Optimal Cost/Performance Design of ATM Switches

Paolo Coppo, Matteo D'Ambrosio, and Riccardo Melen, Member, IEEE

Abstract—This paper proposes a methodology for performing an evaluation and optimization of the cost of an ATM switching architecture under performance constraints given in terms of virtual connection blocking probability.

An analysis of blocking networks is developed, and combined with known results concerning nonblocking networks, provides a theoretical model which relates traffic characteristics, network topology and blocking probability in a multirate/multiservice broadband environment.

An analysis of the characteristics determining the cost of a generic ATM switch implementation follows. The model is oriented to optimize both the topological parameters and the speed advantage, with respect to the main cost factors of VLSI-based switching networks i.e., components count and complexity, interconnection costs.

I. INTRODUCTION

ALTHOUGH in these last years a great variety of studies has been carried out on several aspects of the ATM technique, very little work has been done on the topics of blocking probability and cost evaluation.

This paper is focused on the analysis and simulation of the performance of large broadband ATM switches in terms of virtual connection blocking probability; this kind of analysis applies to an important class of buffered ATM switching structures proposed in the literature and being implemented by several switch manufacturers. In a second part of the paper the parameters determining the cost of a generic ATM switch implementation are considered; an attempt is made at defining a cost measure having technological relevance, i.e., related to the main hardware costs of modern digital switching networks: VLSI components and interconnection. An example is developed to illustrate theoretical issues and to provide the basis for an integrated cost—performance evaluation.

A. Research Scope

Scope of the research is a particular class of ATM switches, i.e., those with large multistage buffered interconnection networks, providing multiple paths between any inlet/outlet pair. In the paper we point our attention toward Clos networks, but the methodology is general, and can be immediately applied to other multistage topologies. The practical importance of this family of networks in the ATM field is demonstrated by their use in most of the current switch prototypes [1], [12], [14], [17], [29], [30].

Multistage networks are mandatory even for small size switches, due to technology limitations, topological considerations and cost. In some cases multipath fabrics employ internal datagram routing, i.e., each cell is independently routed to its destination, while in other solutions all cells of a particular connection follow the same path ("connection oriented" fabrics): in these cases virtual connection blocking is possible.

Although the theory of call blocking in circuit switching telephone networks is well established, these results can not be applied directly to derive virtual circuit ATM blocking probability, as the multirate environment introduces new variables that lead to different results.

In our analysis we do not take into account problems related to cell loss probability, bandwidth allocation for variable bit rate connections and real time back pressure congestion control. In the following sections, we will assume that each connection requires a certain amount of bandwidth, either being the actual bandwidth of a constant bit rate (CBR) connection, or the "effective" bandwidth of a variable bit rate (VBR) connection determined by some resource allocation algorithm.

Congestion control is realized by the resource allocation policy, assigning a certain bandwidth to VBR connections; policing devices located at the user interface are responsible for enforcing this resource allocation policy, while buffers and back pressure mechanisms are used to absorb temporary statistical overloads.

B. Methodology

The methodology followed in this paper is aimed at deriving results useful to the system engineer starting from the specification of the quality of service (QOS) required.

First of all, a theoretical analysis is needed relating traffic characteristics, network topology and blocking probability in a multirate/multiservice broadband environment. Both known results concerning non blocking networks and the analysis of blocking networks are considered in the paper.

Various simulations have been performed to validate theory, and upper bound approximate formulas have been derived when considered relevant.

These studies can be used to choose the basic switch architecture and engineer it to match services requirements; an example is used to illustrate these issues.

A parametric cost analysis of the ATM switch implementations in a reference technology must then be performed; in the example reported in this paper data have been derived from known experiments and scalability considerations.
A global optimization of the cost of an implementation providing a certain QOS links the two aspects of the analysis. A final example thoroughly reviews all covered issues.

II. VIRTUAL CONNECTION BLOCKING
IN ATM SWITCHING NETWORKS

ATM multistage networks with internal routing determined at connection setup must be studied also in terms of their virtual connection blocking characteristics. For circuit switching networks a well established theory exists which allows to characterize the blocking properties of a network [2] (i.e., if blocking states can arise and in which conditions) and its blocking probability [32]. The foundations of this theory rely on the possibility of modeling the occupancy state of a circuit with a binary variable assuming the states "busy" and "idle."

Moreover, because it was developed to model space division electromechanical switches, a cost function was associated to it that defines the cost of an N by M module to be $E = NM$, the crosspoint count, the overall cost of the network being the sum of the costs of its modules.

