DYNAMIC ROUTING ALGORITHMS IN VP-BASED ATM NETWORKS

Hon-Wai Chu and Danny H. K. Tsang

Department of Electrical & Electronic Engineering
The Hong Kong University of Science & Technology,
Clear Water Bay, Kowloon, Hong Kong

Abstract — We consider a VP-based ATM network supporting multiple traffic classes with different traffic characteristics but with identical end-to-end quality of service (QoS) requirement. The concept of effective bandwidth is used to determine the required bandwidth to guarantee the QoS requirement of each traffic class. We study the problem of using dynamic routing to VP-based ATM networks by transforming it into an equivalent multi-rate circuit-switched network problem. To further simplify the analysis, we restrict the choice of path to single-link and two-link routes. We propose a dynamic routing algorithm based on the Least Loaded Routing (LLR) with packing. Simulation results are used to compare the performance of this algorithm with other dynamic routing schemes such as LLR.

I. INTRODUCTION

We consider the problem of routing in Virtual Path (VP) based ATM networks. The network selects a path for the new connection, i.e. find a set of VPs to connect the source to the destination. Most of the previous works on routing in VP networks are restricted to homogeneous traffic class. That is, only a single traffic class is supported in each VP network. In [4], LLR-based routing algorithms for VP networks supporting homogeneous traffic class with one quality of service (QoS) requirement are proposed. In [5][6], the study is extended to the case in which each VP can support two QoS requirements and MDP-based routing algorithms are also developed.

In this paper, we propose a dynamic routing algorithm that can handle heterogeneous traffic and is implementable in real-time. In Section II, we introduce the network model to be considered for VP-based ATM networks. We then discuss the routing problem for such VP networks in Section III. In Section IV, we discuss several previously proposed real-time dynamic routing schemes and also propose our own dynamic routing scheme. Simulation results are presented in Section V to illustrate the relative performance of the different routing algorithms. In Section VI, we draw conclusion in this study.

II. NETWORK MODEL

Virtual Path (VP) concept is introduced to simplify traffic control and routing. We assume that the ATM network consists of several VP subnetworks, for which nodes are interconnected by VPs. Let us consider a 4-node VP subnetwork as shown in Fig. 1. To establish a VC between the switches a and b , there are four possible routes: $\{1\}$, $\{5, 2\}$, $\{4, 6\}$ and $\{4, 3, 2\}$. If a

route consists of one VP, it is a direct route; otherwise, it is an alternative route. Following the practice of dynamic routing in telephone networks, we exclude routes that consist of more than two VPs, i.e., $\{4, 3, 2\}$ in this example. Hence, the legitimate routes are $\{1\}$, $\{5, 2\}$ and $\{4, 6\}$ only. The heuristic idea is that it wastes too much network resources to route call over a path with more than two VPs. It is better off to reject the call so that the spare resources can be used to support other users. Furthermore, Krishnan [9] shows that restricting to 2-link routes imposes no penalty on network cost in broadband ATM networks. Consequently, we shall consider only 1-link direct VP and 2-link alternative VP in our routing decision. We shall use the words VP and link interchangeably in this paper.

Note that call admission control and routing are coupled problems since the network accepts a call only if it can find a suitable path. The bandwidth requirement for a call is first calculated at call setup with knowledge of the traffic characteristics, such as average bit rate, peak bit rate, and mean burst length. This predetermined bandwidth commonly known as effective bandwidth in the literatures [2][7] is then used for decision of call admission and routing.

Suppose that there are K types of traffic, and n_k sources of type k ($k = 1, \dots, K$) are multiplexed into a single large buffer B with deterministic service rate C . Let W be the stationary buffer occupancy. For asymptotically large B , it is shown in [7] that:

$$\sum_k n_k e_k(\delta) \leq C \Leftrightarrow \lim_{B \rightarrow \infty} \frac{1}{B} \log P(W \geq B) \leq -\delta \quad (1)$$

where $e_k(\delta)$ is the effective bandwidth for the type k source corresponding to δ and δ is a parameter determined by the cell loss probability requirement.

