# Dynamic Routing Based on a single parameter: Link Congestion Probability. [1]

Jose Luis Marzo-Lázaro*, Jordi Domingo-Pascual**, Ramón Fabregat-Gesa*, Josep Sole-Pareta**

* Universitat de Girona. EPS. Girona (Spain). Tel.: +34 72 418484. Fax +34 72 418399. E-mail marzo@ei.udg.es
** Universitat Politècnica de Catalunya. DAC. Barcelona (Spain). Tel.: +34 3 4016982. Fax +34 3 4017055.

## Abstract

In this paper we propose an implementation for dynamic routing with alternative routes of calls based on anticipated calculation. The routing decision criteria basically depends on the type of source of the new connection and the current network loading. We suggest an algorithm based on the Probability of Congestion in the link as an optimal control parameter for routing decision. The calculation of the Probability of Congestion is based on the New Convolution Approach.

The objectives of our proposal are: to obtain a real-time connection setup with the routing algorithm, to define an optimal Routing Control Parameter, to maximize network resources utilization and to minimize call blocking probabilities. In order to obtain fast connection setup when a new connection demand arrives, each node already has a pre-computed Probability of Congestion table for each type of source and for each output Virtual Path. To obtain the improvement of resource utilization we propose a dynamic routing scheme with alternative routes. The method proposed for the selection of the route is based on the well-known Least Loaded Routing (LLR) algorithms used in circuit switched networks. Different routing algorithms are compared when changing the criteria by the selection of available alternative routes and when changing the choice of alternative route.

## 1. Introduction

Network Provisioning (NP) is the set of the long-term control actions that determine the physical quantities of the resources to be placed in the network. Given the network resources by a suitable NP, Network Resource Management (NRM) is the set of control functions related to establish VPs and to allocate bandwidth, which are performed by the network in order to optimize performance and utilization objectives, namely to provide the required Quality of Service (QOS) and to maximize utilization, with simple Connection Admission Control (CAC) and Routing procedures. The NRM can be conceived as the set of functions required to allocate bandwidth and buffer to groups of connections at the network nodes. To establish a connection between two arbitrary nodes, an appropriate path through the network must be identified (routing), and sufficient resources must be secured at every intermediate node along the selected route (resource allocation). The drawback of many existing routing algorithms is the lack of cooperation between congestion control and routing. NRM and CAC are closely related. CAC decides on the acceptance of new connections based on the resources made available to the various paths, while NRM must change the allocated resources to the paths according to the demands placed by the users on CAC. Routing could also be considered as belonging either to NRM or to CAC.

The basic objective of a bandwidth management and traffic control strategy in Asynchronous Transfer Mode (ATM) network is to allow for high utilization of network resources (transmission bandwidth, buffer space,etc.), while sustaining an acceptable QOS for all connections. Several network traffic control functions, such as congestion control and routing, depend on the characterization of the bandwidth required for each individual connection and the resulting load on network links and buffer utilization.

When the calls are not admitted according to their peak bandwidth requirements all the existing connections on a link are statistically multiplexed and the efficient use of transmission capacity is obtained. Some authors assign for each source an equivalent bandwidth that reflects its characteristics. In these cases, the accuracy of the bandwidth evaluation is seriously affected especially with heterogeneous traffic. To provide traffic control functions in real-time is a major challenge. Usually this involves a reduction in the complexity, and therefore a reduction of the accuracy of the evaluation models. In [9] an approach is proposed that combines two basic approximated models to study Bandwidth Allocation: the fluid flow model, which estimates the equivalent capacity when the impact of individual connection characteristics is critical, and the stationary model, which is representative of bandwidth requirements when the effect of statistical multiplexing is significant. By the second

[1] This work has been supported by CICYT (Spanish Education Ministry) under contract TIC92-1289-PB.

model, the convolution approach is the most accurate method used in bandwidth allocation when a priori traffic estimation is considered. In fact, the convolution approach obtains a probability density function for the offered system load, expressed as the probability that all sources together are emitting traffic at a give rate. However, the convolution approach has a considerable computation cost and a high number of accumulated calculations. Nevertheless, in critical near-congestion situations, the convolution is an algorithm that gives sufficiency accuracy. One important aspect of Bandwidth Allocation is CAC.

CAC addresses a set of actions taken by the network at the connection setup phase (or during the connection re-negotiation phase) in order to establish whether a new connection can be accepted or rejected. The decision is made based on the knowledge of the current network loading, on the traffic descriptors and on the cell loss and/or cell delay performance requirements (QOS) of the new connection and those of the connections already existing. In [14] and [8] we propose the use of the New Convolution Approach (NCA). These works have studied the usage requirements (complexity as processor capacity and memory required to do the calculations) for bandwidth allocation based on the convolution approach. These papers show reasonable evaluation cost and storage requirements whenever the multinomial distribution function and some cut-off mechanisms are used. In this case the NCA offers a possible real-time evaluation without accuracy loss. The adopted CAC policy has a strong influence on the routing performance. Routing algorithms define the procedure by which a path is sought and defined for network wide connections. Routing algorithms are based on the network topology, on the current state of the loading and on fast signalling techniques to perform the control.

***ROUTING IN ATM NETWORKS***

The routing in ATM networks consists of assigning a path to the incoming connection demands during the call setup. For each new call[3] the network must select a path that has sufficient bandwidth available to support the new connection and to guarantee the required Quality of Service (QOS). Otherwise, the new call will be blocked (or rejected). Cells of the connection are transferred along the assigned route. The traffic produces random variations in the link occupancies in the network. The aim of a routing scheme is to induce, as fas as to the possible, these random variations in the pattern of link occupancies to maximize network resources utilization and to minimize network blocking (the probability of an arriving call finding no suitable idle path for its connection). The routing policy attempts to find the path that causes the least damage to the network.