With broadband systems based on ATM the basic hypotheses related to circuit switching have to be changed. The foundations of a theory capable of describing blocking properties of ATM fabrics [23]-[25] will be reviewed in the next subsection; the extension of the techniques used for estimating the blocking probability of a blocking network presents more problems, and a few results will be presented in a following subsection. The problems caused by this assumption and the reasons for maintaining it will be discussed in the last section of the paper.

A. Nonblocking Networks

New aspects need to be considered. First, an ATM fabric is multirate: each connection induces a load on an ATM multiplex which depends on its bandwidth characteristics. This will be modeled by associating a weight $0 \leq \omega \leq 1$ to each connection; the maximum load carried by an internal fabric link will be 1, so that a new connection with weight $\omega$ will be able to pass through a link if the load that it is already carrying is $\leq 1 - \omega$.

Two possible constraints on the traffic will be considered. First, the weight of the connections will be restricted to the interval $(b, B)$. Also, the sum of the weights of connections involving an input or an output will be limited to $\beta$. Note that $0 \leq b \leq B \leq \beta \leq 1$. The practical effect of a restriction on $\beta$ is to require that the internal data paths of a network operate at higher speed than the external ports. In other terms, $\beta^{-1}$ is the speed advantage of the fabric. Two particular choices of parameters are of special interest. The traffic condition characterized by $B = \beta$ and $b = 0$ will be referred to as unrestricted packet switching, and the condition $b = B = \beta = 1$ as pure circuit switching. It is expected that theorems for the multirate case include known results for circuit switching as a special case.

A network will be strictly nonblocking for particular values of $b$, $B$ and $\beta$ if for all sets of connections for which the connection weights are in $(b, B]$ and the input/output weights are $\leq \beta$, the network cannot block. The definitions of rearrangeably nonblocking and wide-sense nonblocking networks are extended similarly.

A three-stage Clos network with $N$ inputs and $N$ outputs (see Fig. 1) has $N/n$ modules of size $n \times m$ at the first stage, $m$ modules of size $N/n \times N/n$ in the central stage and $N/n$ modules of size $m \times n$ in the third stage; it will be indicated with $C(N, n, m)$. For this topology, the following theorem holds in a multirate environment: a three-stage Clos network is strictly nonblocking if

$$m > 2 \max_{b \leq 2n} \left\{ \lim_{\epsilon \to 0} \frac{\beta n - \omega}{\max(b, 1 - \omega + \epsilon)} \right\}$$

where $[\alpha]$ is the integer part of $\alpha$.

Two things must be noted: first, in the pure circuit switching case the condition is $m \geq 2n - 1$, exactly as in [5]; second, for unrestricted packet switching we have $m \geq 2 \frac{n^2}{1 - \beta} (n - 1)$, an expression which goes to infinity for $\beta = 1$. This is true in the more general framework of strictly nonblocking networks [24].

Two ways are left to obtain nonblocking networks at reasonable cost: internal speed advantage (i.e., $\beta < 1$) and traffic segregation creating subnetworks where $b > 0$ or $B < 1$. For instance, in a three-stage Clos network if half of the middle stage switches are reserved for "large" connections ($\omega > 1/2$) and half for "small" connections ($\omega < 1/2$) the network is nonblocking if $m = 4n$: note that in this way a wide-sense nonblocking network is obtained. It is clear, and this qualitative consideration carries over to blocking fabrics, that in multistage connection oriented fabrics the flexibility of ATM does not come for free: larger or higher speed networks are necessary to obtain the same performance of circuit switching.

In the rest of the paper the result above will be used. Several other related results are reported in [23]-[25].

B. Blocking Probability in Multirate Networks

Several analytical methods have been developed for evaluating connection blocking probability in circuit switching [32]; among these Lee's method is the most popular and conceptually the simplest [19]. It is based on a static model of a switching network and gives conservative bounds on blocking. This method can be easily extended to deal with broadband communication networks. In this subsection we will consider a model of multirate networks, at which Lee's method can be immediately applied; we will also introduce some modifications of such method leading to more accurate results.

![Fig. 1. Clos network C(N, n, m).](image-url)
The application of Lee's method requires that the probability distribution of the link occupancy be known. Assuming that the traffic offered to a link was composed of several service classes, with Poisson arrivals and general service times, it is possible to derive the link occupancy distribution using an easy recursive numerical computation scheme [15, 28, 8, 21]. Here we are proposing an alternative method to evaluate the link occupancy distribution based on the carried load and not on the offered load, therefore correctly matched with the assumption underlying Lee's method.

