtained from simulation. Packet load on link 3 = 2.0
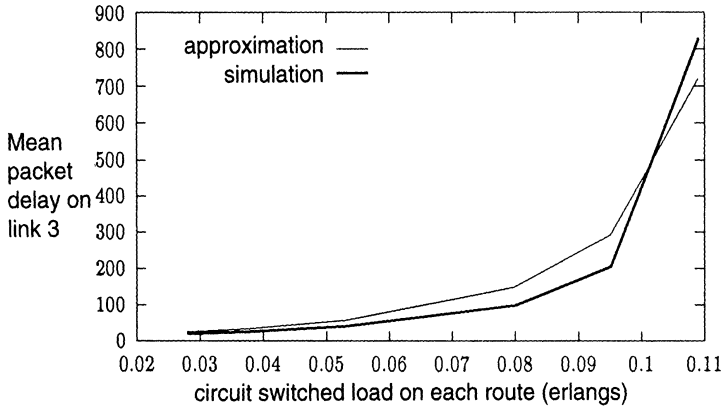
Figure 7: Mean packet delay on link 3 vs. circuit switched load on each route, in a 6-node linear network with 3 channels for circuits and 1 channel for packets on each link. Estimates from approximation are based on link blocking probability obtained from simulation. Packet load on link 3 = 2.5

REFERENCES

[Atmaca and Pujolle, 1992] T. Atmaca and G. Pujolle, "Performance Analysis of an Access Control Strategy in Integrated Networks", *Computer Network and ISDN Systems*, Vo.24, 1992, pp421-434.

[Bertsekas and Gallager, 1987] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall Inc., Englewood Cliffs,N.J, 1987.

[Burke, 1956] P.J.Burke, "Output of a Queueing System", *The Journal of the Operations Research Society of America*, 4, 1956, pp. 699-704.

[Burman et al, 1984] D.Y.Burman,J.P.Lehoczky and Y. Lim, "Insensitivity of Blocking Probabilities in a Circuit-Switching Network", *Journal of Applied Probability*, Vol.21, No.4, December 1984, pp. 850-859.

[Fischer and Harris, 1976] M.J.Fischer and T. C. Harris, "A Model for Evaluating the Performance of an Integrated Circuit- and Packet-Switched Multiplex Structure", *IEEE Transactions on Communications*, Vol.COM-24, No.2, February 1976, pp 195-202.

[Babu, 1991] T.V.J. Ganesh Babu, "Analysis of the End-to-end Performance of Integrated Services Networks", M.Sc(Engg) thesis, Indian Institute of Science, Bangalore, India, 1991.

[Leon-Garcia and Aboul-Magd, 1986] A.Leon-Garcia and O.S.Aboul-Magd, "Performance Analysis of Integrated Voice and Data Tandem Link Network", *Proc.IEEE Globecom*, 1986, pp. 944-948

[Gaver, 1962] D.P.Gaver,Jr, "Waiting Line with Interrupted Service Including Priorities", *Journal of Royal Statistical Soceity*, Vol.b24, 1962, p73-90.

[Kelly, 1986] F.P. Kelly, "Blocking Probabilities in Large Circuit-Switched Networks", *Adavances in Applied Probability*, Vol.18, No.2, June 1986, pp. 473-505.

[Kraimeche and Schwartz, 1984] B. Kraimeche and M. Schwartz, "Traffic Access Control Strategies in Digital Integrated Networks", *INFOCOM' 84, San Francisco*, CA, April 1984.

[Kraimeche and Schwartz, 1985] B. Kraimeche and M. Schwartz, "Analysis of Traffic Access Control Strategies in Integrated Services Networks", *IEEE Transactions on Communications*, Vol.Com-33, No.10, October 1985, pp. 1085-1093.

[Lavenberg, 1983] S.S.Lavenberg, Editor, *Computer Performance Modelling Handbook*, Academic Press INC, 1983.

[Li and Mark, 1985] S.Q.Li and J.W.Mark, "Performance Of Voice/Data Integration on a TDM System", *IEEE Transactions on Communications*, Vol.Com-33, No.12, December 1985, pp. 1265-1273.

[Li and Mark, 1988] S.Q.Li and J.W.Mark, "Simulation Study of a Network of Voice/Data Integrated TDM's", *IEEE Transactions on Communications*, Vol.36, No.1, January 1988, pp. 126-132.

[Rosenberg and Le Bon, 1993] C. Rosenberg and A. Le Bon, "Performance Models for Hybrid Broadband Networks", *Computer Networks and ISDN Systems*, Vol 25, 1993, pp1155-1163.

[Ross, 1983] S.M.Ross, *Stochastic Processes*, John Wiley & Sons INC, 1983.

[Schwartz, 1987] M.Schwartz, *Telecommunication Networks: Protocols, Modeling and Analysis*, Addison Wesley Publishing Company, 1987.

[Sriram et al, 1983] K.Sriram, P.K.Varshney and J.G.Shanthikumar, "Discrete-time Analysis of Integrated Voice/Data Multiplexers with and without Speech Activity Detectors", *IEEE Journal on Selected Areas in Communications*, Vol. SAC-1, No.6, December 1983, pp 1124-1132.

[Sriram and Lucantoni, 1989] K.Sriram and D.M.Lucantoni, "Traffic Smoothing Effects of Bit Dropping in a Packet Voice Mutiplexer", *IEEE Transactions on Communications*, Vol.37, No.7, July 1989, pp. 703-712.

[Weinstein et al, 1980] C.J.Weinstein,M.L.Malpass and M.J.Fischer, "Data Traffic Performance of an Integrated Circuit- and Packet-Switched Multiplex Structure", *IEEE Transactions on Communications*, Vol.COM-28, No.6, June 1980, pp 873-877.

[Williams and Garcia, 1984] G.F.Williams and A.L.Garcia, "Performance Analysis Of Integrated Voice and Data Hybrid-Switched Links", *IEEE Transactions on Communications*, Vol.COM-32, No.6, June 1984, pp. 532-539.

[Wolff, 1989] R.W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, N.J, 1989.

[Yang and Mark, 1991] O.W.Yang and J.W.Mark, "Queueing Analysis of an Integrated Services TDM System Using a Matrix-Analytic Method", *IEEE Journal on Selected Areas in Communications*, Vol.9, No.1, January 1991, pp. 88-94.

BIOGRAPHIES

T.V.J. Ganesh Babu has a B.E. from Madurai University, and a M.S. in Electrical Communication Engg. from the Indian Insititute of Science. He is currently pursuing a Ph.D. in Concordia University, Montreal, Canada.

Anurag Kumar has a B.Tech. in E.E. from the Indian Institute of Technology at Kanpur, and a PhD from Cornell University. He was a Member of Technical Staff at AT&T Bell Labs, Holmdel for over 6 years. Since 1988 he has been with the Indian Institute of Science (IISc), Bangalore, in the Dept. of Electrical Communication Engineering, where he is now Associate Professor. He is also the Coordinator at IISc of the Education and Research Network Project, which has set up a country-wide computer network for academic and research institutions, and conducts R&D in the area of communication networks. His own research interests are in the area of modelling, analysis, control and optimisation problems arising in communication networks and distributed systems.