For a call accepted into the network, only the effective bandwidth is guaranteed. If we focus only on the call-level performance, such as call blocking probability, and assume that the effective bandwidth for the same type of traffic is fixed, then the routing mechanism using this predetermined bandwidth can be modelled as a multi-rate circuit-switched network.

In traditional circuit-switched networks, we do not distinguish direct path and alternative path because they require the same amount of bandwidth along the path but it is *not* the case in ATM networks. For instance, suppose we want to establish a VC between a and b with a required cell loss probability, say ϵ , in Fig. 1. If we route the call in the direct path $\{1\}$, the cell loss probability in VP should be less than or equal to ϵ . However, if the alternative route $\{5, 2\}$ is used, cell loss

can occur at either of the two VP input buffers along the route and thus the sum of the cell loss probabilities of VPs, $\{5\}$ and $\{2\}$, cannot exceed ϵ . Hence, the individual cell loss probability in each VP must be less than ϵ , say $\epsilon/2$. This more stringent cell loss probability implies that a larger effective bandwidth is needed in each link.

In this paper, we consider ATM networks supporting heterogeneous traffic with the same end-to-end QoS requirements (e.g., identical cell loss probability). We apply the effective bandwidth method [2] to find the bandwidth required to support the calls. We learned from the previous discussion that there are two different QoS requirements for the calls using direct and 2-links alternative routes respectively. In order to support two levels of QoS requirements at a node, namely, ϵ and $\epsilon/2$, we allocate dedicated buffers to each class of QoS. We assume that the traffic characteristics of a call do not change as it passes through a VP along its route. This assumption is also adopted in [5].

III. VP ROUTING PROBLEM

Let us consider a network consisting of a set of nodes N , a set of links (or VPs) L , and a set of possible Origin-Destination (O-D) pairs W . We assume that there exists a unique direct link between any two O-D pairs. Associated with each O-D pair w , $w \in W$, is a set of possible paths. We assume that for any link l ($l = 1, \dots, L$) its capacity is a fixed value C_l . Thus, we have $L = N(N-1)$ in a fully connected VP network.

There are K classes of traffic labelled $k = 1, \dots, K$ entering the network. Each traffic has its own traffic characteristics but all have the same end-to-end QoS requirements. We can transform this information into effective bandwidth requirements. As mentioned previously, the effective bandwidth for direct VCs and alternative VCs are different. Let the class k call require b_k^d units of bandwidth for direct path and $b_k^a (> b_k^d)$ units of bandwidth for alternative path. Without loss of generality, we assume that $b_1^d \leq b_2^d \leq \dots \leq b_K^d$. The class k call arrives to O-D pair w according to a stationary Poisson process with mean arrival rate λ_k^w . The call holding time for class k calls is assumed to be independent and exponentially distributed with mean $1/\mu_k$.

When a call arrives to the network, it will be either carried on a path or lost, depending on the routing policy being used. For each class k call carried on the network, we assume that it brings $r_k = b_k^d/\mu_k$ units of revenue to the network. In other words, we lose revenue r_k when the class k call is blocked. The objective of a routing scheme is to minimize the expected revenue loss of the network.

The network is described by a Markov chain with a finite number of states, where each state corresponds to the number of directly and alternatively routed VCs on the L VPs. The minimum revenue loss routing problem can be formulated as a Markov Decision Process (MDP). Recent papers by Hwang

[5][6] show that the MDP-based routing algorithms only slightly outperform the LLR-based routing algorithms. Although [5] [6] only cover the case of supporting homogeneous traffic class in VP-based ATM networks, this may give a hint on the performance of using MDP routing in multiple traffic case. Due to the simple implementation of LLR, we propose a modified LLR algorithm in the next section to handle the multi-class environment.

IV. DYNAMIC ROUTING ALGORITHMS

A. Least Loaded Routing

Least Loaded Routing (LLR) is the most popular real-time adaptive routing scheme in traditional single-rate circuit-switched networks because of its simplicity. In [4] and [6], LLR algorithm is applied to VP-based ATM networks supporting homogeneous traffic.