Let us consider a network having internal links carrying \( k \geq 1 \) (\( k \) is an integer called "multiplexing factor") units of bandwidth; we suppose that the network supports a traffic formed by services with bandwidth multiple of \( 1/k \). Let us indicate a channel of bandwidth \( 1/k \) as minimal channel. We say that a connection with capacity \( w/k \) (\( w \) integer such that \( 1 \leq w \leq k \)) has multiplicity \( w \), or that it is a \( w \)-connection. Moreover, for the \( w \)-connection class, it is useful to consider a link as a set of \( \left[ \frac{k}{w} \right] \) identical resources or channels of multiplicity \( w \) (or \( w \)-channels). It is quite clear that, to realize a \( w \)-connection between an input and an output of the network, it is necessary to allocate a path having a free \( w \)-channel on each of its links. The blocking probability of a \( w \)-connection is the probability that such a path cannot be found.

To determine the blocking probability of a link for a \( w \)-connection we need the link occupancy distribution; this one depends on the bandwidth distribution of the incoming connections. Proceeding in analogy with Lee's derivation for circuit switching, we will assume the independence of the \( w \)-channels when considering the class of \( w \)-connections, and we will deduce the occupancy distribution of the internal links of a switch, when the network background traffic is composed of \( v \geq 1 \) service classes. Let \( \lambda \) be the average traffic load due to all the service classes. Let \( \alpha_i \) the share of class \( i \) (\( i = 1 \)) \( \sum \alpha_i = 1 \), \( X_i \) the random variable representing the number of class \( i \) connections on the link, \( r_i \) their multiplicity, \( k \) the link multiplexing factor and \( Y \) the random variable representing the link occupancy.

The method described here is based on a procedure to generate a random state of the link: we order the service classes according to their multiplicity: let then be \( r_1 > r_2 > \cdots > r_v \). According to the notation of the previous section, here the traffic is restricted to the interval \( [b, B] \) where \( b = r_v/k \) and \( B = r_1/k \). At first, we consider the link as an aggregate of channels of multiplicity \( r_1 \); each channel is busy with probability \( \alpha_1 \lambda \); this corresponds to the allocation of connections of multiplicity \( r_1 \) for an average load equal to \( \alpha_1 \lambda \); then we allocate in the same way the connections of multiplicity \( r_2 \) for an average load equal to \( \alpha_2 \lambda \) and so on till the connections of multiplicity \( r_v \). This procedure guarantees that the constructed state bears a traffic correspondent on the average to that specified. Had we allocated the connections in any other order, there would have been no guarantee that the larger connections could have been loaded over the link proportionally to their quota. It is obvious that, in a real network, the call allocation process will lead to different distributions, but this procedure is basically a tool to generate random states corresponding to certain carried (not offered!) load distributions. The analysis can follow the same path.

Let \( Y_u \) be the random variable representing the link state after the allocation of the connections belonging to the \( u \) classes with larger multiplicities:

\[
Y_u = \sum_{i=1}^{u} r_i X_i \quad 1 \leq u \leq v
\]  

Obviously \( Y_v = Y \). The distribution of \( Y_1 \) is:

\[
P[Y_1 = y_1] = \begin{cases} 
\frac{\binom{y_1}{\left[ \frac{k}{w} \right]} x_1 \left(1 - \frac{x_1}{y_1} \right)^{\frac{x_1}{y_1} - 1}}{\binom{\left[ \frac{k}{w} \right]}{1}} & \text{if } y_1 \mod \left[ \frac{k}{w} \right] = 0 \\
0 & \text{elsewhere}
\end{cases}
\]

Assuming a binomial distribution for the random variable \( X_1 \), we have:

\[
P\left\{ X_1 = \left[ \frac{y_1}{\left[ \frac{k}{w} \right]} \right] \right\} = \frac{\binom{\left[ \frac{k}{w} \right]}{1} \left(1 - \frac{x_1}{y_1} \right)^{\frac{x_1}{y_1} - 1}}{\binom{\left[ \frac{k}{w} \right]}{1}}
\]

Note that the mean traffic loaded on the link after this first step is exactly \( \alpha_1 \lambda \) if \( k \) is a multiple of \( r_1 \) (closer agreement can be obtained if all the multiplicities \( r_i \) are submultiples of \( k \) and \( r_1 \) is a multiple of \( r_j \) for \( j > i \)).

