IFIP Workshop TC6

IFIP Working Groups 6.3 and 6.4

FOURTH WORKSHOP on PERFORMANCE MODELLING and EVALUATION OF ATM NETWORKS

Craiglands Hotel
Ilkley, West Yorkshire, UK

8 – 10th July 1996

PARTICIPANTS PROCEEDINGS

Edited by
Demetres Kouvatsos
University of Bradford, UK

With the support of the Performance Engineering Groups of the British Computer Society and British Telecom (BT), Skelton GmbH, Telematics International Ltd, Departments of Computing, Electrical Engineering and Mathematics, University of Bradford and Engineering and Physical Sciences Research Council (EPSRC), UK.

General Chair

Demetres Kouvatsos, Bradford, U.K.

Scientific Committee

Åke Arvidsson, Karlskrona, Sweden
Monique Becker, Evry, France
Chris Blondia, Antwerpen, Netherlands
Alexandre Brandwajn, Santa Cruz, U.S.A.
Herwig Bruneel, Ghent, Belgium
Nigel Burton, GPT Ltd., U.K.
Olga Casals, Catalunya, Spain
Marco Conti, CNUCE, Italy
Laurie Cuthbert, London, U.K.
Tadeusz Czachórski, Gliwice, Poland
Lorenzo Donatiello, Bologna, Italy
Serge Fdida, Paris, France
Georges Fiche, ALCATEL CIT, France
Erol Gelenbe, Duke, U.S.A.
Nicolas Georganas, Ottawa, Canada
Richard Gibbens, Cambridge, U.K.
Fred Halsall, Wales, U.K.
Peter Harrison, London, U.K.
Boudewijn Haverkort, Twente, Netherlands
Christoph Herrmann, Philips, Germany
Ilias Iliadis, IBM Zürich, Switzerland
László Jereb, Budapest, Hungary
Gunnar Karlsson, Kista, Sweden
Peter Key, BT, UK
Ernest Koenigsberg, Berkeley., U.S.A.
Daniel Kofman, Paris, France
Ulf Körner, Lund, Sweden
Paul Kühn, Stuttgart, Germany
Phil Mars, Durham, U.K.
Marco Ajmone Marsan, Torino, Italy
Isi Mitrani, Newcastle, U.K.
Nicholas Mitrou, Athens, Greece
Arne Nilsson, North Carolina, U.S.A.
Raif Onvural, IBM, U.S.A.
Achille Pattavina, Milano, Italy
Harry Perros, North Carolina, U.S.A.
Michal Pióro, Warsaw, Poland
Ramón Puigjaner, Illas Balears, Spain
Guy Pujolle, Paris, France
Douglas Reeves, North Carolina, U.S.A.
Doug Shepherd, Lancaster, U.K.
Andreas Skliros, Telematics Int. Ltd., U.K.
Geoff Smith, Strathclyde, U.K.
Otto Spaniol, Aachen, Germany
Ioannis Stavrakakis, Boston,, U.S.A.
Yutaka Takahashi, Kyoto, Japan
Don Towsley, Massachusetts, USA
Phuoc Tran-Gia, Würzburg, Germany
Yannis Viniotis, North Carolina, U.S.A.
Hideaki Yamashita, Tokyo, Japan

Local Committee

Irfan-Ullah Awan
Rod Fretwell
Hamid Ullah Jan
Charlotte Johnson
Jean Lees
Danielle Roche
Charalabos Skianis
Vassilios Tsiantos
Jeffrey Wilkinson

Dynamic Route Selection at Call Set-up Level in ATM Networks

R. Bolla, A. Dalal'ah, F. Davoli, M. Marchese

Department of Communications, Computer and Systems Science
DIST - University of Genoa
Via Opera Pia 13, I-16145 Genova, Italy

Abstract

A new policy for routing in ATM networks is introduced. Virtual circuits are routed at call set-up level, by combining Call Admission Control (CAC) and routing decisions in a hop-by-hop fashion at the nodes along a path. The network traffic is divided into classes, homogeneous with respect to performance requirements and statistical characteristics, which are dynamically assigned bandwidth partitions over the network links, by means of scheduling algorithms at the switching nodes. At each node traversed, a call request of a certain class is first assigned a subset of outgoing links towards the destination, upon which the resources necessary to maintain the required level of QoS are available; then, a routing decision is taken, by choosing the least "cost" link, according to a specific criterion. The link costs are dynamically updated for all traffic classes, by means of local information and of aggregate information exchanged among neighbouring nodes, and are based on a measure of the link "saturation", in terms of some "distance" from the boundary of the load region in call-space that the multiplexer serving the link can support. In case that no route is available at a node, the call request is traced back to the least-loaded node (according to the measure chosen) among the ones already visited along the path. This strategy is analysed and compared by simulation, in terms of performance and of the capability of reacting to network failures, with a previous one that immediately drops the call at the node where blocking occurs.

