An Integrated Model for Reliability and Synthesis of Multi-Rate and ATM Networks with SVC Connections

Brunilde Sansò

GERAD and École Polytechnique de Montréal
Department of Electrical and Computer Engineering
2900 Édouard-Montpetit P.O. Box 6079, Station Centre-ville
Montréal, Québec, H3C 3A7, Canada
bruni@crt.umontreal.ca

André Girard

GERAD and INRS-Télécommunications
Place Bonaventure, 900 de la Gauchetière Ouest, Niveau C
C.P. 644, Montréal, Québec, H5A 1C6, Canada
andre@inrs-telecom.uquebec.ca

Florent Mobiot

GERAD and École Polytechnique de Montréal
Department of Electrical and Computer Engineering
2900 Édouard-Montpetit P.O. Box 6079, Station Centre-ville
Montréal, Québec, H3C 3A7, Canada
florent@crt.umontreal.ca

August, 1999
Revised: June, 2000

Les Cahiers du GERAD

G–99–35

Abstract

There is an increasing interest for efficient synthesis methods (routing and dimensioning) of large and robust multi-service networks. In the case of ATM networks, the synthesis requires performance guarantees at the cell level (QoS) and at the call level (GoS). Another important feature is reliability, defined as the network robustness in the event of failure. In fact, the planners seek cost-effective networks with guaranteed GoS even in the event of important failures. In this paper we present an optimization model based on cost minimization of multi-rate and ATM networks with SVC connections. The model integrates the notions of QoS, GoS, failures and failure propagation between the physical and the logical level as well as circuit routing.

Keywords: Network reliability, optimization, failure propagation, multi-service, call routing, circuit routing, ATM, network synthesis, SVC, QoS, GoS.

Résumé

L'intérêt est croissant pour des méthodes efficaces de synthèse (acheminement et dimensionnement) de grands réseaux multi-services robustes. En ce qui concerne les réseaux ATM, cette synthèse exige des garanties de performance ou qualité de service tant au niveau cellule qu'au niveau appel. Une autre caractéristique importante dont il faut tenir compte a trait à la fiabilité définie comme la robustesse du réseau en cas de panne. Les planificateurs recherchent en effet des réseaux rentables fournissant des garanties de QS même en présence de pannes importantes. Dans cet article, nous présentons un modèle d'optimisation basé sur la minimisation des coûts pour les réseaux multi-débits et ATM avec des connexions à commutation de circuits virtuels (SVC). Ce modèle intègre les notions de qualité de service aux niveaux cellule et appel, de pannes et de propagation de panne de la couche physique à la couche logique ainsi que de routage de circuits.

Mots-clés : Fiabilité de réseaux, optimisation, propagation de pannes, multi-service, acheminement, routage de circuits, ATM, synthèse de réseaux, SVC, QS.

1 Introduction

Reliability has always been a major concern for the telecommunication industry, as witnessed by the nearly flawless operation of telephone networks. This issue of reliability is becoming even more important in the context of ISDN networks where the increased demand for broadband services and the introduction of fiber-based transmission systems with enormous capacity, a few physical links can now contain a very large part of the demand thus causing a major damage when they happen to fail.

1.1 Traditional Design

Traditionally, reliability was taken into account during the design of the transmission networks since the most frequent failures are transmission cable cuts [1]. Various techniques have been designed over the years to reroute transmission channels, e.g., a T1 trunk group and all of the 24 calls that it carried, around a failed transmission link and to provision sufficient reserve capacity to be able to make these reroutings when they are needed. In many cases, networks were designed to be able to reroute *all* the channels affected by a single link failure (for 100% rerouting capability) so that a customer would be unaware of a failure except perhaps for the very short time required for restoring the transmission path.

This design process was carried out in two broad stages. First, the switched traffic network was designed. The traffic demands, expressed as the number of call requests or packets per unit time, were used as input to an optimization procedure that yields as output the capacity of the links connecting the switches (*trunk groups*), expressed typically as the number of voice circuits or as the service rate of a packet queue, and the routing of this traffic in the network. This was done taking into account an appropriate measure of Grade of Service and some simple cost model for a unit of bandwidth on the trunk groups.