Let us assume that the connections of the first \( u \) classes have been loaded and the link occupancy distribution at this step, \( P\{Y_u = y_u\} \), has been calculated.

In the following step the connections of multiplicity \( r_{u+1} \) must be allocated. Let \( P\{X_{u+1} = x_{u+1}|Y_u = y_u\} \) be the probability of adding \( x_{u+1} \) connections of class \( u+1 \) when there are already \( y_u \) minimal channels busy on that link. At this point there are \( (k - y_u) \) minimal channels free for the connections of class \( u + 1 \). We indicate with \( \pi_{u+1} \) the probability that a channel of multiplicity \( r_{u+1} \) among those that are idle is allocated at this step on the link: that is:

\[
\pi_{u+1} = \frac{\alpha_{u+1} \lambda}{1 - \sum_{i=1}^{u} \alpha_i \lambda}.
\]

The distribution of \( X_{u+1} \) conditioned to \( Y_u \) is assumed again to be a binomial one; thus:

\[
P\{X_{u+1} = x_{u+1}|Y_u = y_u\} = \binom{k-y_u}{x_{u+1}} \pi_{u+1}^{x_{u+1}} (1 - \pi_{u+1})^{k-x_{u+1}}
\]

Now we can calculate the distribution \( P\{Y_{u+1}\} \):

\[
P\{Y_{u+1} = y_{u+1}\} = \sum_{S} P\{X_{u+1} = x_{u+1}|Y_u = y_u\} P\{Y_u = y_u\}
\]

where \( S = \{x_{u+1}, y_u\} \).

The distribution \( P\{Y_{u+1} = y_{u+1}\} \) at the last step is the link occupancy distribution. From this distribution, it is immediate to find the blocking probability of the link \( l \) for connections of multiplicity \( w \):

\[
p_l(w) = \sum_{y=k-w+1}^{k} q(y).
\]
Then Lee's method can be applied like in the case of circuit switching [19], [27]. Considering for instance a three stage Clos network $C(N, n, m)$ (see Fig. 1), its internal blocking probability for a connection of multiplicity $w$ is:

$$P_B[C(N, n, m), w] = (1 - (1 - p_l(w))^2)^m.$$  

With this expression we can obtain upper bounds for the blocking probability; often, there can be a relevant difference between these values and real ones. This occurs because this method does not take into account the statistic dependence of a stage on the previous one. To obtain more precise results the method can be modified introducing the reduced load $\lambda_r$: the reduced load is the channel occupancy probability that we obtain leaving out the load of the connection to be routed from the switch at stage $s$; if a connection of weight $\omega$ ($0 < \omega \leq 1$) must be routed through a link coming out of an $N \times M$ crossbar at stage $s$ of the network, when the mean load at the inlets is $\lambda$, we have:

$$\lambda_r = \frac{N \lambda - \omega}{M}.$$  

The reduced load can be used in the method just described instead of $\lambda$, and then the probability $p_l(w)$ becomes a function of the switch size stage by stage; this technique is an attempt at modeling the statistical dependencies of the load on adjacent stages, and offers more accurate results than the original method, also for circuit switching networks ($k = 1$).

To summarize, we have described a method that is useful to deduce the link (and switching network) blocking probability for connections of multiplicity $w$ when the following parameters are specified: the link multiplexing factor $k$, the total traffic load $\lambda$ carried on the link and, for each service class $i$, its traffic share $\alpha_i$ and its multiplicity (bandwidth) $r_i$.

1. Results: The analytical model was verified by simulation. For this purpose, a three-stage Clos network $C(128, 8, 8)$ has been chosen; this network has no expansion and generates relatively large values of the blocking probability in the simulations. The complex nature of the problem considered and the dependence of the results on a large number of assumptions lead us to the decision of using two different simulators. The former is a static simulator; it generates random states of the network one at a time allocating connections with the procedure followed in the analysis and random routing. The latter is a discrete event simulator. Its input parameters are the mean interarrival time and the mean holding time for each service class (the distributions of interarrival and holding times are assumed to be exponential). The traffic load $\lambda$ is measured inside the network itself during the simulation. For multirate networks the blocking probability was found to be extremely sensitive to the routing policy (see also [33]). Under the “random routing” hypothesis the analytical results were found quite accurate with respect to the simulations, especially when using the “reduced load” $\lambda_r$. For interested readers, a detailed comparison of analysis and simulation results are reported in [6]. A few results on the routing policy are also discussed there.