In ATM networks, a Virtual Path route is established by setting the ATM routing tables cross-conected between Virtual Path Terminators. The information for path routing is obtainable simply by changing routing information in the path connection tables at nodes along the path. Each input has its own Virtual Path Identifier conversion table. VP routing is easily controlled by changing table contents. Static and dynamic routing policies can be considered. To perform dynamic routing policies, each node must maintain a table specifying, at each moment, the path selected which depends on the state of the network at the instant of call arrival. In static routing, this table is fixed.

*ALTERNATIVE ROUTING*

Dynamic routing gives a better chance of success to an individual call by increasing the number of ways it can traverse the network. When a new call requests establishment between a pair of switches, it is possible to consider routing schemes in VP networks which always offer the direct VP first; then this path/route will be used with priority. If a call is blocked on the direct VP, the call is setup on an allowable alternative route. Otherwise, the call is blocked. The routing algorithms differ in how they choose from among the set of allowable alternative routes. By providing alternative paths, ATM networks will be able to achieve lower connection blocking probabilities and higher network throughput, while still satisfying various QOS requirements through the rigorous application of CAC. The blocking probability can be decreased by using alternative paths. [21] .Several studies exclude alternative routes that consist of more than two VPs. Thus, every pair of terminals has an associated set of routes, where each route consists of one direct VP, or two VPs as alternative routes [3], [11], [19], [18], [15], [20]. The two VP constraint simplifies routing implementation and ensures loop-free routing.

Removing the restriction to two VPs (Links), of course, allows a wider choice of alternative routes. One effect of this is that it gives more flexibility in routing and, thus, tends to reduce the blocking probability. Another effect is the reduction of the effective capacities of the physical links. This effect tends to increase the blocking probabilities. This result implies that the effect of the increase of routing flexibility is overweighed by the effective capacity reduction. However, in general, there is a price to be paid for the use of multiple VP (Links)

---

2 Throughout this paper the terms “path” and “route” are interchangeable.

3 Throughout this paper the terms “call”, “connection demand” and “Virtual Circuit” are interchangeable.

for a single call, viz., the fact that it constitutes an inefficient use of network resources to tie up several VP (Links) for the sake of a single call, while those VPs (Links) might be used to complete several separated calls [16] , [17] . Moreover, the cost of routing the call over a multiple-VP route is the sum of costs of routing the call over its constituent VPs [15] . In [17] and [6] the effects of limiting the maximum number of physical links for each path are discussed. It is shown in [1] that providing even as few as 2 or 3 multiple paths per source-destination pair will result in a marked decrease in call blocking probability. Superior results in terms of a higher percentage of offered traffic satisfied are obtained with a greater number of total VPs in the network, fewer VPs per path, and fewer physical links per path.

When a call is to be routed on an alternative path consisting of two VPs, we again encounter the following problems: how to characterize the traffic in a VC as it travels along the alternative path and how to assign the end-to-end QOS requirement to each link of the alternative path; note that for a call that is alternatively routed, cell loss, cell delay, etc., can occur at either of the two VPs along the route. When a VC traverses more than one VP, we assume that the traffic characteristics offered by the VC to all VPs on the route are the same, i.e., the traffic characteristics of a single VC are not changed as it passes through a VP along its route. Two classes of QOS, one for direct calls and one for alternative calls, need to be established on a VP that contains alternative routed calls. When a VC is routed on an alternative path, it requires a more stringent QOS at each VP on this alternative path, since QOS is an end-to-end measure and multiple VPs are now the end-to-end path.

*BACKGROUND STUDIES*

In [11] , two properties, heterogeneous sources (yes or no) and statistical multiplexing inside the network (yes or no), lead to four classes of ATM networks. When all sources have the same traffic characteristics and QOS requirements (homogeneous sources) and the VCs are statistically multiplexed, various Least Loaded Routing (LLR) algorithms are proposed and developed to reduce the network call blocking probability. The work assumes that the QOS requirement only involves cell loss, although a previous study by the same author [10] examined QOS requirements based on delay as well as cell losses. The interest in developing dynamic routing algorithms which incorporate the diverse nature of calls is addressed in [12] .

Other works seek to route each call in such a way that it minimizes the risk of blocking future calls, responding to the current state of the network on the basis of certain assumptions about future traffic demands. These routing schemes use information about the network state at the time of a call arrival to select a route for that call. This can be characterized as a Markov Decision Process (MDP), i.e., as the probabilistic control of a system whose load can be modelled by the evolution of a Markov chain. It is possible to define an optimum routing rule as one that minimizes the expected steady-state blocking rate in the network. There are some equations which allow us to attach a cost to each state. When a call reaches a given state, out of the possible measures, the one that leads to the state with the smallest cost should be chosen [13] .

In [20] an optimal control parameter, $\beta$, is developed by optimal control theory for ATM networks. The parameter is related to link utilization, buffer occupancy, system throughput and cell rejection characteristics. This single parameter is a function of instantaneous cell loss, buffer occupancy, utilization and user requirements. It reflects the critical elements of ATM networks and is proposed to be applicable to network performance management systems. The parameter $\beta$ is applied to route selection and admission control.

In [7] we propose anticipated calculations by CAC and adaptive routing with a pre-evaluated scheme. The behavior on a link of the selected route using the pre-evaluated scheme is verified by comparison to ordinary and optimum evaluation. Incoming connection and disconnection demands are used under different traffic characteristics assumptions.

This paper is organized as follows. Section 2 gives a detailed explanation of the proposed routing scheme based on convolution approach. Section 3 deals with simulation models and results, followed by conclusions and further works in Section 4.

## 2. Proposed Routing Algorithm

The objectives of our proposal are: to obtain a real-time connection setup with the routing algorithm, to define an optimal Routing Control Parameter, to maximize network resources utilization and to minimize call blocking probabilities. To obtain an accurate real-time response with the routing algorithm, pre-evaluated calculations of the CAC are needed. To obtain the improvement of resource utilization we propose a dynamic routing scheme with alternative routes. The Probability of Congestion (PC) in the VP (link) is the control parameter selected for routing decision. Most algorithms handle bursty services assigning an equivalent bandwidth (or effective bandwidth) to each source which reflects their characteristics, presuming a linear relationship between the capacity of the VP and its call carrying capability. This assumption implies a reduction of the multiplexing gain because the interactions between the different classes of traffic are not considered. In [8] we show that the New Convolution Approach (NCA) despite this drawback.