Having determined the size of the trunk groups, it still remained to determine from which transmission systems these bandwidths should be taken. This was known as the *circuit-routing* problem where the input was the set of trunk group sizes obtained from the traffic network dimensioning and the output was a set of routes on which to carry these groups in a given transmission network. The dimensioning of the transmission link, expressed as the number of transmission systems of a given type to be installed between two cross-connect points, was also computed by this procedure. It is at this stage that the question of network reliability was considered with the addition of some reserve capacity and the definition of a rerouting policy in case of failure of some transmission equipment.

Note that except for the use of the results of the trunk group sizes obtained from the traffic network design, these two stages were carried out independently of each other. More specifically, the notion of reliability never occurred during the design of the switched network but appeared only during the circuit routing stage.

This approach was well suited to hierarchical telephone networks, where the routing of calls was somewhat limited by the hierarchy, or to low-speed data networks. The introduction of adaptive nonhierarchical routing methods in the early 1980's added another dimension to the problem. Because of the fast response times and adaptability of the new switches, it now became possible to reroute *individual calls* around failures[1]. This notion applies not only to telephone networks but to any traffic network with adaptive routing such as switched VC ATM networks or IP networks. An interesting question is then to what extent it is possible to use the adaptation in the traffic layer to reduce the amount of reserve transmission equipment required to maintain a particular measure of reliability. This has had two consequences. First, it has brought an important change in the notion of reliability itself. The second effect is that it has become necessary to take reliability into account during the design of the traffic network because of the so-called failure propagation effect which in turn leads to the notion of integrated design.

1.2 Definition of Reliability

As mentioned above, another consequence of the evolution of telecommunication technology has been the need to redefine the notion of reliability itself. Some definitions of reliability, mostly in the Operations Research community, were traditionally associated to the notion of *connectivity* [2]. Here, reliability would be defined in graph-theoretic terms, for instance as the fraction of terminals that would remain connected given some failure pattern. As we mentioned above, another definition was commonly used in the telecommunications industry by the transmission engineers, where the number of *channels* between any two end points would not fall below a certain percentage of its nominal value, often 100%, in the case of a single link failure. Associated with this was a particular rerouting technique used in real-time to reconnect broken channels and the provisioning of a backup network.

It is now clear that such notions are not well suited to the nature of modern telecommunication networks for two reasons. First, they are not directly connected to the Grade of Service perceived by the users who are only interested in the possibility of getting a connection or of having their packets delivered to their destination and do not care at all about how the channels that carry these connections are actually implemented. In other words, graph-theoretic definitions do not relate well to customer perception. The same is true for the addition of reserve capacity in transmission networks except perhaps for the case of 100% rerouting capability. As we said before, the other weakness of these approaches is that they do not take advantage of the adaptability of routing in the traffic networks which could permit savings in the amount of backup equipment. This is particularly true for full rerouting capability which may not actually be needed if call rerouting is taken into account.

Therefore, another definition of reliability as *the expected lost traffic or expected delay due to failures* in the network has been proposed in [3, 4, 5, 6, 7]. This is more realistic since it is directly related to the grade of service perceived by the user (i.e., blocking and delay) and is also related to the effectiveness of the call routing algorithm. This is the definition of reliability that we will be using in this paper.

[1]Even though rerouting of calls in progress was possible, it was technically difficult and was never used so that rerouting only applied to calls that arrived after the failure

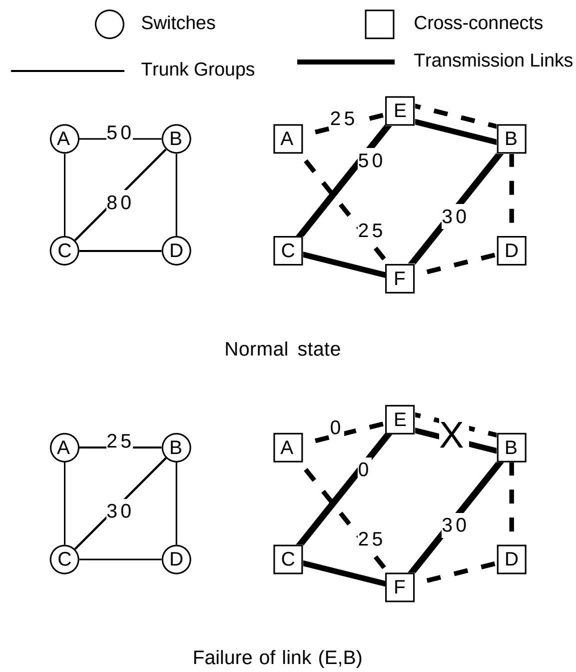


Figure 1: Failure propagation from the transmission network to the switched network