In the following, we describe an extended version of the LLR algorithm to handle VP-based ATM networks with heterogeneous traffic. Suppose there are K types of traffic in a VP subnetwork and a class k ($k = 1, \dots, K$) call arrives at the source node. Let the residual capacity $\hat{C}_l$ be the amount of capacity left unutilized on link l , ($l = 1, \dots, L$), at the call arrival instant. Then the residual capacity on route R at call arrival is defined as $\hat{C}_R = \min \{\hat{C}_l, l \in R\}$. Let T_k be the trunk reservation threshold for class k calls, which is the amount of bandwidth reserved for 1-link direct calls. The purpose of trunk reservation is to avoid the instability problem in LLR algorithm as mentioned in [1]. In the later part of the paper, we shall discuss the choice of T_k . Then the free bandwidth of a route is given by $\hat{C}_R - T_k$, which is the bandwidth available for 2-link class k alternative calls. Let S_k be the set of 2-link alternative paths whose free bandwidth is larger than b_k^a (i.e., $S_k = \{R \mid \hat{C}_R - T_k \geq b_k^a\}$). Hence, S_k is referred to as the candidate set for class k alternative calls.

LLR algorithm for VP-based ATM networks:

1. First attempt to set up the connection along the direct link if $\hat{C}_l \geq b_k^d$.
2. If there is not enough bandwidth to establish the connection in the direct link, select a 2-link alternative path with the largest free bandwidth in S_k .
3. If there are more than one such path, pick one randomly.
4. If the candidate set S_k is empty, block the call.

LLR tries to evenly distribute the loading in the network and is efficient for single-rate circuit-switched networks. However, this may not be the best strategy in a multi-rate environment because there are calls with diverse bandwidth requirements in such networks. It is suggested in [3] that packing of narrowband calls (calls require relatively smaller bandwidth) should be employed on some links to leave room on

other links for wideband calls. Siebenharr [10] proposes a dynamic algorithm known as Minimum Free Capacity Routing (MFCR) for VP-based ATM networks with several direct virtual paths connecting a node pair. The MFCR algorithm picks the direct VP with the smallest residual capacity among those VPs having enough residual capacity. Hence, MFCR tries to aggregate the unused bandwidth on one path.

In this paper, we modify the MFCR to find 2-link alternative paths instead of 1-link direct paths as originally proposed in [10]. The modified MFCR algorithm is very similar to LLR except in step 2. Our MFCR picks the 2-link alternative route with the smallest free capacity in S_k . Although MFCR can achieve maximum packing, it leads to non-uniform distribution of traffic. It is because packing and uniform distribution of traffic are two conflicting objectives and there is a trade-off between them. Gupta [3] proposes a modified LLR algorithm with packing to deal with this problem for circuit-switched network. In the following, we propose a real-time dynamic routing algorithm for VP-based ATM networks with heterogeneous traffic using both the packing and the LLR concepts.

B. Proposed Routing Algorithm

Suppose there is a class k call destined for a particular O-D pair (e.g., from a to b in Fig. 1) but there is not enough capacity to carry the call in the direct link (VP 1 in Fig. 1). Let us further assume that the candidate set S_k of 2-link alternative routes is non-empty. The objective of our routing algorithm is to route the class k call over a suitable alternative path R with minimal effect on future class j ($j > k$) calls. It is because we would like to reserve space for future wideband call. For class K call, no packing is needed because it has the largest bandwidth requirement.

Let $y_R^k = (y_{R,K}^k, y_{R,K-1}^k, \dots, y_{R,k+1}^k)$, $k = 1, \dots, K-1$ be a binary vector associated with route R for class k calls, where $y_{R,j}^k$ is given by

$$y_{R,j}^k = I \left(q \left(\frac{\hat{C}_R}{b_j^d} \right) - q \left(\frac{\hat{C}_R - b_k^a}{b_j^d} \right) \right), \quad j = k+1, \dots, K, \quad R \in S_k \quad (2)$$

where $I(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \end{cases}$

and $q(a/b)$ is the quotient of a divided by b .