It must be noted here that the connections of higher multiplicity have the higher blocking probability, as it is shown in Fig. 2, which illustrates the blocking behavior of a Clos network $C(128, 8, 8)$ with 34 Mbit/s links and three classes of connections (64 Kbit/s, 1 Mbit/s and 2 Mbit/s). For the purposes of network designers, it is therefore sufficient to guarantee a given blocking probability requirement for the connection class with the largest bandwidth foreseen in the network.

III. COST AND IMPLEMENTATION MODEL

As already noted above, the original theory of interconnection networks is based on a cost measure—the number of crosspoints, which is clearly obsolete in the VLSI era, although it can still be useful as a guidance in preliminary comparisons.

It would be very useful in the analysis of ATM fabrics to develop a technologically relevant cost measure and to adopt it together with the theory of multirate interconnection networks outlined above, thus building a consistent theoretical framework for the evaluation of such architectures. The model we are seeking should optimize both the topological parameters and the speed advantage, with respect to the main cost factors of VLSI-based switching networks: component count and complexity, and interconnection cost. To our knowledge, no satisfactory analysis of this kind has yet been carried out for ATM architectures, and the technical difficulties involved, especially in relating the cost measure to technology while keeping it reasonably simple are such that it would probably require an excessive effort to be developed in its full generality.

It is, however, felt that in many cases even a very approximate analysis can lead to interesting results, which help to understand the design tradeoffs which cannot be completely explored by exhaustive simulations or laboratory experiments.
have that the total number of circuits of area $A$ to be processed per working component is $r^{-A}$. Our assumption will be that the cost of a single VLSI component is proportional to $Ar^{-A}$, and that its total area is composed of a part $A_b$ proportional to the size of the buffers and a part $A_p$ proportional to the number of ports. Because of the buffer sharing, for a given design load the memory requirements per output decrease with the increase of $\mu$. We have found that the percentual saving of memory with respect to the case of independent buffers can be approximated for a wide range of loads by the function $f'(\mu) = \frac{0.05\mu+2.45}{\mu+1.5}$ (in other words $A_b^{\text{shared}} = f'(\mu)A_b^{\text{independent}}$). Supposing that in our example technology the two parts of the area are equivalent for $\mu = 8$ ($A_b(8) = A_p(8)$), the cost of a $\mu \times \mu$ shared buffer component will be:

$$C_{\text{VLSI}} = k_s f(\mu) N^{-2f(\mu)}$$

where

$$f(\mu) = f'(\mu) + f'(8) = \frac{0.32\mu + 1.9}{\mu + 1.5}$$

Assuming that the total number of components per module grows quadratically (for instance like in Fig. 3) and that the speed advantage is supported by some form of bit slicing, we are in the position to give approximate expressions for the cost of the various parts of the switching network.

IV. INTEGRATED COST/PERFORMANCE EVALUATION: AN EXAMPLE

In this section we are going to illustrate how theoretical principles can be derived to choose a good cost/performance compromise. The following example deals with the optimization of an ATM switching fabric (excluding the header translation functions). It is admittedly oversimplified for the sake of analytical tractability, but it gives some interesting information about the interaction of the key design parameters when a VLSI-oriented cost measure is used.

First we will consider a nonblocking three-stage Clos network employed in an unrestricted packet switching scenario. In a second time a more realistic scenario with given traffic specifications will be assumed, and optimal blocking and nonblocking architectures will be obtained. For a three-stage Clos network $C(N, n, m)$ the following expressions give the cost of the various parts of the network (obtained by counting the switching modules in each stage and supposing for each one an implementation as in Fig. 3):

$$C_{\text{stage I}} = C_{\text{stage III}} = k_s \frac{N \mu}{n} \frac{m}{\beta} f(\mu)^{-2f(\mu)} \mu$$  \hspace{1cm} (11)

$$C_{\text{stage II}} = k_s m (\frac{N}{n})^2 \frac{1}{\beta N \mu^2} f(\mu)^{-2f(\mu)} \mu$$  \hspace{1cm} (12)

$$C_{\text{interconnection}} = k_f \frac{2N}{\beta} (1 + \frac{m}{n})$$  \hspace{1cm} (13)

A three-stage Clos network employed in an unrestricted packet switching scenario is nonblocking if $m > \frac{N}{2} (n - 1)$, as noted above. Let us assume, with little approximation, $m = \frac{N}{2} (n - 1)$. At this point we have the topological parameter $n$ and the speed advantage factor $\beta$ to play with in the optimization.
process. Considering that \( m = \frac{2\beta}{1-\beta} n \), the normalized cost per port is:

\[
C_{SW} = \frac{C_{SW}^*}{Nk_i} = \frac{2}{1-\beta}(\frac{2n + N}{n})^{-1} f(\mu) r^{-1} f(\omega) \mu + \frac{1 + \beta}{\beta(1 - \beta)} r \frac{1 + \beta}{\beta(1 - \beta)}
\]

where \( k_i = k_i^*/k_s \).