1.3 Failure Propagation

The notion of failure propagation can be best understood from Fig. 1 taken from [4] showing both the switched and transmission layers of a small network. In that example, 25 of the 50 circuits of trunk group between nodes A and B are routed on transmission path (A, E, B) and 25 on path (A, F, D, B). Similarly, 50 of the 80 circuits of trunk group (C, B) are routed on path (C, E, B) and the remaining 30 circuits on path (C, F, B). If the physical link (E, B) fails, transmission paths (A, E, B) and (C, E, B) are no longer available. Hence, the capacity of *both* trunk groups is reduced: Trunk group (A, B) now has only 25 circuits and (C, B) 30 circuits.

1.4 Integrated Design

The need for an integrated design method, where both the switched and the transmission layers are designed at the same time, can then be understood from this example. Coming back to the example of Figure 1, we could protect the network against the failure of transmission link (E, B) by installing 75 protection circuits on another path, say $(ECFB)$, at some extra cost. Note however that in the traffic network, calls from A to B that would normally be offered to the (A, B) link could be offered to the alternate paths (A, C, B) or (A, C, D, B) in the event that transmission link (E, B) failed. This in turn would reduce the need to install the reserve circuits on the transmission path $(ECFB)$. Note however that these reserve circuits may be used as a backup for *other* failures so that the precise tradeoff between call rerouting and dimensioning of backup facilities quickly becomes very complex. An important consequence is that it is now necessary to take into account the relationship between the switched and the transmission levels during the network synthesis. In other words, *integrated* network design is now required.

1.5 Our Problem

A number of models have been proposed [8, 9] integrating the design of the switched layer and the circuit routing. Some of them are based on an accurate model of the routing [9, 10] e.g., fixed alternate routing and often make up a large fraction of the computational requirements of the overall algorithm. Others [8] use a simplified model of the routing with limited accuracy.

For this reason, this paper concentrates on the design of the switched network but as a part of an integrated method for reliable design. We extend to the case of multi-rate and ATM networks a flow-based algorithm previously designed for single-rate circuit-switched networks. We also present numerical results on a realistic backbone network to show that the computation times are reasonable and that the method can indeed be used for integrated design.

From now on, any mention of network synthesis should be understood to apply to the switched network. The paper is divided as follows. Section 2 is devoted to the presentation of the synthesis model itself, effective bandwidth, QoS, GoS and blocking probabilities as well as the overall formulation of the problem. Section 3 contains the solution approach based on a problem decomposition and use of the flow deviation method. The numerical results are presented in Section 4. Conclusions and further work follow in Section 5.

2 The Synthesis Model

The switched network synthesis problem is defined as finding the capacities of the trunk groups and the optimal flows carried by the links. It can be divided into two subproblems:

1. the call routing subproblem (RP) which is to find the flows on each link to minimize the total lost traffic;
2. the dimensioning subproblem (DP) where we compute the capacities on each link to minimize the total cost of the network.

We assume that the switched network is represented by an undirected graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ where $\mathcal{V}$ is the set of switching nodes and $\mathcal{E}$ the set of logical links or *trunk groups*. If the switching technology is ATM, as is the case here, then the set of logical links can be seen as the set of VPs. In that context, the synthesis problem typically refers to the optimal dimensioning of VPs and the optimal routing of connections in the VP network. A similar problem has been much studied for circuit-switching with non-hierarchical networks with dynamic call routing algorithms and various traffic models [8, 11, 5, 12] and these models have been extended to multi-rate technologies and ATM networks. Most vendors now offer three connection *types*: permanent connections (PVP or PVC), soft connections (semi-permanents), or switched connections such as SVCs. In this paper, we deal with switched virtual circuits (SVCs) which are dynamically established on demand and torn down when transmission is complete.

Even though we concentrate on the synthesis of the switched network, we must nevertheless model the reliability explicitly in that layer. In order to do this, we assume the physical or *transmission* network underlying the logical network is denoted by an undirected graph $G(\mathcal{N}, \mathcal{M})$ where $\mathcal{N}$ is the set of transmission nodes (e.g., cross-connect devices) and $\mathcal{M}$ the set of transmission links. We do not try to optimize the circuit routing but assume that the routing of circuits in the transmission network is given.