1. Introduction

The heterogeneous traffic flows nature and performance characteristics in the Asynchronous Transfer Mode (ATM) technique give rise to several control problems, such as bandwidth allocation, admission control and routing, [1, 2], which still deserve investigation, both in their own respect and in their interaction, even though they have generated lots of effort in the literature (see, for instance, [3-7]). The admission control scheme can be effected at the call- or cell-level, or both [1]. The problems of admission control and routing are obviously related, since they both involve allocation of the same resources; nonetheless, trying to solve them independently not only simplifies the global task, but also yields greater flexibility in combining different schemes.

Some version of distributed shortest path routing has been considered as an upper bound in comparison with newly suggested strategies for routing [1]. In the latter, some hypotheses made in the literature, such as a full connectivity, the constraint of at most two physical hops [4], or even of a homogenous traffic flow [5], may be relaxed to allow an efficient use of the ATM network resources in a real environment.

In the present paper, we propose a new strategy, named Re-attempt Distributed Least Congested Path (R-DLCP) in the following, which operates in a hop-by-hop fashion to route a connection request, with the possibility of tracing back to a previously visited node in case of failure in finding an available route at a certain point along the path. Although this mechanism can be applied to a family of algorithms, the specific strategy that will be described was originally proposed to overcome some shortcomings, in particular with respect to robustness to failures, emerged from the DLCP (Distributed Least Congested Path), a routing method previously proposed by the authors [3], and to enhance, in general, its performance as well. The R-DLCP is a step-backward algorithm, in the sense that, as mentioned above, if the process of establishing a path faces a congested point, it returns back to some point and retries again.

The present work is proposed to be integrated in an overall control scheme, so that the admission control and bandwidth allocation methods used in [3] are preserved. The traffic is divided into classes, characterised by statistical characteristics and Quality of Service (QoS) requirements and, upon a connection request, the admission control and the routing strategy are implemented independently for each class of traffic. The routing decisions are realised in a distributed computational structure.

The rest of the paper is organised as follows. In the next section, the features of the proposed algorithm are discussed; in Section 3, the routing strategy is described; in Section 4, some simulation results are considered and discussed, and Section 5 is dedicated to the conclusions.

2. Features of the proposed dynamic routing mechanism

The traffic is supposed to be divided into H classes, where each class differs from the others for the required QoS, in terms of cell loss and delayed cell rate, and for statistical properties, such as average (or sustainable) and peak cell rate. Each class is dedicated a fixed amount of bandwidth over each link, as in a Complete Partitioning scheme [2]. However, we suppose that a dynamic fair assignment of the bandwidth partitions among the various classes of traffic is maintained by link bandwidth allocators [3], which periodically adjust the bandwidth assignments, to be kept over a fixed period of time. At each node, a connection of a certain class is admitted if the nodal local admission control can find at least an outgoing link over which the call can be accommodated, by ensuring the desired QoS level to the new call as well as to the ones in progress.

At connection set-up, the source sends a special packet, which will be called Resource Reservation Packet (RRP) in the following. When a node receives a RRP, it verifies whether it can accommodate the required resources at one of its outgoing links towards the destination or not. If it turns out that there is at least one available link, the routing algorithm, yet to be described, is used to send the RRP to the successor node; otherwise, in the R-DLCP strategy, the RRP returns back to the least loaded node along the already established path, and it retries again starting from that node. If the RRP cannot find the required resources over one of the outgoing links of the above mentioned node, it will be changed into a Free Resource Packet (FRP), which must be sent back in its turn along the already established path, releasing the resources reserved for the call related to it. With a similar philosophy as in [6], the RRP is sent thus with two objectives, the first one being the purpose of establishing a route, and the second one being its usage in the case of failures. For the sake of non saturating the network, the process of re-routing the RRP back to the least loaded node along the already established path is limited just to a single attempt.

