An Evaluation of Routing and Admission Control Algorithms for Multimedia Traffic in Packet-Switched Networks *

Sanjeev Rampal, Douglas S. Reeves and Dharma P. Agrawal

Contact Author: Dr. Douglas Reeves
Department of Computer Science
Box 8206
North Carolina State University
Raleigh, NC 27695-8206, USA
Tel. (919)-515-2044
Fax. (919)-515-7896
email: reeves@eos.ncsu.edu

Abstract

Support for real-time multimedia applications is becoming an essential requirement for packet-switched networks. In this work, we investigated how routing and admission control influence the performance of packet-switched networks. We propose and evaluate the use of routing algorithms which take into account the constraints imposed by the admission control algorithms. Routing algorithms proposed elsewhere in literature are also evaluated. We show that the new algorithms decrease call blocking probability and increase network utilization. The amount of iprovement depends on such factors as the admission control function, the traffic mix, and the QOS constraints.

We evaluated two deterministic methods of Call Admission Control: Earliest Due Date and Stop&Go. We studied the effect of lossy source traffic shaping on routing. The interaction and relative importance of routing, admission control, and traffic shaping was also studied.

Technical Areas : Routing and control of congestion, admission and flow control

*This work was supported by the Air Force Office of Scientific Research Under Grant F49620-92-J-0441.

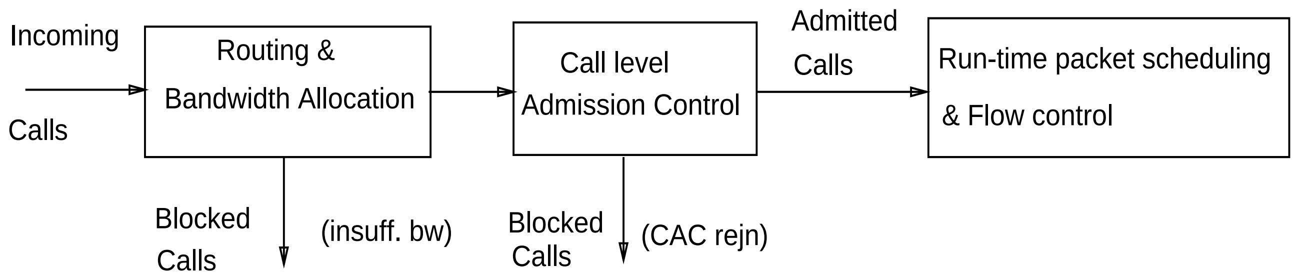

Figure 1: Interaction Between Different Traffic Management Functions

1 Introduction

Packet-switched data networks are increasingly being utilized to carry multimedia traffic. This trend is expected to continue with the deployment of broadband integrated services networks (BISDNs). Delay-sensitive real-time traffic such as voice and video, constitute an important component of such traffic. Such traffic has fairly stringent *quality-of-service* (QOS) requirements particularly in terms of end-to-end packet delay and packet loss. Table 1 shows the data rates and typical QOS requirements of standard traffic models for voice and video [16],[12].

In traditional "best-effort" delivery, the individual user is not given any quantitative network performance guarantees, and the goal is to optimize some aggregate network statistic, such as average network delay experienced by all calls [1]. With multimedia traffic, resources are resered to guarantee each individual call a specified end-to-end quality-of-service (QOS). A *call-level admission control* (CAC) procedure is used for this purpose. The router and CAC functions both play a role in determining network performance, as indicated in figure 1.

[1] We use the terms "call" and "channel" interchangeably throughout this paper

	Peak Rate	Mean Rate	Packet loss probability
Voice	32 KBits/sec	11.2 KBits/sec	.05
Video	11.6 MBits/sec	3.85 MBits/sec	10**(-5)

(a) Data rates of standard Voice, Video models & Typical acceptable loss probability

CCITT G.114 Delay Recommendations	
One-Way Delay	Characterization of Quality
0 to 150 ms	"acceptable for most user applications"
150 to 400 ms	"may impact some applications"
above 400 ms	"unacceptable for general network planning purposes"

(b) End-to-end delay requirements for multimedia traffic

Table 1: Typical data rates and QOS requirements for multimedia traffic

A router which was oblivious to principles of admission control might tend to select paths which have a high probability of being blocked by the admission control method. We investigated this thesis through a series of experiments. The emphasis of this paper is to examine this interaction between routing and call admission for multimedia traffic. Very little work appears to have been done on this problem to date. Some ideas on routing algorithms for real-time channels have been proposed by Parris and Ferrari [14], but not investigated quantitatively. Kompella et al [11] have investigated some related routing problems for multicasting of real-time data. Our investigations have yielded a number of useful insights.

- We show the improvement in call acceptance probability for multimedia traffic that can be obtained by the use of novel, simple to implement, routing algorithms. The improvement in performance depends on the choice of admission control algorithm, traffic mix and QOS requirements. The best performance is achieved by a least-cost routing algorithm, where the cost function models the probability of blocking due to the CAC function. Shortest path type routing algorithms were shown to yield better performance than algorithms similar to those used in circuit-switched networks.

- Our experiments indicate that the the admission control algorithm has a greater impact on overall performance than does the routing algorithm.