We define the *state* of the transmission network by the state (failure or operation) of its components, nodes or links. Although this state space is enormous, in practice, a single cable cut is by far the most likely failure and we consider that a node failure is such a rare event that we do not need to take it into account. With these assumptions, it has been shown in [7] that most of the state space is covered by the no-failure and the single failure states. Thus, only $(|\mathcal{M}| + 1)$ states are necessary to evaluate an expected measure of performance such as the expected lost call traffic proposed in [4, 5, 6, 7] or the expected delay such as in [3].

Because of failure propagation, the effect of transmission link failures on the performance of the switched network is thus faithfully taken into account. For every state, the trunk group sizes can be easily calculated from the circuit routing and are therefore integrated into the synthesis procedure.

2.1 Model Formulation

The approach followed here is to compute the link dimensioning by considering the set of single link failures. For every failure state we must solve a routing subproblem. Then, the result of such subproblems are incorporated into a suitable dimensioning model that takes into account the trade-off between adding capacity at the switched level and degraded GoS in the event of transmission failures. This type of framework has been used in the context of packet and circuit-switched networks [3, 12, 4] but using it for the synthesis of ATM networks requires the development of important modeling details that make it particularly challenging. In what follows, we proceed step by step into the presentation of such details. First we present the notation, then show the overall synthesis problem including the routing subproblem. The details concerning the call blocking probabilities used for SVC connections and the effective bandwidth are provided in the appendices.

2.1.1 Notation In this paper, paths are represented by a triplet of the form (p, c, ℓ) where p denotes the number of an origin-destination pair, c the class of service that uses the path and ℓ is a sequence number for the path in some given list of all paths available to route calls of class c between o-d pair p. With this convention, we can define the variables of the model as follows:

$\mathcal{C}$ set of service classes. Each class c is defined by its rate w_c^u, the number of circuits required by a connection on link u. In the case of multi-rate circuit-switched networks, this number is a constant multiple of a fixed basic bandwidth unit (BBU). For ATM networks, we denote $\mathbf{W}_u$ the vector of service rates obtained from the computation of effective bandwidth on link u;

$\Omega(S_k)$ set of traffic flows so that flow conservation and non-negativity constraints are respected for the state S_k (normal and failure states); a state is defined by a binary vector which describes the condition of all the physical links;

N_u capacity of the link u as determined by the model. It is expressed as a number of circuits, a circuit being the base rate or the basic bandwidth unit (BBU). In this paper, we assume that a circuit corresponds to a 64 Kbps peak rate for a voice call;

$f_\ell^{p,c}(S_k)$ traffic flow of class c in erlangs offered to the path ℓ of the OD pair p for the state S_k. This is the average call attempts on the link. We also use $\mathbf{f}_u$ to refer to the vector of traffic of all classes on link u;

K_u cost of adding one unit of capacity to link u;

$\alpha_k^{p,c}$ cost of losing one unit of flow of class c from OD pair p for the k^{th} state of the network;

D_u capacity already in place on link u;

$d^{p,c}$ average connections demand for the OD pair p and class c;

$Pr(S_k)$ probability of being in state S_k;

$B_u^c = g(\mathbf{f}_u, \mathbf{W}_u, N_u)$ blocking probability function on link u for class c;

$L_k^{p,c}$ GoS of the OD pair p for class c in state k. We refer to $\overline{L}_k^{p,c}$ as the maximum GoS allowed;

$I_{u,(p,c,\ell)}$ arc-path incidence matrix. It takes the value 1 when arc u belongs to the path ℓ for class c and OD pair p and 0 otherwise.

2.1.2 The Synthesis Model The synthesis problem is to dimension the trunk groups at minimum cost in such a way that all GoS constraints are met in all states. Because the actual performance is tied to the routing, the traffic flows also have to be optimally routed in each state. A simple form of the synthesis problem can then be stated as

$$\begin{aligned} \min_{f,N} \; Z_D &= \sum_u K_u N_u \\ L_k^{p,c} &\leq \overline{L}_k^{p,c} \qquad \forall\, k = 1, \ldots \\ f_\ell^{p,c}(S_k) &\in \Omega(S_k) \\ D_u &\leq N_u \\ N_u & \quad \text{integer.} \end{aligned} \tag{1}$$

In this form, the problem is particularly difficult to solve and a number of approximations and simplifications have to be made.