To determine the yield parameter \( r \) and the constant \( k_i \), let us suppose arbitrarily that for \( \mu = 8 \) the yield is 10%, and that for a switch with \( N = 1024 \), \( \beta = 1/2 \), \( n = 32 \), \( m = 64 \) the two components of the cost are the same. In this case we have \( r = 0.619 \) and \( k_i = 24 \).

The whole point of having a model for the cost is to be able to verify its sensitivity to the variations of the design parameters. Let us start by checking what is the optimal architecture given these values.

Let us consider the case with \( N = 1024 \). By deriving with respect to \( n, \mu \) and \( \beta \) we have three equations which give the optimal values. Forgetting by the moment that \( n \) and \( m \) must be integer multiples of \( \mu (\beta^{-1} \text{can actually be not integer if we admit a mixed implementation of speed advantage, with both parallelism and physical speed, and we suppose that the cost function is not changed significantly by that}) \) it turns out \( \mu = 7, n = 22.6, m = 20.5 \) and \( \beta = 0.31 \). A feasible possibility which is close to this one is \( \mu = 8, n = m = 24 \) and \( \beta = 1/3 \). Note that the optimal speed advantage is large (it means running the fabric at 320 Mbit/s with 100 Mbit/s ports) probably rather more than could be expected. Fig. 4 gives optimal values of \( \beta \) and minimal costs for networks of different size (making \( n \) and \( m \) multiple of \( \mu \), with \( \mu \) power of 2). Figs. 5 and 6 show the sensitivity of the cost measure to the parameters \( \mu \) and \( \beta \). It is very interesting to note how all the parameters, and especially the speed advantage \( \beta \), change with the target size. Note that this is related to the more than linear growth of the cost of the components with \( N \), while the interconnection cost grows linearly with \( N \).

In a more realistic scenario the traffic characteristics are known and specified; our assumption is to have three different service classes, each with a different bandwidth (BW): 2, 8 and 32 Mbit/s; these connections are multiplexed over a link with a total bandwidth of 144 Mbit/s (we assume that the overall bandwidth is an integer multiple of the minimal connection rate). The design traffic load (\( \lambda \)) is assumed to be equal to 0.8; this value corresponds to the effective traffic that is carried into the network; \( \alpha \) specifies the share of traffic for each service class. The assumptions are summarized in Table 1.

The switching network to be designed will have a size in the range of a few hundred to a few thousand ports, and we will evaluate both strictly non blocking and blocking topologies.

Let us suppose that no speed advantage is implemented in the switching fabric, i.e., \( \beta = 1 \); moreover, from the traffic characteristics we have that \( b = \frac{1}{2} \) and \( B = \frac{1}{10} \). Applying the result (1) on generic Clos networks with \( \beta = 1 \), we have:

\[
m > 2 \max_{b \leq \omega \leq B} \left\{ \lim_{\epsilon \to 0} \frac{n - \omega}{\max(b, 1 - \omega + \epsilon)} \right\}
\]

where

\[
\lim_{\epsilon \to 0} \frac{n - \omega}{\max(b, 1 - \omega + \epsilon)} = \begin{cases} n - \omega & \text{if } 1 - \omega \geq b \\ \frac{n - \omega}{1 - \omega - 1} \end{cases}
\]

(assuming \( [a] \) to be the "smallest integer greater than or equal to \( a \)).

From direct inspection, we see that in each of the three cases \( \max(b, 1 - \omega) = 1 - \omega \); the maximum of the function
Fig. 6. Cost per port versus speed advantage $\beta$ for a nonblocking three-stage Clos network of size $N$ in an unrestricted packet scenario.