- We also evaluated two standard admission control schemes for real-time traffic (Earliest Due Date (EDD) [5] and Stop&Go [7]). Stop&Go resulted in better performance than EDD for the network and traffic models we studied.

- We investigated the use of traffic shaping (based on the leaky bucket method) and equivalent-bandwidth allocation. End-to-end delay requirements limit the improvements which are possible through their use. The relative goodness of the routing algorithms was not affected by them.

- The type and mixture of traffic strongly impacts the relative performance of the routing algorithms. This is due to the behavior of the call admission function under different traffic conditions.

The organization of the paper is as follows. In the next section we review the two admission control algorithms and traffic shaping function used in our investigation. In section 3, we discuss the routing algorithms which we have evaluated. Section 4 describes our simulation method and assumptions, and section 5 presents the results of our experiments. The last section summarizes our findings and suggests topics for future work.

2 Admission Control Schemes for Real-time Traffic

In this section we review two admission control schemes for real-time traffic: Earliest Due Date (EDD) [5] and Stop&Go [7]. Other methods of admission control are surveyed in [1]. These two methods were chosen because they are deterministic, which means that traffic losses in the network are eliminated (and thus don't have to be calculated, or simulated). We then describe the use of traffic shapers with such schemes to allow for improved utilization through selective packet loss.

2.1 Earliest Due Date

The Earliest Due Date (EDD) packet scheduling and admission control policy [5], [10], seeks to meet user specified end-to-end delay bounds by splitting these into local per-link delay bounds. Earliest Due Date packet scheduling is used at each node, based on these local "deadlines". If the K channels using a link n, are ordered in order of increasing values of their local delay bounds at this link (denoted as $d_{i,n}s, i = 1, \dots K$ with $d_{1,n} \leq d_{2,n} \leq \cdots d_{K,n}$), then the following necessary condition must be satisfied,

$$T_{min,i} \geq \sum_{j=1}^{K} Tx_{j,n}, \qquad (i = 1, \cdots K) \tag{1}$$

where $T_{min,i}$ is the minimum inter packet arrival time for channel i and $Tx_{j,n}$ is the transmission time of one packet of channel j on link n. All channels always meet their local "deadlines" at this link if and only if

$$d_{i,n} \geq \sum_{j=1}^{i} Tx_{j,n} + T \qquad (i = 1, \cdots |U|) \tag{2}$$

where T is the maximum packet transmission time of any channel using the link and U denotes the set

$$U = \{i | i = 1, \cdots |U|; d_{i,n} < \sum_{j=1}^{K} Tx_{j,n}\} \tag{3}$$

2.2 Stop&Go

The Stop&Go service and admission control policy [7], defines a framing strategy in which each real-time channel is assigned bandwidth at a certain frame level. The policy requires the source to obey *(r, T) smoothness* where T is the frame interval duration (The source is required to generate no more than r units of data during time frames of length T).

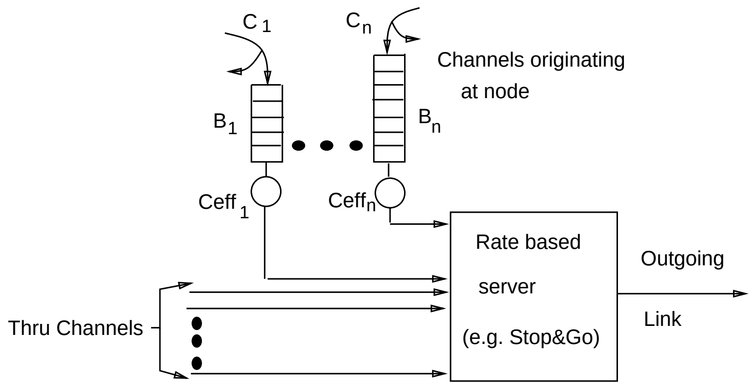

Figure 2: Use of Traffic Shapers with Rate Based Servers (Single output link of a node)

A necessary condition to be satisfied by channels sharing a link is given by the following inequality.

$$-C_n^{g0} + \sum_{g=g0}^{G} C_n^g (1 + \lceil T_{g0}/T_g \rceil) T_g / T_{g0} \leq \begin{cases} C_n - ,_{max}/T_{g0}, & g0 = 2, \cdots, G \\ C_n, & g0 = 1 \end{cases} \quad (4)$$

Here C_n^g is the capacity assigned to frame level g at link n, T_g is the duration of a level g frame, , $_{max}$ is the maximum packetsize over all channels and G is the total number of levels. If the ratios of the frame times are integers, instead of equation 4, a simple capacity relation is all that needs to be satisfied [7].

2.3 Equivalent Bandwidth based Allocation

The equivalent bandwidth (sometimes also referred to as effective bandwidth) of a source represents the service rate that must be provided to it to guarantee its required QOS [8], [4]. For a Markov modulated fluid source characterized by (M, λ) where M is the infinitesimal generator of the controlling markov chain, λ is the generation rate vector, allowable loss probability is p and B is the buffer size, for small p and large B values, the effective bandwidth of the source is given by the maximal real eigenvalue of the matrix $[\Lambda - (B/\log p)M]$, where $\Lambda = diag(\lambda)$ [4].