The decisions of the algorithm are based on the minimisation of the sum of a "local cost" and an aggregate cost, evaluated at the node at the moment of the decision only; thus, there is no need to know the global situation over the network to execute the algorithm at a specific node. More specifically, the information needed to take a decision is limited to the node under consideration and to an information exchange mechanism among adjacent nodes. It was argued in [9] that basing the decision on a similar kind of information makes the routing procedure more practical from an implementation point of view. In the case of a huge network, the overhead incurred in acquiring remote information would be considerable; as will be apparent in the following, the aggregate information exchanged among the nodes in our case is limited to an "indication" of the downstream saturation, without regard to the specific destinations. In this sense, the algorithm can be considered to be scalable (even though, obviously, the "age" of the aggregate information, as an indicator of the network status, grows with the network size).

In general, any routing algorithm must be robust to link or node failures. To illustrate this point, consider Fig. 2, which depicts an arbitrary network, used for simulation and analysis purposes in the present paper. Suppose that the link between node 9 and node 11 is broken or node 9 has failed. In the DLCP algorithm of [3] (where the same information exchange mechanism is used as in the R-DLCP, but a call is rejected if no available route is found at a node, without re-attempting to find an alternative path), the neighbours of node 9, after notification by the signaling protocol (in case of node failure) or by node 9 itself (in case of link failure), would increment their cost. The increment would then be propagated downstream, but it would take some time (depending on the frequency of the cost updating mechanism) to reach the furthermost nodes; in the meanwhile, the latter would be unaware of the congestion, and might still direct calls in the direction of node 9, which might incur higher rejection probability the closer they get to it. This problem might be overcome by using an algorithm as the one proposed in [6]; however, in that approach, a part of the network capacity is dedicated for the re-routed calls after any single failure, which not only dimishes the overall capacity of the given network, but is also hard to compute.

On the contrary, the R-DLCP reacts implicitly in the same way as the DLCP, but the calls attempting to pass through node 9 would be deviated, instead of being refused, by the back-tracking mechanism. Thus, the nodes in the network other than node 9 should not suffer unnecessary call rejection during the transient period, before the network becomes aware of the congestion. In fact, if the other part of the network can accommodate those calls, they will be accepted. In the sense, the R-DLCP better meets the robustness requirements.

3. Routing Strategy

The routing algorithm applied at each node is based upon the computation of a simple cost function, related to each outgoing link, given by the sum of a local cost and an aggregate cost. Let i be the node considered; then the cost of link ij for class h at decision instant k (in slots) is defined as

$$W^{(h,ij)}(k) = W_{loc}^{(h,ij)}(k) + \alpha_j W_{agg}^{(h,j)}(s) \tag{1}$$

where ij is the link with origin i and destination j and α_j is a weighting coefficient. The aggregate cost $W_{agg}^{(h,j)}(s)$ refers to the traffic conditions of node j and its successors, which was evaluated at instant s<k (in slots) and communicated to node i (s then represents the most recent updating instant of the costs in (1)). Using the current values of the local cost and the aggregate cost at the node (which highlights the behaviour of the algorithm as a "greedy" one), has a drawback in the fact that, owing to the propagation delay of aggregate costs, in a large network, several changes may have happened, which are not reflected in the information upon which a decision is based. On the other hand, a large network dimension would also hinder the implementation of a centralized algorithm in the same way.

Let $N_k^{(h,ij)}$ be the number of connections in progress for class h on link ij and $N_{max}^{(h,ij)}$ be the maximum number of connections for class h at link ij that can be accepted, maintaining QoS and according to the admission policy. Moreover, let m indicate the slot when a bandwidth re-allocation was performed, and K the duration (in slots) of the reallocation period (we recall that we are using a Complete Partitioning scheme, where, however, bandwidth partitions of the link or Virtual Path (VP) capacities may be periodically recomputed). We define the local cost of link ij and class h at instant k as

$$W_{loc}^{(h,ij)}(k) = \begin{cases} \dfrac{1}{N_{max}^{(h,ij)}(m) - N_k^{(h,ij)}} & \text{if} \quad N_{max}^{(h,ij)}(m) > N_k^{(h,ij)} \\ Z & \text{if} \quad N_{max}^{(h,ij)}(m) = N_k^{(h,ij)} \end{cases} \tag{2}$$

where $k \in [m, m+K-1]$, and Z is a very large value (Z should be large enough to ensure that no saturated link will be chosen if non-congested links are available). Thus, the local cost is inversely proportional to the "available space" (in terms of the number of acceptable connections on the link), and it is Z, when there is no more bandwidth available; i.e., when no more calls of class h can be accepted on that link.