***NETWORK MODEL***

Given a network topology, a capacity for each physical link and a traffic requirement for each origin-destination pair, a Virtual Path Network embedded in the original network can be defined. A Virtual Path is an information transport path that can be viewed as a single logical direct link between two nodes (source and destination) with a capacity assigned. In order to isolate the advantages of the proposed routing algorithm, only dedicated bandwidth allocation strategy is considered in this paper. Other allocation strategies will be implemented in further works.

The use of VPs in ATM networks reduces the call setup delay, simplifies the hardware in the transit nodes and provides simple virtual circuit admission control. However, it also reduces the degree of capacity sharing and, thus, increases the call blocking rate for most cases. In a network using VPs, two levels of statistical multiplexing are possible. VCs are statistically multiplexed onto VPs and these VPs can be statistically multiplexed onto network links. If both these levels of multiplexing are employed (with due regard to Quality of Service requirements), the capacity required in the network will again be a minimum [4] . There is clearly a trade-off between resources spent on management and carried traffic. We assume that only the VCs statistically multiplexed onto the VP are considered. The VPs statistically multiplexed onto the link will be studied in further works.

The system of virtual paths in an ATM network can be established based on estimated traffic demands. VPN establishment is beyond the scope of this paper. We consider ATM networks where only one VP is established between every origin node of the connection demand and the corresponding destination. This requirement can seem very restricted, yet it is not so if we take into account that we are not concerned with the problem of routing between nodes of different networks (internetworking), but the fact that the origin-destination pair are inside the same network. We assume a semi-permanent VPN topology because it can be updated periodically if changing traffic demands make it necessary.

We assume that the time interval between two VPN topology changes is significantly larger than the call setup time. Under this assumption, routing of virtual circuits can be performed as if the topology of the VP networks were fixed. The VPN under consideration is not fully connected. Several VPs may share a physical link with each VP having resources assigned. We suppose that each VP has a dedicated buffer and its bandwidth is fixed and constant. The sum of the VP capacities is required to be less than the capacity of the link. The set of possible routes consists of one direct VP, or two VPs, as alternative routes, with the QOS requirements divided between the two VPs. The two VP constraint simplifies the routing implementation. The VPN is referred to as a heterogenous VPN in [11] because at least one VP supports calls that have different traffic characteristics and/or QOS requirements.

***TRAFFIC MODEL***

The question of how to multiplex two or more diverse traffic classes while providing different QOS requirements is a very complicated problem. Throughout this paper we have considering heterogeneous (multiclass) traffic which requires different QOS requirements. Each connection demand is characterized by the traffic types, the required QOS, the origin-destination pair, the interarrival time and the duration of the connection (holding time). Results by heterogeneous traffic with identical Probability of Congestion (PC) requirements are presented in [7] .

There are **ST** types of traffic modelled by General Modulated Deterministic Process (GMDP). The specification of the number of traffic types to be supported and their QOS requirements are beyond the scope of this paper. The arrival connection of traffic type **j** is assumed to follow a Poisson process with rate $\lambda_j$. The connection holding time is assumed to have an independent and exponentially distributed function with mean $\mathbf{1/\mu_j}$. We assume that when a call is denied access to the network, it does not retry and is considered lost.

## 2.1. Connection Setup Scheme

To obtain an accurate real-time response with the routing algorithm we propose to have the CAC calculations before the new connection demands arrive. Different evaluation schemes can be considered: on-line pre-evaluation scheme, store the on-line evaluations (storing the past) and off-line pre-evaluation scheme. The first two can be combined.

***ON-LINE PRE-EVALUATION SCHEME***

The interarrival time between calls depends on the type of source, and its duration varies from a few seconds to several minutes; during this time the CAC-evaluation subsystem can be idle. Also, the holding times depend on the type of source and vary from a few seconds up to hours.

The on-line pre-evaluation scheme consists of updating dynamically the status of the output VP when the CAC-evaluation subsystem is idle. For each type of source a new virtual connection must be considered. We could make the calculations in order depending on the interarrival time between calls of each type of source, i.e., sources with shorter interarrival time will be evaluated first. This process allows the system to know whether or not it is possible to allocate this hypothetical new call. Later on, the CAC-evaluation subsystem restores data structures to the real status. Using the pre-evaluated response table, it is possible to know beforehand if this selected output VP has the sufficient bandwidth available to support one new connection of a given type.

*IMPLEMENTATION*

For each output VP on the node a response vector **R(vp)** is used to store results:

$\mathrm{R(vp)} = \mathrm{R}_0, \mathrm{R}_1, \ldots, \mathrm{R}_j, \ldots, \mathrm{R}_{ST-1}$, where **ST** is the number of possible Source Types. **Rj** can be:

- **Y**ES, when a new j-type connection can be accepted because all Individual Cell Loss Probability (Individual CLP) calculated are smaller than the Individual CLP requirements. In this case, the Virtual Path Identifier can be assigned to the new connection.
- **N**O, when a new j-type connection cannot be accepted because some of the Individual CLP calculated are larger than the Individual CLP requirements.
- **U**NKNOWN, when the CAC-evaluation subsystem has not yet evaluated a new j-type connection.