We investigated the use of the two rate based flow control schemes along with traffic shaping at the inputs. Figure 2 shows the model of one output link of a switch in such a scheme. The simple leaky bucket [13] is used for traffic shaping (packets queue but not tokens). The token generation rate of a leaky bucket is set to the equivalent rate of the source required for a given size of the shaper buffer and given allowable loss. The shaper buffer size is dictated by allowable delay considerations as we explain later. The equivalent bandwidth is then used by the admission control policy. We note that the traffic shaper introduces an additional delay which must be subtracted from the specified end-to-end delay bound to yield the allowable network delay.

3 Routing methods

The routing problem has been investigated extensively for both circuit-switched and packet-switched networks. We concentrate on virtual circuit based routing since all resource allocation techniques are designed around the assumption of a single path for a channel.

We developed and evaluated some shortest path type algorithms which are modified to account for the constraints of the CAC policy. For our studies, end-to-end delay was the main QOS metric, while packet loss was handled using traffic shaping as discussed earlier. Link bandwidth was the main network resource sought to be optimized; we assumed flow control would make buffer space a less important consideration. Hwang et al [9] have suggested the use of some circuit-switching routing techniques for high speed multimedia networks; we also evaluated two of thos algorithms.

3.1 Shortest Path Type Routing Algorithms

For these algorithms, for each new channel admission request a reduced network graph is formed. This is done by eliminating links at which the necessary constraint of the CAC is not satisfied (equations 1 and 4 respectively in case of the EDD and Stop & Go servers). The minimum delay bound that can be "offered" to this new channel without causing any existing channels to miss their delay bounds can be calculated for each link, using procedures outlined for these server policies. The different routing algorithms discussed below formulate different link length and/or cost measures and execute some form of shortest path calculation on the reduced network graph.

The Shortest Path router (SP) This algorithm seeks to meet the QOS requirement (end-to-end delay in our case) in a greedy manner. The minimum delay bounds in the reduced network graph are used as the link lengths in a shortest path calculation. The motivation for this approach is that a path with minimal delays corresponds to a lightly loaded path, and hence load balancing is also achieved. The SP algorithm was first proposed in [14]. However, no quantitative evaluation was presented.

The Shortest Cost router (SC) An algorithm which seeks to meet end-to-end delays may not be suited for a situation where delays are not critical. For instance, packet transmission times drop in high speed networks and it may not be very difficult to meet end-to-end delay requirements. The constraints imposed by the flow control policy may become the dominant reasons for call blocking. From equation 1 it is clear that the probability of a link getting blocked increases as the values of the r.h.s approaches that of the l.h.s.. Hence for an EDD based scheme, we use the cost function

$$EDD : Cost_n = 1/(T_{min,n} - \sum_{j=1}^{K} Tx_j) \tag{5}$$

where $T_{min,n}$ is the minimum of peak packet inter-arrival times over all channels using link n (including the incoming channel). Similarly, from equation 4, it is seen that the blocking probability at a link increases as the quantity on the l.h.s. approaches that on the r.h.s. i.e. the link capacity. The cost formulation used is hence

$$SAG : Cost_n = C_n - (-C_n^{g0} + \sum_{g=g0}^{G} C_n^g(1 + \lceil T_{g0}/T_g \rceil)T_g/T_{g0}) \tag{6}$$

The SC algorithm uses these link costs as link lengths in the reduced network graph and then performs a shortest path calaulation. No consideration is given to meeting the end-to-end delay bound at all.

Other Routing Algorithms The **Minimum Hop with Delay Bounds (MHDB)** algorithm finds a minimum hop path which also satisfies the end-to-end delay constraints. (this can be done using the Bellman-Ford algorithm [14]). To accomodate delays as well as scheduling costs, we would like to have a routing algorithm which finds a path with minimal delay as well as minimal total cost. Unfortunately this problem is NP-complete [6]. Instead an approximation called the **Min Max Cost with Delay Bound (MMCDB)** algorithm was developed. This algorithm finds a path which has minimum value of the maximum cost of any link in the path, while also satisfying the end-to-end delay bound. Details are withheld for lack of space. To investigate the performance of an algorithm which seeks to perform load balancing, the **Lowest Maximum Utilization (LMU)** algorithm was evaluated. This algorithm simply chooses a path with minimal value of maximal utilization of any link in the path.

Conventional Routing Algorithms C_STAT and **C_DYN**, two conventional routing algorithms were also evaluated in order to compare the performance of the real-time routers. The C_STAT algorithm is a static router which selects the preset minimum hop path for a channel. The C_DYN algorithm is a dynamic router in which link utilizations are used as the link lengths.

3.2 Algorithms based on Circuit-switched network routing

Hwang et al [9] propose the use of routers similar to those used in circuit switched networks for routing in integrated services networks. Rather than perform a shortest path calculation, these algorithms send the request message outwards from the origin node. The path traversed by the request message till it reaches the destination is chosen as the selected path. In the **Sequential Office Control (SOC)** algorithm, the call is blocked if all the outgoing links from a node are blocked, while the **Crankback (CB)** algorithm allows backtracking. Details are given in [9].