First, we preferred not to consider the GoS constraints (1) explicitly in our model. We chose to associate penalty costs $\alpha_k^{p,c}$ to the lost traffic. These costs can be perceived as actual costs or some Lagrangian multipliers of the dualized constraints (1).

2.1.3 Flow Model For most real-time routing methods, the $L_k^{p,c}$ are complex functions of the link blocking probabilities B_u^c and are given implicitly as the solution of a system of nonlinear equations, the so-called Erlang Fixed-Point system. This makes the optimization much more complicated and instead of using Equations (1), we assume that traffic flows are conserved inside the network and that the total lost traffic is given by the total traffic blocked on the links. Under these assumptions, the f variables become standard multi-commodity flows. The approximation is certainly reasonable for small values of blocking. In the case of failures, however, we cannot assume that the blocking will be small and the accuracy of this approximation has to be evaluated. It was shown in [4] that this is indeed the case for single-service networks. One of the results of this paper will be to show that this conclusion also holds for multi-rate and ATM networks. Another advantages to multi-commodity flows is that they are very well known and efficient techniques are available to compute numerical solutions.

The modeling of the switched network depends very much on the type of control (centralized versus decentralized), the type of switching (per call, per service, etc.) and the measure of performance or QoS (Quality of Service) that must be achieved. A new typology to compare the different methods and an exhaustive literature review is provided in [13]. A key element in all cases is the real-time routing algorithm that is used. Most optimization models in the literature were produced with a specific routing algorithm in mind before specific features concerning connection establishment were actually adopted and implemented by ATM switch vendors. Modeling the routing as a multi-commodity flow has the added advantage that it is independently of any specific real-time implementation. The basic assumption is that in a well-dimensioned network, the average behavior of a particular real-time algorithm will not be significantly different from the traffic patterns produced by the flow model. This assumption has turned out to be sufficiently accurate for the design of circuit-switched networks [4] and one objective of this paper is to verify whether this is still true for ATM networks.

Given these assumptions, the arc-path formulation of the synthesis model then becomes

$$\begin{aligned}
\min_{f,N}\ Z_D &= \sum_u K_u N_u + \sum_{k=0}^{|\mathcal{M}|} \Pr(S_k) \sum_u \sum_c w_u^c(S_k) \qquad (2)\\
&\quad \left\{ \sum_p \alpha_k^{p,c} \left(\sum_\ell f_\ell^{p,c}(S_k) I_{u,(p,c,\ell)}(S_k) \right) \right\} B_u^c(S_k)\\
f_\ell^{p,c}(S_k) &\in \Omega(S_k) \qquad (3)\\
D_u &\leq N_u \qquad (4)\\
N_u &\quad \text{integer.} \qquad (5)
\end{aligned}$$

In this synthesis model, the objective function is composed of two terms: the capacity cost term plus a term that reflects the expected cost of losing traffic in the event of failures. The flow values that appear in the objective function are computed from the call routing subproblem solved for every state of the network (3). Constraints (4) say that capacities installed must not be lower than those already in place before the optimization. Both sets of integer constraints (4 and 5) can be relaxed for broadband networks where huge capacities are involved. Although the structure of the model is similar to the one proposed in [4], specific modeling details are not. They are of prime importance to grasp the key features of multi-service and ATM networks. Such details relate to the call blocking probabilities function and the effective bandwidth, and are given in the following subsections. Moreover, such details greatly increase the resolution complexity of the problem.

2.1.4 Call Blocking Probabilities For multi-rate circuit-switched and ATM networks with effective bandwidth, the GoS is the end-to-end blocking probabilities which can be computed from the link blocking probability. As seen in the previous subsection, this calculation is a key element in multi-rate and ATM networks synthesis. Therefore, it is crucial to use a formula for link blocking probabilities that is accurate and fast enough to compute. We refer the reader to [4] for more details concerning the computation in the MF model of end-to-end loss probability given the link blocking probabilities.