The following *background process* is run to fill all the elements of the response vector R(vp) with the value YES or NO. The elements that are equal to UNKNOWN are the only ones evaluated.When an incoming call or a disconnection arrives the CAC-evaluation subsystem interrupts the background process and processes the new event. The *disconnection process* consists of the re-evaluation of all the elements of the response vector R(vp) that are equal to NO. After the call termination the responses of R equal to NO are transformed into UNKNOWN in order to be re-evaluated. Note that YES responses do not need to be updated because the system load has decreased. These responses will continue to be equal to YES. However, the *connection setup process* consists of a re-evaluation of all the elements of R(vp) that are equal to YES, so that those elements equal to YES are transformed into UNKNOWN. In this case the NO responses do not vary because the system load has increased.

If the interarrival time and the duration of the connection are long and the CAC-evaluation subsystem is still idle most of the time, several hypothetical new calls can be considered to update the status of the output VP. If the interarrival time and the duration of the connection are different for each class of service only the more frequent combinations should be considered.

***STORING THE PAST***
As the probability of congestion only depends on the current traffic demands and the value of the VP capacity assigned to the VP and as we suppose that the VPN is semi-permanent, often the same situations may be calculated. Therefore, an immediate response can be obtained if all evaluated results are stored. Moreover, the status of the output VP exhibits a locality, meaning that the status of the system does not suffer drastic changes and for a relatively long period of time the number of connections on the VP are only a relatively small fraction of all possible cases. When the entire possible status is in the memory, the CAC-evaluation subsystem will offer a response without making any calculations. If the available memory is too small to hold the entire possible status, the CAC-evaluation subsystem will make many calculations.

***OFF-LINE EVALUATIONS***
If there is sufficient memory to hold the entire possible status, all the calculations can have been made previously. This evaluation scheme is useful for comparing the performance of the proposed routing algorithms with works that can obtain a small connection setup delay because linear equivalent bandwidth is assigned.

## 2.2. Optimal Routing Control Parameter

QOS is normally expressed in terms of CLP, cell delay variation and maximum cell delay, in which the maximum cell delay can easily be imposed by having a finite buffer to limit the maximum queue length and consequently the maximum delay. Therefore, some authors, [2], [3], [5], [11], choose the CLP as the QOS measure for routing decision.

If statistical multiplexing is considered it may happen that the bandwidth required by the accepted connections exceeds the capacity of the link. The probability of the bandwidth required by the accepted connections exceeding the capacity of the link is evaluated with the Probability of Congestion (PC). For this reason the PC is a value of the link utilization and it can be related with the residual capacity used in some works. Using the New Convolution Approach the PC, the Total CLP and the Individual CLP can be calculated. In this paper the performance criteria by CAC is Individual CLP and the PC was applied to route selection.

There are a trade-off between the utilization of the VP and the cost-penalty when two VPs of an alternative route with a more stringent QOS are considered. Therefore, a Routing Control Parameter is defined. An alternative route is only considered if the PC on each VP of the alternative route is smaller than the Routing Control Parameter.

In the implementation explained above, for the possible VPs than can aceept the new connection, a numerical value indicating the PC can be stored instead of YES. When the new connection cannot be accepted the value stored is 1. When the response is unknown, which means that it has not been evaluated, any value larger than 1 must be stored. Storing the PC in the response table R, the disconnection process must be modified. The YES responses need to be updated because the system load, and therefore the PC, is smaller.

## 2.3. Alternative Routes

To obtain the improvement of resource utilization we propose a dynamic routing scheme with alternative routes. Each node must maintain a table specifying the PC of each possible output-VP and destination pair node. In dynamic routing algorithms the new connection is first offered to a direct VP, which acceps it if all Individual CLP calculated are smaller than the Individual CLP requirements. Otherwise, an alternative route is chosen. Different routing algorithms can be implemented by changing the choice from the set of alternative routes. One possibility should choose the alternative route so that the congestion on the VP is minimized, i.e., the alternative route selected is the one that has the smallest Probability of Congestion. The call is blocked if no alternative routes are available. When a demand is accepted or when there is a disconnection, the Probability of Congestion stored in the adaptive routing tables must be updated according to the new state.

By providing alternative routes, the direct route will be able to achieve lower connection blocking probabilities and higher network throughput. However, the use of alternative routes entails the utilization of other VPs and, therefore, a reduction of their capacity and an increase of their blocking probabilities. Link congestion can occur at any VP along the route and more stringent QOS requirements must be considered. Therefore, the call blocking probability increases, as because calls that can be accepted on these VPs along a direct route are rejected with the new QOS requirements. As two classes of QOS need to be established on a VP that contains alternatively routed calls, the CAC must calculate the PC for the two QOS. Moreover, if the route has several VPs, the PC calculated on the origin node of each VP must be sent to the origin node of the previous VP.

## 3. Performance of Routing Algorithm

In this section, we present computer simulations to illustrate the behavior of the routing algorithm proposed by using the PC, calculated with the New Convolution Approach, as an optimal Routing Control Parameter. Let us now consider routing-call admission control in ATM networks where a call may be setup via a number of VPs and nodes. The alternative route is limited to a maximum of two VPs. In this case, if the alternative route is selected, the new connection must be accepted in each VP with a more stringent QOS (Individual CLP / 2) What is the performance in this case? What cost-penalty will be incurred?

To answer this question the next routing algorithm is considered:

**choose** direct route **if** PC(direct route) < 1
**else** out of all available alternative routes **choose** alternative route with min(PC(alternative route))
**else reject**

where PC(alternative route) = PC(first VP) + PC(second VP)
and one alternative route is available
if PC(first VP) < Routing Control Parameter and PC(second VP) < Routing Control Parameter.

Routing algorithms differ in how they choose from among the set of allowable alternative routes. We study three algorithms: *Direct*) only direct route, each call is allowed to use exactly one VP; *1 alternative*) only one alternative route; and *2 alternatives*) choose available alternative route which has the smallest PC.

In order to isolate the advantages of the proposed routing algorithm, only dedicated bandwidth allocation strategy is considered in this paper. Another allocation strategies will be implemented on further works.

***THE SIMULATION MODEL***
The simulation model is comprised of a network consisting of many nodes and links. The sample network chosen for simulations is shown in the figure 1 where the link capacity of all the links is 50 Mbits/s.