As regards the aggregate cost, we have chosen the following

$$W_{agg}^{(h,j)}(s) = W_{agg,node}^{(h,j)}(s) + \beta_j W_{agg,succ}^{(h,j)}(s) \tag{3}$$

where β_j is a weighting coefficient. $W^{(h,j)}_{agg,node}(s)$ represents the average situation of the node with respect to its congestion state, and $W^{(h,j)}_{agg,succ}(s)$ is an aggregate information on the average congestion of its successor nodes. More specifically, we have defined

$$W^{(h,j)}_{agg,node}(s) = \frac{1}{L_j} \sum_{n \in Succ(j)} W^{(h,jn)}_{loc}(s) \tag{4}$$

$$W^{(h,j)}_{agg,succ}(s) = \frac{1}{L_j} \sum_{n \in Succ(j)} W^{(h,n)}_{agg}(s) \tag{5}$$

When a node receives a RRP, whose requested destination is d, the node scans a list (already maintained by the access controller) of the links and their corresponding cost values (in non-decreasing order), and a set 1^d for each destination d, and stops at the first link $i\hat{j}$ with $\hat{j} \in 1^d$. If $W^{(h,i\hat{j})}(k) = Z$, the connection request is referred to the least loaded node r_k along the path already done, and retries under the same conditions (but excluding the link that was previously chosen at r_k and led to node i), starting at node r_k; if the connection is accepted at r_k, say over link $r_k s$, then the resources are reserved on that link and the connection attempt proceeds further; otherwise the connection request is refused and a FRP is sent back to release the already allocated resources. Whenever a connection is accepted on link ij, $W^{(h,ij)}$ is updated, by adding 1 to $N^{(h,ij)}_k$, and it is placed in the list in the correct position; then, the RRP proceeds to a successor node, until either it is refused or the destination is reached. In the latter case, the resources must be freed after the connection is closed by either the origin or the destination.

We consider also a slightly different variant of the R-DLCP, where, upon reaching the re-attempt node, the first next hop is chosen from a list of the links in arbitrary order, as the first one that can accommodate the connection. This variant, which will be indicated as RF-DLCP (Re-attempt First-fit DLCP), is aimed at avoiding the use of the cost in the first choice, in order to reduce the possibility that a prevalence of the aggregate information (not yet updated) would lead again toward the same region of the network.

The updating of the aggregate parts of the cost can be effected in different ways. In the simulations reported in the next section, a periodic synchronous information exchange (with period T [slots]) was assumed, whereby each node updates its aggregate cost after receiving the updated cost by its downstream neighbours. The mechanism is synchronous, in the sense that all the nodes at an equal number of steps from the destination act simultaneously. This would be impractical in a real implementation; however, an asynchronous updating mechanism is possible, just in the same way as in distributed shortest path algorithms [11].

4. Simulation results

The performance of the proposed routing scheme has been tested by simulation on a simple twelve-node network. Two traffic classes, a "reallocation interval" $K = 8 \cdot 10^7$ cells and a channel transfer capability C = 150 Mbits/s, with a related slot duration $T_s = 2.83 \cdot 10^{-6}$ s (53 bytes/cell), have been used.

The quantities $N^{(h)}_a$ [Erlangs], h=1, ..., H, represent the global average traffic intensities offered to the network; the call arrival processes follow independent Poisson distributions. All other parameter values are shown in Table 1. The parameters $\varepsilon^{(h)}$, $\delta^{(h)}$ and $D^{(h)}$ represent the cell-level QoS requirements, namely, the upper bound on cell loss probability (P_{loss}), the upper bound on the probability (P_{delay}) of exceeding a delay $D^{(h)}$, and the dealy constraint, respectively.

TRAFFIC CLASS: h	h=1	h=2	h=3
PEAK BANDWIDTH: $P^{(h)}$	1 Mbits/s	2 Mbits/s	10 Mbits/s
BURSTINESS: $b^{(h)}$	2	5	10
AVERAGE BURST LENGTH: $B^{(h)}$	100 cells	500 cells	1000 cells
AVERAGE CONNECTION DURATION	20 s	15 s	25 s
Ploss UPPER BOUND: $\varepsilon^{(h)}$	0.0001	0.0001	0.0001
Pdelay UPPER BOUND: $\delta^{(h)}$	0.001	0.001	0.001
DELAY CONSTRAINT: $D^{(h)}$	400 slots	200 slots	100 slots
BUFFER LENGTH: $Q^{(h)}$	20 cells	15 cells	10 cells

Table 1. Parameter values.