4 Simulation Model and Experimental Method

Simulation techniques were used to evaluate the routing algorithms. The simulations were performed in a static environment with the assumption of infinite call holding times and negligible call processing times. The figure of merit used to evaluate the routing algorithms was call acceptance probability[2] (conversely, the call blocking probability) This was estimated by counting the number of accepted channels in an appropriately sized 'sliding window'. Confidence intervals were calculated to ensure the validity of the results.

Each time a real-time channel was blocked, the reason for being blocked was recorded. An analysis of the relative fraction of the causes of blocking is useful in understanding the performance of a routing algorithm. A channel can be blocked for one of three reasons:

1. <u>No Path Available</u> No path exists between the source and destination with sufficient residual bandwidth.
2. <u>Violation of Server Necessary Condition</u> No path exists for which the necessary condition imposed by the CAC function is met at all links along the path.
3. <u>Admission Control Rejection</u> The best QOS (in our case, end-to-end delay) that the network can offer to the new channel violates the user-specified bound.

Figure 3 shows the topology of the network we used (first used by [5] to investigate the EDD admission control method). Packet sizes were uniformly set to the ATM cell size of 53 bytes. Link bandwidths in different experiments were chosen variously as 1.5 Mb/s, 45 Mb/s and 155 Mb/s. The sum of propagation and processing delay at each link was chosen to be uniformly one millisecond. Two standard markov modulated fluid models were chosen for source models of voice and video traffic (from [16] and [12] respectively). Figure 4 shows the model parameters used. Allowable packet loss was taken as specified in Table 1 and allowable end-to-end delay was taken to be 350 ms. For the simulation results reported, the set of frame sizes used by the the Stop&Go scheme was 2.5ms, 5ms, 20ms, 50ms and 150ms.

5 Simulation Results

In this section we present the results of our simulations. The performance of the routing algorithms is investigated under different admission control policies and traffic mixes.

[2] A different figure of merit which models the effect of accepting a call on *expected revenue* was also used. The conclusions from those experiments were essentially identical to those reported here.

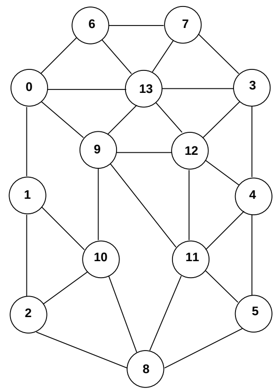

Figure 3: Graph of network used for simulations (from [5]).

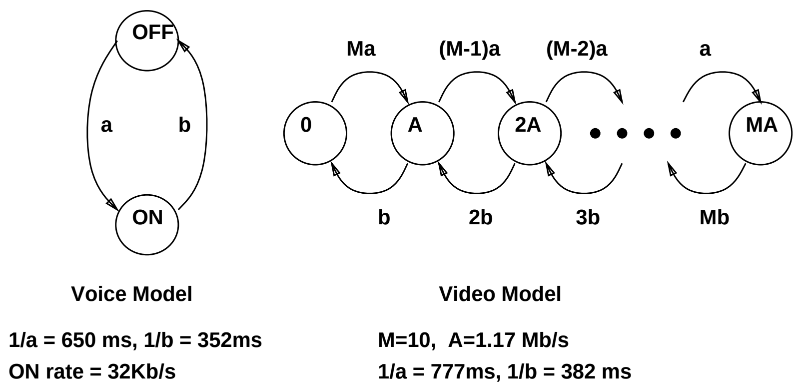

Figure 4: Traffic Models Used for Voice and Video

5.1 Performance of Routing Algorithms with Homogeneous Traffic

We first examine the call acceptance probabilities of the different routing algorithms with peak rate based bandwidth allocation and homogeneous traffic (i.e. only voice or only video channels). Figure 5 and Figure 6 plot these as a function of accepted load for the EDD and Stop&Go servers respectively, for voice-only traffic at T1 link speeds. For most of the experiments, the performance of the LMU, MMCDB and SOC algorithms was found to be consistently poorer than that of the other dynamic algorithms. Also, the performance of the MHDB algorithm was found to closely follow that of the SP algorithm. Hence we do not plot the curves corresponding to these algorithms in this paper. For all plots, the maximum magnitude of the 95% confidence intervals (omitted to retain clarity) was within 5-6% of the values plotted.

The performance of all the dynamic routing algorithms is seen to be fairly similar. For the EDD case, the Shortest Cost and Shortest Path algorithms perform the best, accepting about 10-15% more load than the C_DYN algorithm at acceptance probabilities better than 80%. The difference in performance is eliminated as the call acceptance probability drops below 60%. However, it is not likely that the