Table 1 explains the set of possible connection demand classes, value (*) is different for each load scenario. All connection demand classes have an holding time equal to 10. Both, interarrival time and holding time have an exponential distribution. All time is expressed as a normalizated unit time.

Traffic types are modelled by General Modulated Deterministic Process (GMDP) and grouped into traffic types which are shown in the table 2. The set of possible connection demands and available routes is shown in table3.

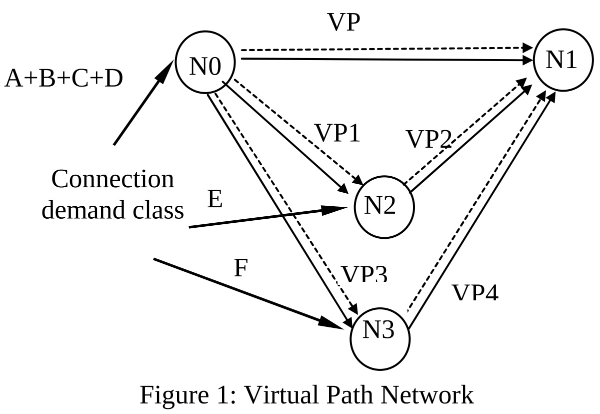


Figure 1: Virtual Path Network

| Conecction demand | origin | dest. | Traffic type | interarr. time |
|---|---|---|---|---|
| A | N0 | N1 | 0 | 0,2 |
| B | N0 | N1 | 1 | 0,8 |
| C | N0 | N2 | 0 | (*) |
| D | N0 | N3 | 0 | (*) |
| E | N1 | N3 | 0 | (*) |
| F | N2 | N3 | 0 | (*) |

Table 1: Demand Classes.

| traffic type | 0 | | 1 | |
|---|---|---|---|---|
| | **rate** | **prob.** | **rate** | **prob.** |
| **state 0** | 0 | 0,9 | 0 | 0,6 |
| **state 1** | 4 | 0,1 | 4 | 0,4 |
| **mean rate** | 0,4 Mbit/s | | 1,6 Mbit/s | |
| **burstiness** | 10 | | 2,5 | |
| **CLP req.** | 1.0 E-5 | | 1.0 E-6 | |

Table 2: Traffic types.

| conecction demand | direct route | set of allowable alternative routes |
|---|---|---|
| **A,B** | VP0 | {VP1+VP2 ,VP3+VP4} |
| **C** | VP1 | -- |
| **D** | VP3 | -- |
| **E** | VP2 | -- |
| **F** | VP4 | -- |

Table 3: Routing Table.

***OFF LINE EVALUATIONS***

If the Probability of Congestion of all possible combinations has been precalculated before starting the routing simulation, the connection setup delay does not influence the results of the routing algorithms

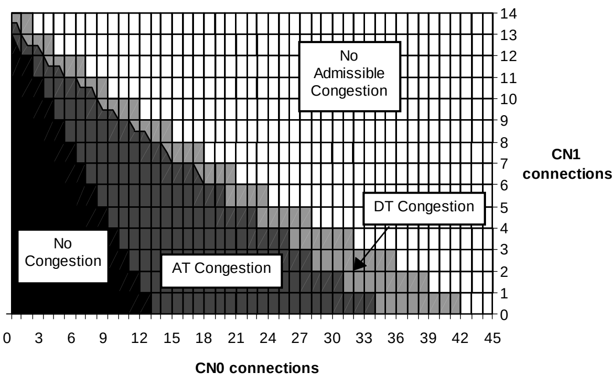


Figure 2: Probability of Congestion (CLP/2).

Criteria by CAC: Individual CLP requeriments / 2.

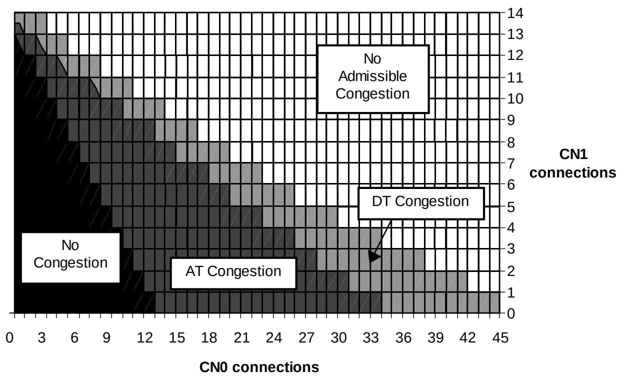


Figure 3: Probability of Congestion (CLP).

Criteria by CAC: Individual CLP requeriments.

In figures 2 and 3, we compare the maximum number of traffic type 0 and traffic type 1 connections that can be supported on a link with 50 Mbits/s of capacity under different Individual CLP requirements according to table 2. In figure 2 this requirements are more stringent. A point (CN0,CN1) on the shadowed area can be interpreted as follows: CN0 traffic type 0 connections and CN1 traffic type 1 connections can be supported to guarantee the required QOS. On black shadowed area the capacity required does not exceed the link capacity. On AT & DT congestion areas the offered traffic exceeds the capacity of the link but the connections are accepted. Moreover, on AT congestion area the PC is less than the Routing Control Parameter, connections over VPs with alternative traffic are admitted. But on DT congestion area only direct traffic is considered. The Routing Control Parameter considered in the figure is 1.0 E-5, and therefore the AT congestion area is the same in both figures. In the white area (no admissible congestion) connections can not be supported with the Individual CLP requirements.

Figure 4 shows the cost, in time units, needed to evaluate the Probability of Congestion using the NCA when CN0 connections and CN1 connections are considered. When the point (CN0,CN1) is not accepted, (CN0 + i, CN1) and (CN0, CN1 + j) are ignored. The figure is plotted by Individual CLP requirements. Due to space limitations, figures by Individual CLP requirements / 2 were omitted.