### Table I

**Characteristics of Traffic Classes**

<table>
<thead>
<tr>
<th>Class</th>
<th>BW</th>
<th>Multiplicity</th>
<th>Weight</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class_1</td>
<td>2 Mbit/s</td>
<td>1</td>
<td>$\frac{A}{2}$</td>
<td>0.5</td>
</tr>
<tr>
<td>Class_2</td>
<td>8 Mbit/s</td>
<td>4</td>
<td>$\frac{A}{4}$</td>
<td>0.3</td>
</tr>
<tr>
<td>Class_3</td>
<td>32 Mbit/s</td>
<td>16</td>
<td>$\frac{A}{16}$</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Table II

**Values of $m$ for a Nonblocking ATM Clos Network**

<table>
<thead>
<tr>
<th>$m_{\text{min}}$</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>24</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{\text{min}}$</td>
<td>5</td>
<td>9</td>
<td>19</td>
<td>41</td>
<td>61</td>
<td>81</td>
</tr>
</tbody>
</table>

is reached for $\omega = B$, in which case we have:

$$m > 2 \left[ \frac{n - B}{1 - B} - 1 \right]$$

Computing this formula for "interesting" values of $n$, we derive the nonblocking minimum values for $m$ ($m_{\text{min}}$), shown in Table II.

So far we have made no hypotheses on the size of the switch, nor on technological issues; we want now to study how the cost is influenced by the basic switching element size (represented by parameter $\mu$ in the cost formula) and by the switch size $N$ for a nonblocking architecture.

The general expressions for the cost of the various parts of the fabric are still (4.4,4), with $\beta = 1$.

The total cost of the switch is hence:

$$C_{\text{SW}} = \frac{C_{\text{VLSI}}}{\mu^2} N m_1 (2 + \frac{N}{n^2}) + 2k_i' N (1 + \frac{m}{n}).$$
TABLE IV

<table>
<thead>
<tr>
<th>N</th>
<th>m</th>
<th>m_{th}</th>
<th>m</th>
<th>P_{N/32 Mbit/s}</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7.60 × 10^{-6}</td>
</tr>
<tr>
<td>128</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>4.80 × 10^{-6}</td>
</tr>
<tr>
<td>128</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>2.96 × 10^{-11}</td>
</tr>
<tr>
<td>1024</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>3.14 × 10^{-3}</td>
</tr>
<tr>
<td>1024</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>2.25 × 10^{-6}</td>
</tr>
<tr>
<td>1024</td>
<td>24</td>
<td>26</td>
<td>26</td>
<td>1.91 × 10^{-10}</td>
</tr>
</tbody>
</table>

of the 32 Mbit/s connections, that is the larger and more disadvantaged service class. Results are shown in Table IV.

It is clear that for \( m = n + 2 \) both architectures can be considered virtually nonblocking for the specified traffic. Since both \( n \) and \( m \) must be multiple of the switch basic size \( \mu \), the relationships between \( m \) and \( n \) will be:

\[
m = n + \mu i \text{ and } m = i \mu (i \in \mathbb{N}).
\]

Then we can follow the same methodology as for nonblocking networks: (7) will be used instead of (1) to generate the possible set of values for \( n \) and \( m \); for each value of \( \mu \) a new optimal value for \( n \) will be identified, with the corresponding cost.

Fig. 6 shows the minimal costs as a function of \( \mu \) for different switch dimensions \( N \); blocking probabilities are always close to \( 1 \times 10^{-10} \) for \( \mu = 2 \) or below that value for \( \mu > 2 \). It turns out that the best choice for networks of small size is that with \( \mu = 4 \), while \( \mu = 8 \) is better for networks with more than 1024 ports.

As far as the parameter \( n \) (and the respective \( m \)) is concerned, costs are not very sensitive to its variations: Fig. 7 shows the variation of the cost with \( n \) for different sizes of a blocking switch and the same optimal \( \mu \) (\( \mu = 4 \)), with \( m = n + \mu \) as in (7). It can be noted that, in the neighborhood of the optimal value, curves are very flat, and therefore a certain degree of flexibility in design choices is allowed; this is also true in nonblocking networks even if the optimal values of \( n \) are not always coincident for both architectures.

Fig. 7 shows how the minimal costs change with the network size \( N \). The graph is useful to compare the three cases we have considered. For the unrestricted packet case the best values provide configurations with speedup \( \beta < 1 \); on the contrary, for the cases with the traffic specifications of Table I it can be shown by analysis that a speedup is not convenient. Although it may be misleading to draw general conclusions from this example, at least it shows that more attention should be paid to the interplay of the various parameters and to the convenience of high-speed links within the fabric. It is felt that this effect carries over to other cases, including other topologies.

V. CONCLUSIONS AND FURTHER RESEARCH

In the paper a methodology has been outlined which permits to perform an evaluation and optimization of the cost of an ATM switching architecture under some performance constraint given in terms of virtual connection blocking probability. There are several observations which can be made about the work done and its possible developments.