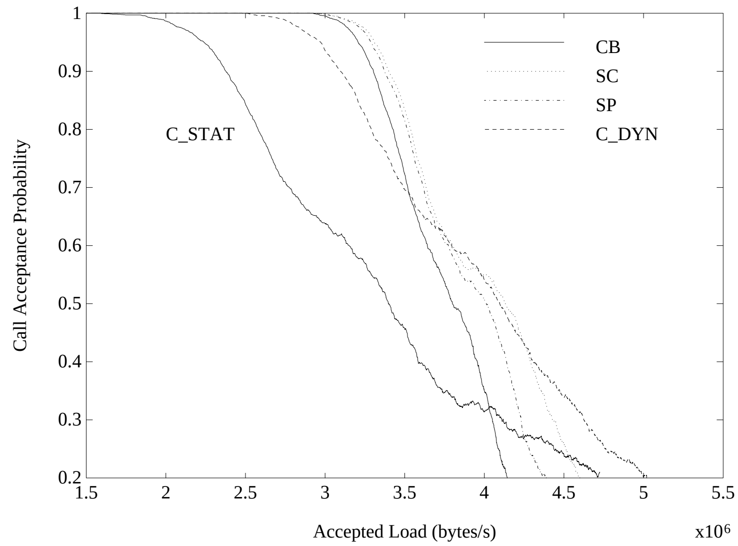

Figure 5: Call acceptance probabilities for homogeneous traffic (voice only), EDD server, 1.5 Mb/s link speeds

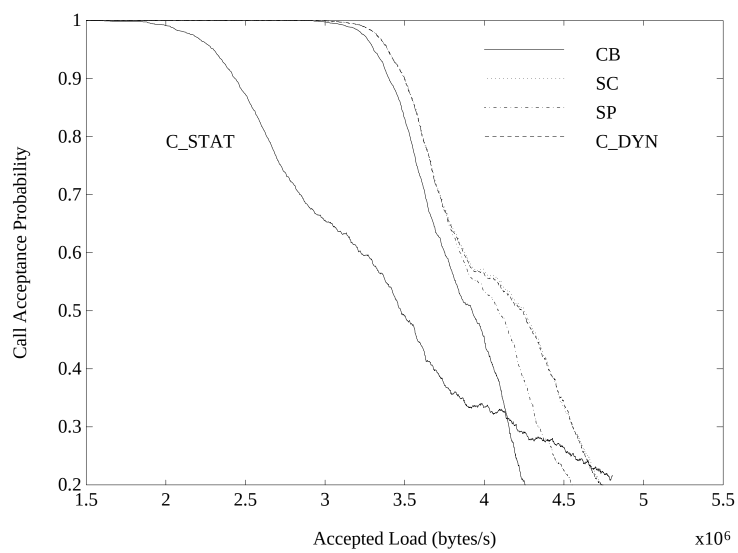

Figure 6: Call acceptance probabilities for homogeneous traffic (voice only), Stop&Go server, 1.5 Mb/s link speeds

	LMU	MMCDB	MHDB	SOC	CB	SC	SP	C_DYN	C_STAT
Avg. Path Lengths	3.01	2.72	2.08	2.19	2.27	2.02	2.08	1.84	1.77
Peak Utilizations (%)	95.4	94.6	89.2	91.9	92.7	88.5	89.3	82.3	70.7

Table 2: Average path lengths of accepted channels, and peak link utilizations observed, averaged over all network links (EDD , voice only, at 20% acceptance probability)

network will be operated at acceptance levels less than 50%. The performance of the best real-time routing algorithm is identical to that of the C_DYN algorithm in the Stop&Go case. The improved performance of the conventional algorithms at very low acceptance probabilities is due to their more conservative use of network resources initially. Qualitatively similar results were observed for the case of homogeneous traffic with only video channels and with higher link speeds.

It was also observed that the two circuit-switched routing algorithms do not have the best performance (the performance of the SOC was consistently poorer than that of the Crankback algorithm and is not shown here). This was observed to be consistently true even for other experiments.

Table 2 lists the average path lengths and peak link utilizations respectively for the different routing algorithms. It is clear that algorithms that select shorter paths perform better overall.

5.2 Performance of Routing Algorithms under Heterogeneous Traffic

Figure 7 and Figure 8 show the call acceptance probability plots for the EDD and Stop&Go servers respectively for the case of heterogeneous traffic. Link speeds are 45 Mb/s, and 20% of the total incoming calls are video channels while the other 80% are voice channels. The oscillatory nature of the acceptance curves reflects the fact that at first only video calls experience blocking, while voice calls start getting blocked at higher loads. (Hence the flat portion of the curve corresponds approximately to the fraction of voice calls in incoming traffic).

The acceptance probabilities show a very different router behaviour for the EDD case. At intermediate loads, the C_DYN algorithm outperforms all real-time routing algorithms except the two cost-based routing algorithms, viz. the MMCDB (not shown here) and SC algorithms. The performance of the SC router is significantly better than all other routing algorithms at all loads. At 75% call acceptance probability, the SC router accepts about 80% more load than the C_DYN and SP algorithms. For the Stop&Go case, routing performance is qualitatively very similar to that seen for the homogeneous traffic case.