Note that $q(\hat{C}_R/b_j^d)$ is the maximum number of class ($j > k$) direct calls that can be carried on route R . Hence, $y_{R,j}^k$ is an indicator value to show whether there is a decrease in the maximum number of class j direct calls that can be accepted on route R after a class k call is alternatively routed on R . The vector y_R^k shows how the other class j direct calls will be affected if a class k call is alternatively routed on R . For example, suppose we have four different classes of calls in a network and an arriving class 2 call is to be considered for

alternative routing on route R_1 . Further, let the vector $y_{R_1}^2 = (1, 0)$, which means that there is a decrease in the maximum of number of class 4 direct calls but no decrease for class 3 direct calls, if the class 2 call is alternatively routed on R_1 .

In order to compare different y_R^k associated with different alternative routes $R \in S_k$, we assign a route weight z_R^k to each y_R^k as follows:

$$z_R^k = \sum_{j=k+1}^K y_{R,j}^k 2^{j-k-1}, \quad k = 1, \dots, K-1 \quad (3)$$

Hence, $z_R^2 = 2$ for $y_{R_1}^2 = (1, 0)$ in the above example. In essence, we try to minimize the effect of the alternatively routed call on other wideband calls. We also implicitly assume that the revenue generated by a class j call is larger than that by a class $j-1$ call. Consequently, we assign a larger weighting factor for class j calls than class $j-1$ calls to reflect this intuitive reasoning. If this is not the case, we can easily modify the above weighting function to reflect the new situation.

We summarize our proposed routing policy in the following.

Proposed routing algorithm:

1. A class k call is first offered to direct route l if the unutilized bandwidth on the direct route $\hat{C}_l \geq b_k^d$.
2. If the direct route cannot carry the call, the call is then offered to an alternative path in S_k .
 - (a) If $k = K$, select a 2-link alternative path with the largest free bandwidth in S_k . If there are more than one such paths, then pick one randomly.
 - (b) If $k < K$, select a 2-link alternative route R with the smallest value of route weight z_R^k in S_k . In case of ties (two or more routes having the same z_R^k), the call is offered to a candidate route that has the largest free bandwidth. If there are more than one such paths, then pick one randomly.
3. If the candidate set S_k is empty, block the call.

Example: Consider a VP subnetwork supporting three classes of calls with bandwidth requirements of $b_1^d = 1$, $b_1^a = 1.3$, $b_2^d = 5$, $b_2^a = 7$, $b_3^d = 10$, $b_3^a = 12$. Suppose that a class 1 call arrives but it cannot be carried in the direct link. Further, assume that there are four candidate routes in the set $S_k = \{R_1, R_2, R_3, R_4\}$ with free bandwidth $\hat{C}_{R_1} = 36$, $\hat{C}_{R_2} = 37$, $\hat{C}_{R_3} = 39$, $\hat{C}_{R_4} = 40$. The results of the calculations are summarized in Table 1. The alternative path is R_4 using the LLR policy, while the MFCR algorithm would pick R_1 . According to our proposed algorithm, the class 1 call should be alternatively routed to R_3 because it

has the smallest value of z_R^1 and the largest free bandwidth among the candidate routes in S_1 .

In our previous discussion, we learn that an alternatively routed call consumes more resources on each of the two links than a direct call. To reduce the chance of alternative routing of wideband call, we set different values of trunk reservation threshold T_k for different classes of calls. In addition, [3] suggested that this may improve either the efficiency of the routing algorithms or the fairness of utilization. Since we would like to limit the chance of alternative routing of wideband calls, we propose that the trunk reservation thresholds should be chosen such that $T_1 \leq T_2 \leq \dots \leq T_K$. In addition, the value of T_k should be an integer multiple of b_k^d .