We refer to the traffic flow generated by the above data as an offered load 1, with a fixed value $N_a^{(1)}=120$; $N_a^{(2)}=100$; $N_a^{(3)}= 15$ (traffic reference); an offered load "x" corresponds to the same data, except for the traffic intensities $N_a^{(h)}$, h=1, 2, 3 which are multiplied by x. The coefficients α_i and β_i, $i=0,...,11$, are considered the same in each node, that is $\alpha_i=\alpha$ and $\beta_i=\beta$, $\forall i$.

The topology of the network that has been used in the simulations is shown in Fig. 2, and is composed of twelve nodes, only one of which (node 11) is a destination .

This Section is dedicated to a comparison between the two strategies R-DLCP and RF-DLCP introduced above and also with the DLCP strategy, already presented by the authors in previous works [3]. The advantages and the drawbacks of each strategy are analysed.

The choice to have a single destination may seem a strong restriction; however, the topology chosen can be seen as a particular network "view" of node 11 and, as such, it allows a simple analysis of network parameters and a simple comparison among the presented routing strategies, which is the main objective of this paper. Values ($\alpha=1$, $\beta=1$) of the weighting coefficients and an updating time of K/T=2 have been chosen for the simulations. The criterion for stopping the simulation is that the width of the 95% confidence interval should be less than 3% of the value of the sample average of the quantity of interest.

The α and β values are not the best ones for this topology. A simulative analysis, whose results have not been shown here, has verified that small values of α and β provide the best results; in fact, since in this network the average number of hops to get to the destination is relatively short (see Fig. 5), the importance of the aggregate cost is minimized and the network topology greatly enhances the importance of the "local" part of the cost.

However the use of higher values of α and β is reasonable if the network in Fig. 2 is considered to be a subnet of a larger one, where the choice of "optimal" values of the weighting coefficients for each node would be too difficult to manage, even if theoretically possible (actually, the general framework in Section 3 considered possibly different coefficients α_i and β_i, for each node have been used).

A comparison among the three routing strategies is shown versus the traffic load in Fig. 3. It can be noted that there is a load interval (0.8-1.1) where the advantage of using R-DLCP and, in particular; RF-DLCP is more evident. In Fig. 4, the 'gain' of the R-DLCP and RF-DLCP is depicted versus the traffic load, taking the DLCP values as a reference. It can be seen that there is a load interval (0.8 - 1.1) where the advantage of using R-DLCP and RF-DLCP is more evident. It has to be remembered that the variability of the presented values is really low, due to the high degree of confidence (3% confidence interval).

The increased computational burden is not so higher and the time to get to know if a new connection has been accepted not so longer for R-DLCP and RF-DLCP with respect to DLCP. This sentence is justified by the values in Fig. 5 where the average number of hops to get to the destination is shown versus the traffic load for the three strategies.

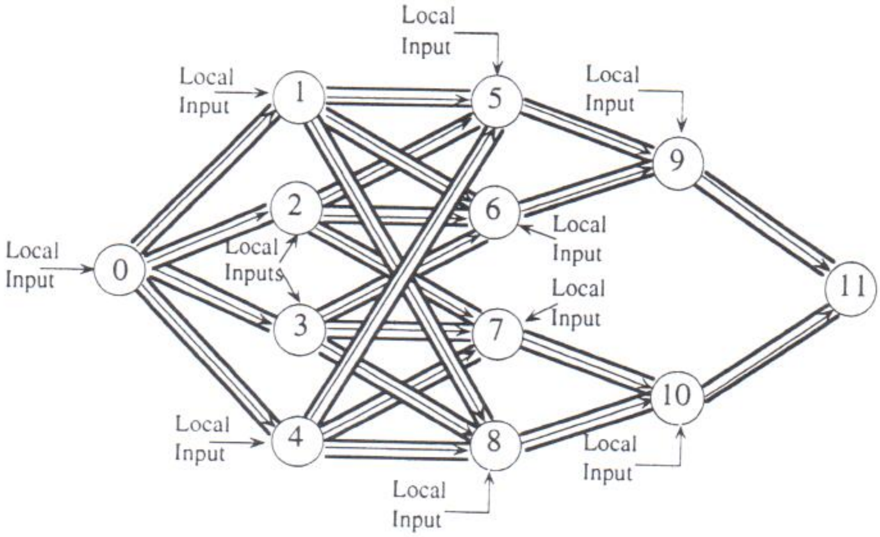

Figure 2. Topology of the test network.