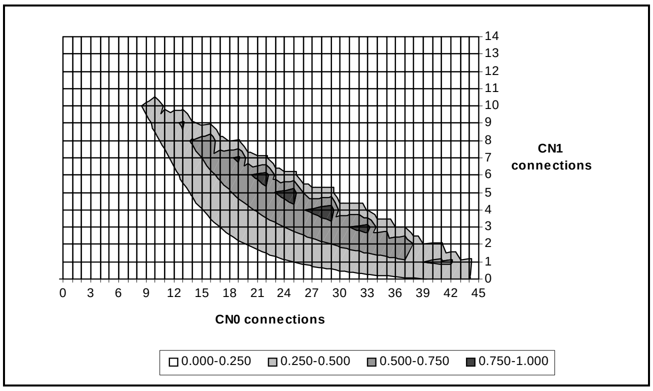


Figure 4: Criteria by CAC: Individual CLP requeriments.

***LOCALITY***

Figure 5 shows the number of instances that the system holds in a given state during the simulation. An state is defined by the number of existing connections of each type. This system state varies dynamically according to the load.

As figures 6, 7, and 8 show, the system state is normally in a reduced subset of all possible states. Therefore, the system state moves in a small area.

Locality implies different effects depending on the Connection Setup Scheme. For the On-Line Pre-Evaluation Scheme, locality permits to the CAC-evaluation subsystem to calculate the new state because there is a great probability that the next change of the system will be close to the real state.

In case of Storing the Past and Off-Line Evaluations, locality permits storing a reduced set of evaluations. Fig 6 shows a large period of simulation and the needed area is smaller than the area for all possible states[4] In this example, for type-0 traffic, the mean number of connections is 36 being standard deviation 5, and for type-1 traffic the mean is 3 being standard deviation 1.

The results shown above correspond to a stationary load of the system, but the global status of the system changes slowly during long periods of time, and the CAC-evaluation subsystem can accomodate to these changes.

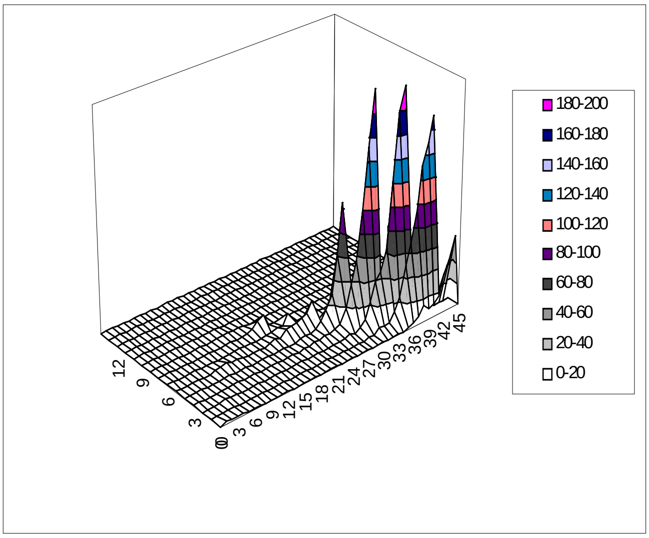


Figure 5: Locality, all period.

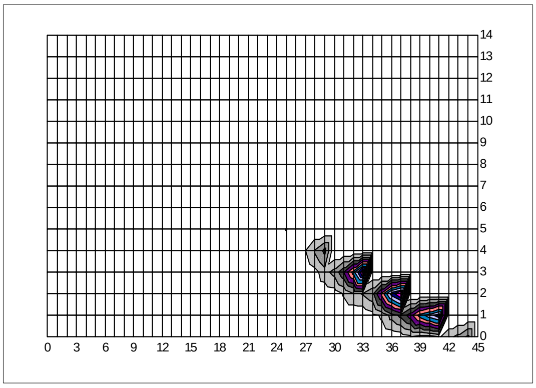


Figure 6: 75 %.

[4] See figure 2 and 3.

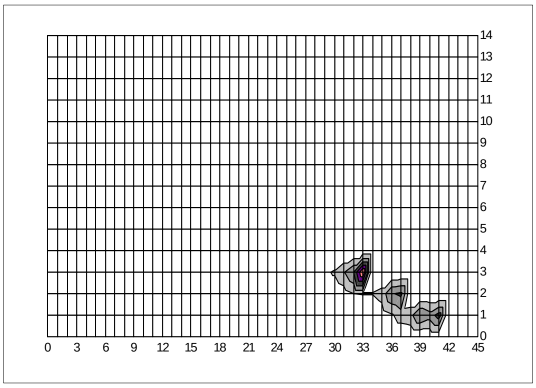



Figure 7: 30%.

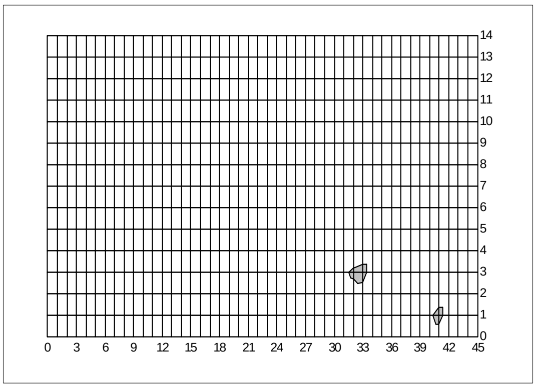



Figure 8: 10%.

The performance of the network varies with the load of the system, with the routing algorithm and with the value of the Routing Control Parameter.

Figure 9 shows the results of simulations for homogeneous traffic. In this case connection demand B is not considered.. All connections demands have the same holding time. The period of each test equals 1500 unit times.

Different values to Routing Control Parameters are considered: 1.0E-6, 5.0E-6 and 1, as third value we can considered any value larger than the maximum PC admited. The legends show the results by different routing algorithms: *Direct*, *1 alternative* (N0-N2-N1) and *2 alternative* (N0-N2-N1 and N0-N3-N1).