V. NUMERICAL STUDIES

In this section, simulation results are used to compare the performance of different dynamic routing algorithms in a VP-base ATM network supporting heterogeneous traffic. However, we assume that the different traffic sources all have the same end-to-end QoS requirement (e.g., cell loss probability). As a result, we can represent the traffic sources by their corresponding effective bandwidths as mentioned in Sections II & III. Henceforth, we assume that each class k call is described by its effective bandwidth requirements for direct VP, b_k^d , and for alternative route, b_k^a , respectively.

We consider a fully connected and symmetric network which has $N = 6$ nodes. Each VP has capacity $C = 100$ bandwidth units. There are two types of traffic classes: $b_1^d = 1$ unit, $b_1^a = 1.3$ units, $\mu_1 = 1$; $b_2^d = 5$ units, $b_2^a = 7$ units, $\mu_2 = 0.5$. It is found in [8] that the optimal trunk reservation problem in traditional circuit-switched networks can be formulated as a Markov decision process. However, the searching of the optimal values in multi-rate case is very difficult. To simplify the problem, the trunk reservation thresholds are set to the same values for all the schemes here. The trunk reservation thresholds for the two classes are $T_1 = 15$ units and $T_2 = 30$ units respectively. For each set of parameters, simulation is repeated with 10 independent runs. Each run corresponds to approximately 10^5 call arrivals to the system. For each run, the initial 10% of the simulation is discarded to account for transient effects. We consider five different routing algorithms in this numerical study, namely, direct routing, LLR, MFCR, Gupta's algorithm [3] and our proposed algorithm. Direct routing means that each call is either carried on the direct link or blocked, and no alternative routing is allowed. Gupta's algorithm is designed for multi-rate circuit switched network. Nevertheless, we extend it to VP-based ATM networks with slight modifications to take into account the difference in bandwidth required by the direct and alternative routes.

In Section IV, we mentioned that the desired routing

algorithm should minimize the expected revenue loss. Let β_k be the blocking probability of class k calls. For each class k call being blocked, we assume that the network loses $r_k = b_k^d/\mu_k$ units of revenue. Then the expected revenue loss $E(\text{loss})$ is given by

$$E(\text{loss}) = \sum_{k=1}^K \left(b_k^d/\mu_k \right) \lambda_k \beta_k \quad (4)$$

We compare the expected revenue loss of different algorithms in Table 2. The traffic load is varied from low (70%) to high (100%) and a 40% overload case is also tested. The performance of direct routing is used as our reference. The relative improvements of different routing schemes over direct routing are reported in Table 2. In the lightly-loaded to moderately-loaded region, there are significant reductions in the average revenue loss using alternative routing. It also shows that the idea of packing can improve the performance of LLR-based algorithms in multi-rate setting. In the highly loaded region, the probability of alternative routing is very small. Therefore, there is not much improvement using alternative paths and the performance of the four schemes are very closed to each other. However, our proposed algorithm in general performs slightly better than the others in this numerical example.

We now examine the performance of the algorithms with a larger set of call classes. Suppose we have four classes of calls in an ATM network and the traffic characteristics of the four classes of call are shown in Table 3. Two different traffic loading situations are tested and their performances are summarized in Tables 4. The numerical results show similar conclusions as in the two-traffic case. All the dynamic routing schemes outperform the direct routing and show approximately the same performance. This agrees with the results shown in [6] which indicate that the performance differences of various routing algorithms (LLR and MDP) are very small.

VI. CONCLUSION

We have discussed the dynamic routing problem in VP-based ATM networks. We propose a dynamic routing algorithm based on LLR with packing and compare its performance with other routing algorithms. Our routing algorithm basically uses packing to find a smaller candidate set of alternative paths and then applies LLR routing to this set. Although the idea is very intuitive, there are still some problems with it. In all the routing algorithms discussed, only the effective bandwidths of various calls are used but the information of arrival rates λ_k and mean call holding time $1/\mu_k$ are ignored in the routing decision. Hence, the arrival rates and the call holding time information should also be considered in the decision process in future studies.