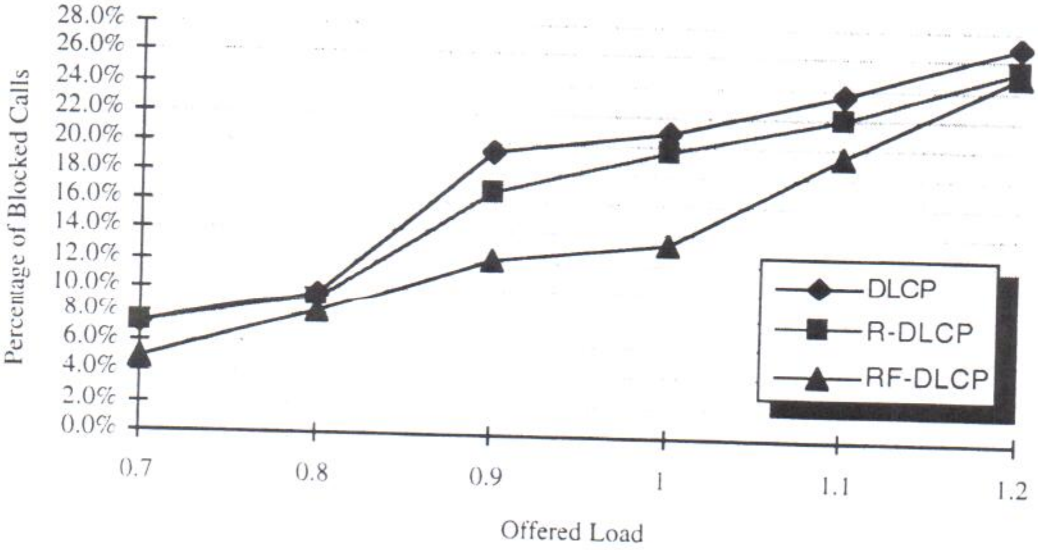

Figure 3. Total percentage of blocked calls versus the offered load (DLCP, R-DLCP, RF-DLCP).

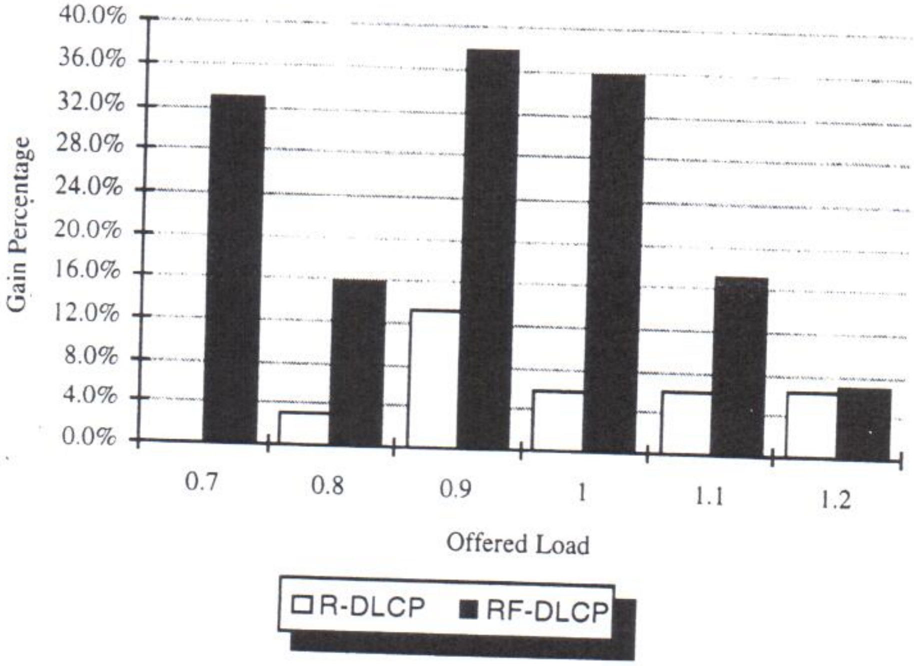

Figure 4. Gain percentage in the total blocked calls with respect to DLCP versus the offered load (R-DLCP, RF-DLCP).