It was observed that most of the EDD call rejections were due to violation of the necessary condition of the server (i.e. condition (2) from section 4) and no rejections occured due to inability to meet end-to-

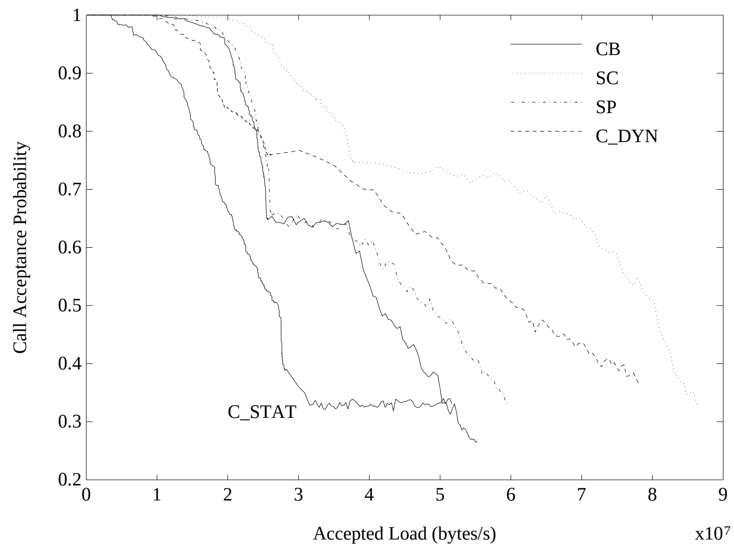

Figure 7: Call acceptance probabilities for heterogeneous traffic (20% video), 45 Mbits/sec link speeds, EDD server

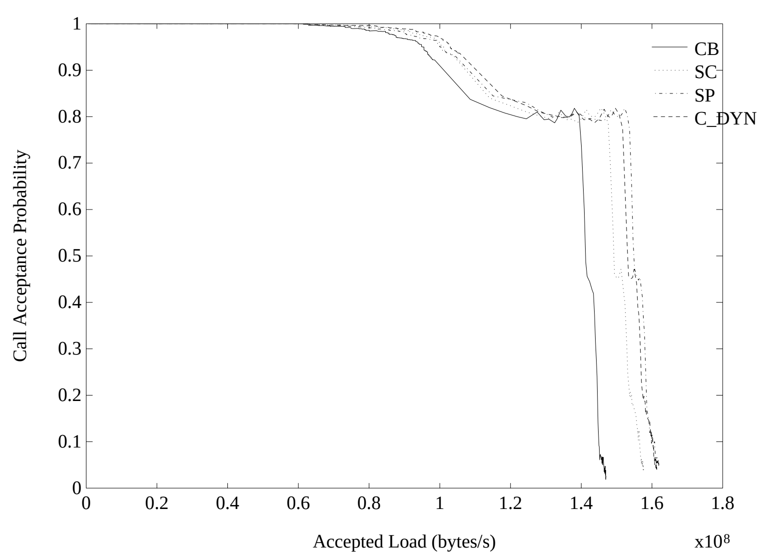

Figure 8: Call acceptance probabilities for heterogeneous traffic (20% video), 45 Mbits/sec link speeds, Stop&Go server

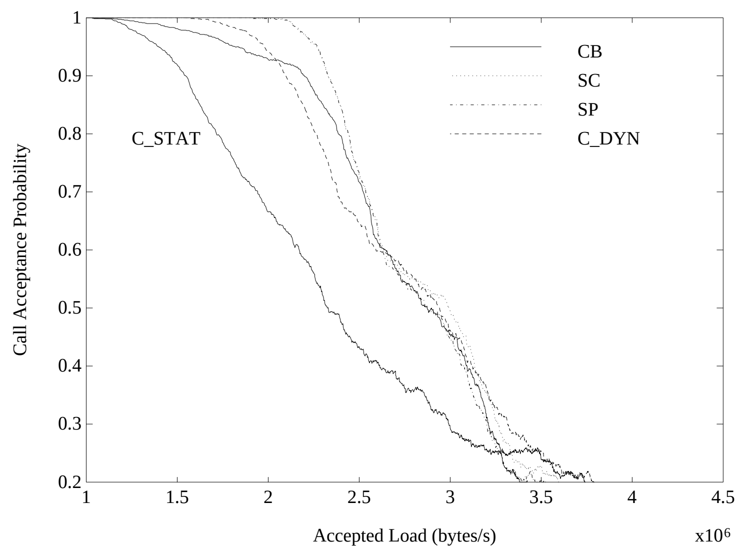

Figure 9: Comparison of routing algorithms, Stop& Go server with irregular frame sizes, voice only traffic, 1.5 Mb/s link speeds

end delay bounds. From the necessary condition for the EDD policy, (i.e. equation 1) it can be seen that when the $T_{min,i}$ values for all channels are similar (i.e. homogeneous traffic), the necessary condition reduces to a simple constraint on utilization. Since the dynamic routing algorithm C_DYN is utilization-based, and meeting end-to-end delays is not a problem, it performs quite well. With heterogeneous traffic, however, call admission is determined by the channel with minimum value of T_{min}, from among all channels sharing the link. The CAC function can block calls even when link utilization is low. The performance of the cost based routing algorithms is particularly good since these are formulated to satisfy the necessary constraint of the CAC policy. In the case of Stop&Go, however, even with heterogeneous traffic the server constraint (equation 4) remains essentially a utilization requirement; hence, the real-time routing algorithms do not perform significantly better than the C_DYN algorithm. For these experiments, the Stop&Go frame sizes were chosen to be interger multiples of each other. In this case (as mentioned earlier) the server constraint is simply a utilization constraint.

We would expect to see some difference if the frame sizes were chosen in a way that equation 4 differs significantly from being just a capacity constraint. Figure 9 plots the results of an experiment in which the frame sizes were chosen to be 150ms, 100ms, 70ms, 40ms, 25ms. The results agree with this expectation.