First of all, it seems that the cost of the fabric should be a negligible fraction of the cost of the whole switch, dominated by the line interfaces for what concerns the hardware, and in general by software. It is however true that the fabric is the area where the most advanced technology is employed, and the critical point in determining the Quality of Service delivered. The lack of a precise knowledge of the role of the various
parameters in determining the cost and the performance of the switch can still lead to an excess over sizing of the hardware or to bad performance in some traffic condition.

An example will make this point clear. If we consider the Clos switching network examined in the last section, with \( N = 128, n = 16 \) and \( m = 24 \), it seems that we have a system with exceptionally good performance, and this is certainly true if the assumed traffic conditions are maintained. But if we want to route through the fabric also some 100 Mbit/s connections, we find the perhaps unexpected result that they are almost always blocked: the blocking probability is close to 1! It is clear that in a multi-rate environment the assumptions about the traffic are extremely critical, and trusting any a priori traffic model can be very dangerous, in consideration of the little knowledge existing on real ATM traffic.

Comparing the cost of this solution with the design of a fabric being strictly non-blocking under any traffic condition (unrestricted packet case) we find that having a non-blocking switch with speed advantage costs about three times as much: this can be a reasonable price to pay, but at this point the cost of the fabric is probably not negligible any more.

A second point regards the "effective bandwidth" abstraction which has allowed all these considerations to be made. This is certainly a strong assumption to make, and it is likely to make peak bandwidth allocation necessary within the fabric for a large class of connections. It is however necessary to follow this procedure if we want to build and engineer safety buffered fabrics with internal virtual connection routing. Note however that this also demonstrates that this kind of architectures, in spite of their implementation advantages, are far from being free of problems. The quota of bandwidth resources that a connection is employing on a link depends on the multiplexing environment of that link, i.e., on the characteristics of the connections active on that link: trivially, a bursty connection with average bitrate 0.5 Mbit/s and peak bitrate 1 Mbit/s will occupy 1/100 of a 100 Mbit/s link which it shares with a 99 Mbit/s continuous bitstream connection, but roughly 1/200 of the same link shared with other connections of the same kind.

This means that the "effective bandwidth" \( \omega \) just introduced is an approximation which can be rather crude. The reasons for maintaining this assumption and developing a theory based on it are the following:

- it is valid in a good number of realistic multiplexing environments [7];
- it allows the comparative analysis of competing fabric topologies and important qualitative considerations can be derived from this analysis;
- the resulting formulas are fast and easy to use, and can be complemented by a simulative approach when specific solutions have to be assessed.

A final consideration about the cost model: the assumptions made were very approximate: this is an area where a lot of work can be done to find the correct tradeoff between model simplicity and significance.

REFERENCES


Paolo Coppo received the degree in electrical engineering from Politecnico di Torino, Torino, Italy in 1988. He joined CSELT in 1988 and since then has been working on various aspects of the ATM technique, including performance evaluation of interconnection networks, ATM chip design, service definition, network architecture and management aspects. He spent 1991 as a resident visitor at Bellcore on a research agreement focusing on broadband circuit-switched support on heterogeneous broadband network protocol involved in the standardization scenario, actively participating to ETSI (European Telecommunications Standards Institute), EURESCOM (European Institute for Research and Strategic Studies in Telecommunications) and the ATM Forum.

Mr. Coppo is a member of the IEEE Communications Society.

Matteo D’Ambrosio received the degree in computer science from Universita’ di Torino, Torino, Italy in 1989. He joined CSELT in the same year, and since then he has been working on the performance evaluation of ATM switching systems; his current research interests include traffic characterization received his degree in Computer Science from Universita’ di Torino, Torino, Italy in 1989. He joined CSELT in the same year, and since then has been working on the performance evaluation of software platforms for telecommunication systems.

Riccardo Melen (M’92) received his degree in electrical engineering from Politecnico di Torino, Torino, Italy in 1981. In 1983 he joined CSELT, the research laboratory of the Italian operating company, where he was formerly involved in the design and performance evaluation of interconnection networks for multiprocessors. Since 1985 he has been working on telecommunication systems, especially dealing with broadband packet switching systems. During his activity in CSELT, he was in charge of the research area dealing with high speed data networks and B-ISDN. Since 1992, he has been an Associate Professor at Politecnico di Milano. His current research interests include innovative switching systems architectures and software for telecommunication systems.