When the routing algorithm is only direct route the Blocking Probability is a constant value because alternative routes are not considered and the Routing Control Parameter is ignored. The Blocking Probability of *Direct* decreases with the Interarrival Time.

When the interarrival time equals 0.3 time units, the routing algorithm and the value of the Routing Control Parameter have little influence. The Blocking Probabilities obtained are very similar to the Blocking Probability obtained with the direct route. When the interarrival time equals 0.35 time units the routing algorithm and the value of the Routing Control Parameter have great influence. The Blocking Probability decreases with the use of alternative routing schemes and it also decreases when the value of the Routing Control Parameter equals 1 due to the low network load. When interarrival time equals 0.25, the network is more loaded and the Blocking Probability depends on the routing algorithm and on the Routing Control Parameter.

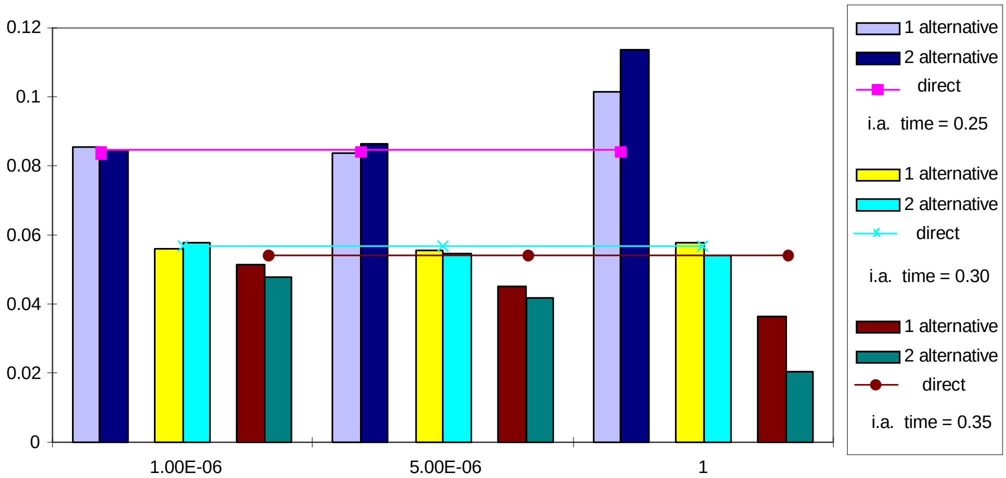

Figure 9: Blocking Probability vs. Routing Control Parameter by different Routing Algorithm and different Interarrival Time

In figure 9, Blocking Probability of the network for all VPs are shown, but VPs have diferent Blocking Probabilities. Figures 10 and 11 show the Individual Blocking Probability for each VP, when the interarrival time is 0.35 in an homogeneous scenario

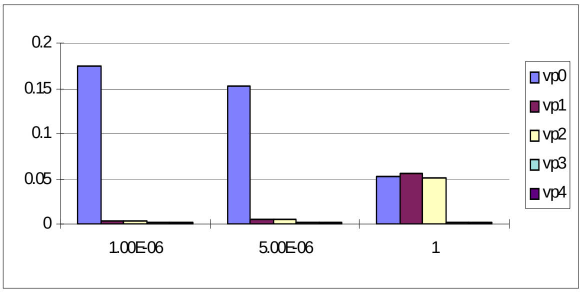


Figure 10: Blocking Probablity 1 alternative

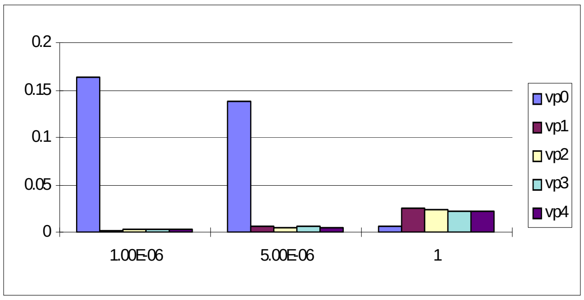


Figure 11: Blocking Probablity 2 alternatives

As expected, using alternative routes the Blocking Probability of VP0 decreases, but depending on the Routing Control Parameter, the Blocking Probability of the alternative VPs used increases dramatically. This drawback is shown by figure 11, when the Routing Control Parameter equals 1, the Blocking Probability of the alternative VPs is larger than the Blocking Probability for VP0.

## 4. Conclusions and Further Works

We have proposed an implementation for dynamic routing based on anticipated calculation. We suggest an algorithm based on the Probability of Congestion in the link as control parameter for routing decision. The Probability of Congestion is evaluated using the NCA, which permits real-time calculation. The main objectives of our proposal are: to maximize network resources utilization and to minimize call blocking probabilities.

Differents routing algorithms are compared. Results show different network performance by changing the Routing Control Parameter and the load of the network, especially the load of the alternative route. Using alternative routes does not always decrease the blocking probability of the network.

One important aspect of future works is to determinate the best Routing Control Parameter based on the load of the network. Also, other allocation strategies and VPs statistically multiplexed onto the link will be studied in the future. We are continuing our investigations on routing in different ways: First, working with more realistic approaches on other traffic mixes. Second, by looking at the performance of these algorithms for centralized and distributed routing. Moreover, we are working in the application of these algorithms in a multicast implementation where results show that the use of pre-computed routing tables improves the performance of the multicast.


***ACKNOWLEDGEMENTS***

The authors thank Dr. Ian F. Akyildiz, Georgia Institute of Technology (Atlanta), for his many constructive observations and comments in relation with this work.