5.3 Effect of Traffic Shaping

Figure 10 is a plot of the equivalent bandwidth of our voice and video source models, versus the delay introduced by the the shaper, for two different loss rates.

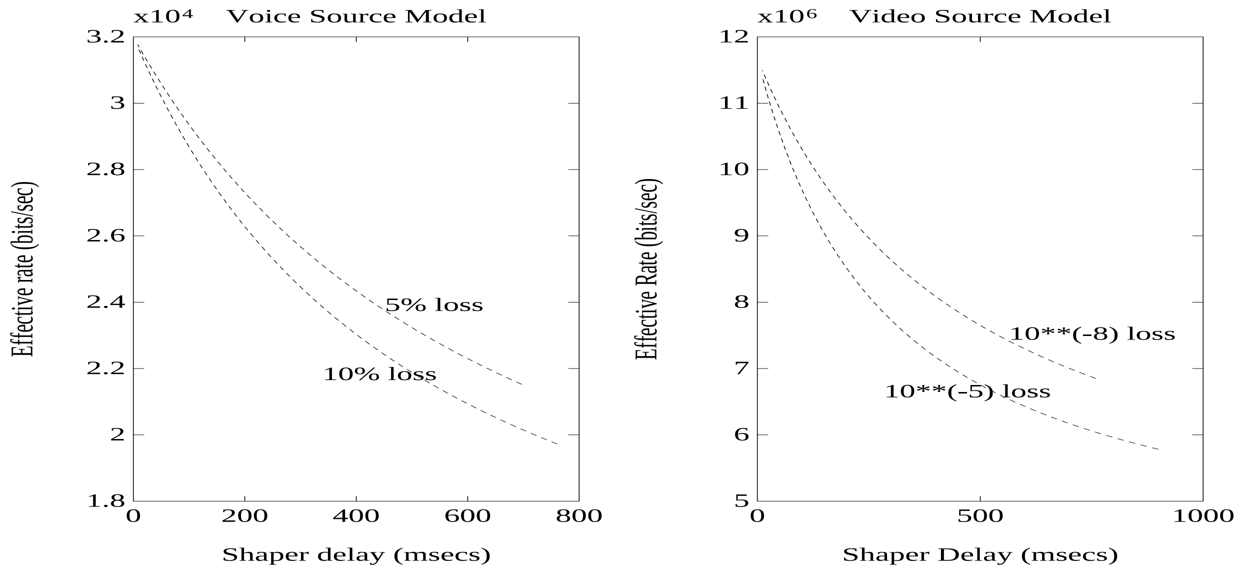

Figure 10: Effective Bandwidth versus shaper-introduced delay for voice and video source models (Voice model: Peak rate 32 Kb/s, Mean rate 11.24 Kb/s, Video model: Peak rate 11.7 Mb/s, Mean rate 3.85 Mb/s)

50 ms. was found to be sufficient for queueing and propagation delays in our experiments. This would permit the shaper to introduce as much as 300 ms of delay and still meet the delay targets. Figure 10 shows a reduction of 20% for the equivalent bandwidth of a voice source for this shaper delay. The reduction of a video source is approximately 35%.

The performance of the Stop&Go scheme with such shapers was investigated. As seen from a comparison of Figure 6 and Figure 11, all the routing algorithms accepted about 25% more load using equivalent bandwidths. However, a comparison of the two figures shows very little change in the relative performance of the routing algorithms. Similar results were observed for the case of video traffic.

5.4 Comparison of Admission Control Algorithms

These experiments enabled us to evaluate the relative performance of EDD and Stop&Go. Stop&Go was found to yield better performance for typical multimedia traffic and network models. The chief advantage of the EDD discipline is greater flexibility in accepting calls with stringent end-to-end deadlines. This did not occur in the situations we modelled, for which the necessary constraint of the CAC policy was the dominant reason for call blocking. The EDD policy was shown to perform poorly in the presence of a heterogeneous traffic mix; the Stop&Go policy was much less affected by the traffic mix. Note that we did not implement the full EDD scheme, because of its exponential worst case time complexity [5]. The full EDD scheme will undoubtably perform significantly better.

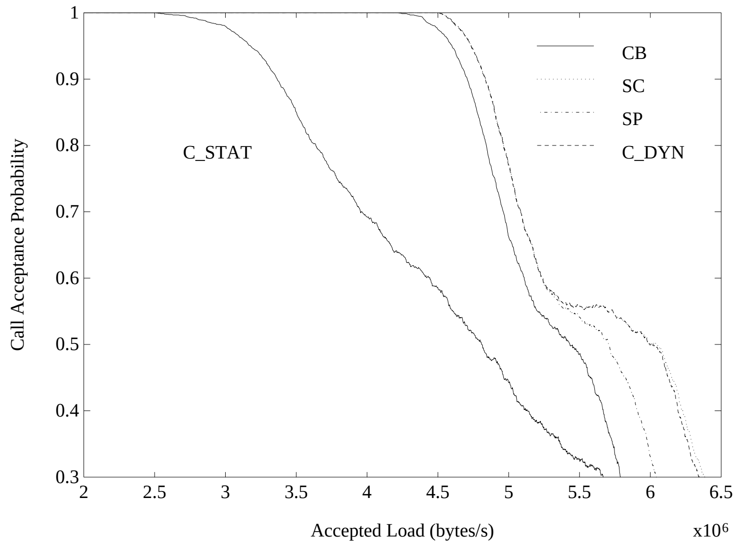

Figure 11: Call acceptance probabilities for homogeneous traffic (voice only), Stop&Go server with Traffic Shaping, 1.5 Mb/s link speeds