REFERENCES

- [1] J. M. Akinpelu, "The overload performance of engineered networks with nonhierarchical and

hierarchical routing”, AT&T Bell Labs Technical Journal, vol. 63, pp. 1261-1281, 1984.

- [2] R. Guerin, H. Ahmadi, and M. Naghshineh, “Effective capacity and its application to bandwidth allocation in high-speed networks,” IEEE J. Select. Areas Commun., vol. 9, no. 7 pp. 968-981, 1991.
- [3] S. Gupta, P. Gandhi, “Dynamic Routing in Multi-Class Non-Hierarchical Networks”, Proceedings IEEE ICC 94, pp. 1390-1394, April 1994.
- [4] S. Gupta, K. Ross and M. El Zarki, “On Routing in ATM Networks”, IFIP Transactions C-15, North Holland, pp. 229-239, 1993.
- [5] R. H. Hwang, J. F. Kurose and D. Towsley, “MDP Routing in ATM Networks Using the Virtual Path Concept”, Proceedings IEEE Infocom 94, pp.1509-1517.
- [6] R. H. Hwang, “LLR Routing in Homogeneous VP-based ATM Networks”, Proceedings IEEE Infocom 95, pp. 587-593, April 1995.
- [7] G. Kesidis, J. Walrand and C.-S. Chang, “Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources”, IEEE/ACM Trans. on Networking, vol. 1, no. 4, pp. 424-428, Aug. 1993.
- [8] P. B. Key, “Optimal Control and Trunk Reservation in Loss Networks”, Probability in the Engineering and Informational Sciences, vol. 4, pp. 203-242, 1990.
- [9] K.R. Krishnan and R.H. Cardwell, “Routing and Virtual-Path Design in ATM Networks”, Proceedings IEEE Globecom 94, pp. 765-769, Dec. 1994.
- [10] R. Siebenhaar, “Simulative Comparison of Call Routing Algorithms in Virtual Path Based ATM Networks”, Proceedings of the 4th Open Workshop on High Speed Networks, Brest, France, Sept. 7-9, 1994.

ACKNOWLEDGEMENT

This research was supported in part by Hong Kong Telecom Institute of Information Technology grant: HKTIIT 93/94.EG01

Fig. 1. A 4-node VP subnetwork

route R	free capacity $\hat{C}_R$	indicator vector $y_R^1 = (y_{R,3}^1, y_{R,2}^1)$	route weight z_R^1
R_1	36	(0, 1)	1
R_2	37	(0, 0)	0
R_3	39	(0, 0)	0
R_4	40	(1, 1)	3

Table 1: Routing computation in our proposed algorithm

Arrival Rate (λ_1, λ_2)	Percentage reduction in average revenue loss over that using direct routing			
	LLR	MFCR	Gupta	Proposed algorithm
(50, 2)	38.05%	44.03%	47.53%	48.12%
(50, 3)	15.52%	16.54%	17.48%	18.64%
(50, 4)	4.78%	4.97%	5.20%	6.13%
(50, 5)	1.35%	1.58%	1.43%	1.75%
(30, 5)	9.30%	10.06%	8.99%	10.22%
(40, 5)	5.36%	5.06%	5.06%	5.37%
(45, 5)	3.74%	3.74%	3.53%	4.17%
(70, 7)	0.06%	0.10%	0.12%	0.15%

Table 2: Relative Performance of Routing Algorithms

class k	b_k^d (units)	b_k^a (units)	μ_k	trunk reservation T_k (units)	λ_k	
					Data set I	Data set II
1	1	1.3	1	20	20	26.667
2	5	7	1	25	4	5
3	10	13	0.5	30	1	1.267
4	20	23	0.25	40	0.267	0.267

Table 3: Traffic characteristics

data set	Percentage reduction in average revenue loss over that using direct routing			
	LLR	MFCR	Gupta	Proposed algorithm
I (81% load)	4.11%	3.33%	4.16%	5.01%
II (98% load)	1.78%	1.29%	1.64%	2.10%

Table 4: Relative Performance of Routing Algorithms