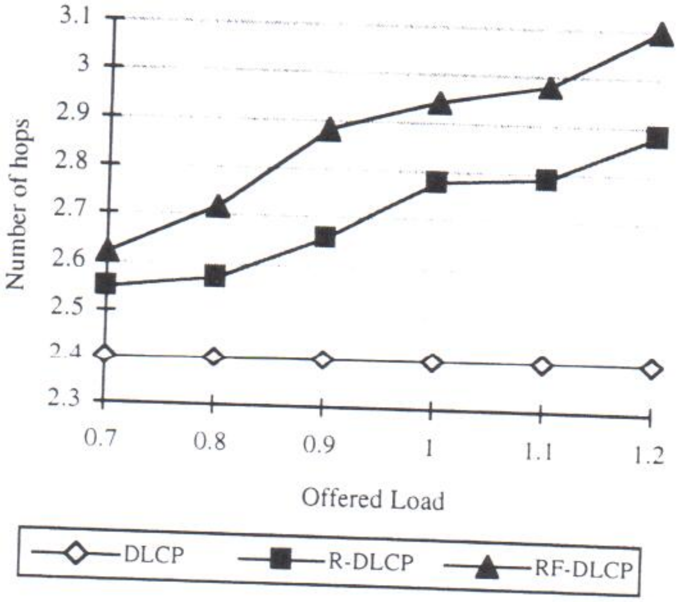

Figure 5. Number of hops to get the destination versus the offered load (DLCP, R-DLCP, RF-DLCP).

5. Conclusions

A dynamic routing scheme has been introduced, which is based on the combination of a distributed algorithm and a re-attempt strategy in case of failure to establish a connection for the first time. The distributed algorithm is based on an information exchange among adjacent nodes and on a metric represented by the "available space" on each ougoing link. The re-attempt procedure has been introduced to increase the possibility of a faster response to changes in the load or the network topology; two slightly different methods have been considered.

Some preliminary simulation results have been used to discuss the different aspects of the schemes, also in comparison to the situation where no re-attempt is allowed.

References

[1] R.O. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues*, Artech House, Inc. Norwood, MA, 1994.

[2] K.W. Ross, *Multiservice Loss Models For Broadband Telecommunication Networks*, Springer-Verlag, London, 1995.

[3] R. Bolla, F. Davoli and M. Marchese, "A Global Control System for Integrated Admission Control and Routing in ATM Networks", *Proc. IEEE GLOBECOM'95*, Singapore, Nov. 1995, pp. 437-443.

[4] S. Murthy and J.J Garcia-Luna-Aceves, "Dynamics of a Loop-Free Path-Finding Algorithm", *Proc. IEEE GLOBECOM'95*, Singapore, Nov. 1995, pp. 1347-1351.

[5] K. Begain, "Discrete Performance Model of Multiplexer in ATM", *Proc. IEEE ICECS'95*, Amman, Jordan, Dec. 1995, pp. 233-239.

[6] R.Dighe, Q. Ren and B. Sengupta, "A Link Based Alternative Routing Scheme for Network Restoration under Failure", *Proc. IEEE GLOBECOM'95*, Singapore, Nov. 1995, pp. 2118-2123.

[7] C. Alaettinoglu, I. Matta and A. Shankar, "A Scalable Virtual Circuit Routing Scheme for ATM Networks", *Proc. IEEE ICCCN'95*, Las Vegas, NV, Sept. 1995, pp. 630-637.

[8] G. Oberhauser and R. Simha, "Fast Data Structures for Shortest Path Routing: a Computer Evaluation", *Proc. IEEE ICC'95*, Seattle, June 1995, pp. 1597-1601.

[9] A. Shaikh, S. Lu and K.Shin, "Localized Multicast Routing", *Proc. IEEE GLOBECOM'95*, pp. 1352-1356.

[10] W. Zaumen and J.J. Garcia-Luna-Aceves, "Dynamics of Distributed Shortest-Path Routing Algorithms", *Proc. ACM SIGCOMM'91*, Zurich, Switzerland, Sept. 1991, pp. 31-42.

[11] D. Bertsekas, R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, NJ, 1987.