Link utilisation	EDD Server	Stop&Go Server	Stop&Go w/ Effective BWs
Peak/effective	75.9%	78.2%	80.7%
Average	26.6%	27.4%	39.1%

Table 3: Peak/ Effective & Average Link Utilizations for Shortest Costs router, Voice only traffic, at 50% call acceptance probability

Table 3 link utilizations achieved by the Shortest Cost router for the two CAC policies with peak rate based allocation and that for Stop&Go with effective bandwidth based allocation, at 50% call acceptance levels.

6 Conclusions & Future Work

We investigated the performance of several routing algorithms and two standard admission control policies for real-time traffic in packet-switched networks. Several routing algorithms were designed to account for real-time constraints and hence improve acceptance rates. Routing algorithms proposed elsewhere in literature were also evaluated. We modeled a simple packet-switched network, used standard traffic models of voice and video sources. The algorithms were evaluated with respect to call blocking probability and accepted load. We investigated only algorithms which were practical to implement. The two admission control policies which we studied bound the packet loss rates deterministically.

The dominant cause for blocking of real-time channels was not the attainable quality-of-service (viz. end-to-end delay), but rather the constraints placed on per call traffic by the CAC function. When these constraints become severe, real-time routing algorithms can improve performance significantly. Among the real-time routing algorithms we studied, the best performance was attained by minimizing (heuristically) the blocking probability due to the CAC function. For the EDD-based policy under heterogeneous traffic conditions, this routing algorithm was shown to result in 20% to 80% more accepted load. Overall, real-time routing algorithms yielded only modest performance improvements with the Stop&Go admission control policy. Shortest-path routing methods performed better than circuit-switched routing methods. We also investigated the use of traffic shaping, using a simple leaky bucket. This resulted in about 25% to 30% better link utilizations. The relative performance of the routing algorithms was qualitatively similar to the peak rate case.

Problems for future research include analysis of the dynamics of the routing algorithms, analysis of their performance with other flow control and CAC policies.

References

[1] C.M. Aras, J. Kurose, D. S. Reeves, and H. Schulzrinne, "Real-time Communication in Packet-switched Networks," to appear in the *Proceedings of the IEEE.*

[2] D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, New Jersey:Prentice-Hall, 2nd Ed 1992.

[3] D.D. Clark, S. Shenker and L. Zhang, "Supporting real-time applications in an integrated services packet network: architecture and mechanism," *Proc. SIGCOMM '92*, Aug.1992, pp.14-26.

[4] A.I. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE/ACM Trans. Networking*, Vol.1, No.3, June 1993, pp.329-343.

[5] D. Ferrari and D.C. Verma, "A Scheme for Real-time Channel Establishment in Wide-area Networks," *IEEE Journal on Selected Areas in Communications*, Vol.8, No.3, April 1990, pp.368-379.

[6] M. Garey and D. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman, 1979.

[7] S.J. Golestaani, "A Framing Strategy for Congestion Management," *IEEE Journal on Selected Areas in Communications*, Vol.9, No.7, Sep 1991, pp.1064-1077.

[8] R. Guerin, H. Ahmadi and M. Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks," *IEEE Journal on Selected Areas in Communications*, Vol.9, No.7, Sep 1991, pp. 968-981.

[9] R-H. Hwang, J.F. Kurose and D. Towsley, "The effect of Processing Delay and QOS Requirements in High Speed Networks," *Proc. IEEE INFOCOM '92*, pp. 160-162.

[10] D.D. Kandlur, K.G. Shin and D. Ferrari, "Real-time Communication in Multi-hop Networks," *Proc. IEEE Int'l Conf. on Distributed Computing*, 1991, pp.300-307.

[11] V. Kompella, J. Pasquale and G. Polyzos, "Multicast Routing for Multimedia Communications," *ACM/ IEEE Transactions on Networking*, vol.1, No.3, June 1993, pp.286-292.

[12] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. Robbins,"Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Trans. Communications*, vol.36, no.7, July 1988, pp.834-844.

[13] C. Partridge, *Gigabit Networking*, Addison-Wesley, 1993.

[14] C.J. Paris and D. Ferrari, "A Dynamic Connection Management Scheme for Guaranteed Performance Services in Packet-Switching Integrated Services Networks," Tenet Technical Report TR-93-005, Computer Science Division, University of California at Berkeley, 1993.

[15] A.K. Parekh and R.G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks-the multiple node case," *Proc. IEEE INFOCOM '93.* pp.521-530.

[16] D. Yates, J.F. Kurose, D. Towsley and M. Hluchyj, "On per-session end-to-end delay distributions and the call-admission problem for real-time applications with QOS requirements," *Technical Report*, Dept. of Computer Science, Univ. of Massachusetts, Amherst, 1993.