## 5. References


[1] S. Ahn and others. “Virtual Path Layout Design On ATM Networks”. INFOCOM’94, pp. 192-200

[2] A. Atkinson. “A Traffic Control Schem for Virtual Paths in an Asynchronous Transfer Mode Network”. GLOBECOM’94, pp. 1768-1772

[3] S. Bahk and M. El Zarki. “PREVENTIVE CONGESTION CONTROL BASED ROUTING IN ATM NETWORKS”. ICC’94, pp. 1592-1599

[4] J. Burgin and D. Dorman. “Broadband ISDN Resource Management: The Role of Virtual Path”. IEEE Communications Magazine. September 1991. pp. 44-48

[5] J.H.S. Chan and H.K.Tsang “Bandwodth Allocation of Multiple QOS Classes in ATM Environment”. INFOCOM’94, pp. 360-367

[6] K.T. Cheng and F. Y.S. Lin. “On the Joint Virtual Path Assignment and Virtual Circuit Routing Problem in ATM Networks”. GLOBECOM’94, pp. 777-782

[7] R. Fabregat-Gesa, J. Sole-Pareta, J.L. Marzo-Lázaro and J. Domingo-Pascual. “Adaptive Routing Based on Real-Time Calculations Using the Convolution Approach.”. EFOC’94, nº 69, pp. 88-92

[8] R. Fabregat-Gesa, J. Sole-Pareta, J.L. Marzo-Lázaro and J. Domingo-Pascual. “Bandwidth Allocation Based on Real Time Calculations Using the Convolution Approach”. GLOBECOM’94, pp.

[9] R. Guerin. and others. "Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks". IEEE Journal on Selected Areas in Communications, Vol 9, Nº 7, September 1991, pp 968-981

[10] S. Gupta, K.W. Ross and M. El Zarki. “Routing in Virtual Path Based ATM Networks”. GLOBECOM’92, pp. 571-575

[11] S. Gupta, K.W. Ross and M. El Zarki. “On Routing in ATM Networks”. Modeling and Performance Evaluation of ATM Technology IFIP transactions (C-15). pp. 229-239

[12] S. Gupta and P.P. Gandhi. “Dynamic Routing in Multi-Class Non-Hierarchical Networks”. ICC’94, pp. 1390-1394

[13] R.H. Hwang, J.E. Kurose and D. Towsley. “MDP Routing in ATM Networks Using the Virtual Path Concept”.INFOCOM’94. pp. 1509-1517

[14] J.L. Marzo-Lázaro, R. Fabregat-Gesa, J. Domingo-Pascual and J. Sole-Pareta. "Fast Calculation of the CAC Convolution Algorithm using the Multinomial Distribution Function". Tenth UK Teletraffic Symposium. 1993. pp.23/1-23/6.

[15] A. Kolarov and J.Hui. “Least Cost Routing in Multiple-Service Networks”. INFOCOM’94. pp. 1482-1488

[16] K.R. Krishnan and R.H. Cardwell. “Routing and Virtual-Path Design in ATM Networks”. GLOBECOM’94.

[17] F.Y.S. Lin and J.R. Yee. “A Real-Time Distributed Routing and Admission Control Algorithm for ATM Networks”. INFOCOM’93. pp. 792- 801.

[18] K. Mase and S. Shioda. “Real-Time Network Management for ATM Networks”. ITC-13 pp. 133-140

[19] S. Ohta and K.I. Sato. “Dynamic Bandwidth Control of the Virtual Path in an Asynchronous Transfer Mode Network”. IEEE Transaction on Communications, Vol. 40, nº. 7, July 1992.

[20] T.M. Oser, X. Gu and D. Vaman. “Routing with Admission Control in ATM Networks”. GLOBECOM’94. pp. 1212-1216.

[21] E.M. Spiegel and T. Murase. “An Alternate Path Routing Scheme Supporting QOS and Fast Connection Setup in ATM Networks”. GLOBECOM’94. pp. 1224-1230.

[i] MIRAR EN EL FICHERO TRB_ROUT.DOC EL APARTADO CD [ROUTING ALGORITHM] Y [HOP]

[ii] [BAH94] pp. 1592 (5) deals with/gives a / introduces
[GUP94] pp. 1390
[HYM94] pp. 165

[iii] (When arriving call differ in their bandwidth requirements, it is desirable to employ a routing policy that achieves two main networks objectives. They are (i) efficient sharing of network resources in order to maximize their utilization and (ii) isolation of classes in order to minimize interaction between calls of different classes ) [GUP94] pp. 1391

... the characteristics of load balancing and reduced routing oscillations ... [BAH94] pp. 1593

The performance objetive of the routing is to minimize the control penalty, cell delay and cell rejection [OSE94] pp. 4

[iv] Traffic expressed in Erlangs [ARV94] pp. 10/2

[v] EXPLICAR MEJOR LO DE LA LOCALIDAD. A MEDIDA QUE LLEGAN NUEVAS CONEXIONES / DESCONEXIONES EL NUMERO DE CONEXIONES ACEPTADAS EN EL SISTEMA VA VARIANDO, COMO ESTAS VAN LLEGANDO DE UNA EN UNA, NO SE PRODUCEN CAMBIOS BRUSCOS Y DURANTE UN INTERVALO DE TIEMPO EL NUMERO DE CONEXIONES ACEPTADAS SE MANTIENE DENTRO DE UN RANGO.

[vi] MIRAR EL FICHERO PCONG.DOC

[vii] **REPASAR !!!!!!** The system must now be so designed such that in addition to guaranteeing the QOS (bounds on ~~cell delay and~~ loss probabilities) of each individual call, it must also guarantee the grade of service for each source type (given by the call blocking probabilitis), i.e., the call blocking probability for calls of source type s should be kept below a given $\varepsilon_i$, ($\varepsilon_i$ >0). [GUP93b]

The QOS of VP must meet the most demanding QOS of the connection carried.

If there some VP onto a link, each with a number of accepted connections by a determined PC, the PC of all the connection over the link is smaller that the fixed PC of the VP.

**OCURRE LO MISMO PARA LA PROBABILIDAD DE PERDIDAS ???**
**INTENTAR RELACIONAR LA PC CON LA CAPACIDAD RESIDUAL.**