For a single service, the blocking probability is given by the well known Erlang B function $E(X, N)$ for N circuits and a Poisson distributed offered traffic (X). In the case of multi-service, an exact model for the computation of multi-rate link blocking probability and a recursive algorithm to solve it have been proposed in [14] and [15]. This algorithm is efficient but still inappropriate for large network synthesis. The computation time grows rapidly with the number of servers and more importantly, the model is defined only for an integer number of servers. We would like to avoid using an integer programming technique and for this reason, in this paper, we propose to work with the approximation given in [16], which depends much more weakly on the number of servers and also can be extended to fractional N. The calculation is given in Appendix A.

2.1.5 Effective Bandwidth For ATM networks with VBR traffic, the standard approach has been to model the cell QoS constraints (cell loss probability, cell delay, cell delay variation, etc) via the notion of effective bandwidth [17]. An important difference with multi-rate methods is that this bandwidth is a function of the server rate, that is, the trunk group sizes (See Appendix B for details on the model used here). In a dimensioning algorithm such as the one we are considering here, where these group sizes are design variables, this will introduce new elements in the calculation of the optimality conditions and the calculation of directions of descent. An important question is then to what extent this added complexity will slow down the numerical calculation and how it will affect convergence.

3 Solution Technique

The resolution method (See Fig. 2) is based on the *capacity and flow assignment* algorithm [18, 19, 20, 21]. First, we fix the capacity variables N_u and solve a set of routing problems (6-8), one for each state. Then, we fix the flow variables $\mathbf{f}_u$ and solve the dimensioning problem (2) for the variables N with fixed values of f. Between the resolution of the two subproblems, the multipliers α must be updated in order to meet the required GoS values. We now proceed to get into the details of the call routing and dimensioning subproblems as well as the penalty calculation.

3.1 Call Routing Subproblem (RP)

As previously mentioned, the call routing subproblem domain is defined by $\Omega(S_k)$ for the normal and the failure states. As we said, the formulation of the routing subproblem would

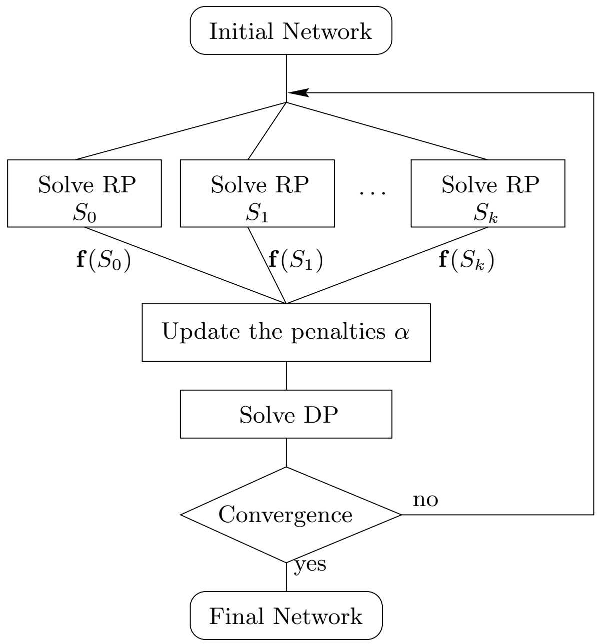


Figure 2: Solution algorithm

change according to the type of routing that is actually implemented. To keep the approach as general as possible, we use as a call routing an adaptation of the multi-commodity flow model (MF) that was used for packet- and circuit-switched networks [4, 6, 7].

Assuming flow conservation in the network, we represent this RP by the following non-linear multi-commodity flow model. It will give the optimal flows $\mathbf{f}_u$ carried on each arc u for all the states of the network.

$$\min_f \ Z = \sum_u \sum_c w_u^c \left[\sum_p \left(\sum_\ell f_\ell^{p,c} I_{u,(p,c,\ell)} \right) \right] B_u^c \tag{6}$$

$$\sum_\ell f_\ell^{p,c} = d^{p,c} \tag{7}$$

$$f_l^{p,c} \geq 0. \tag{8}$$

Equation (6) states that the expected lost traffic in the network is to be minimized. This is an approximation to the real value since it is the sum of the traffic blocked on each link. The accuracy of this approximation has been shown in [4] to be sufficient for the purpose of network design. Constraints (7) are the standard flow conservation equations.

The coupling between the two levels is taken into account by the presence of the state-dependent flows $f_\ell^{p,c}$. Given a particular routing of the circuits on the physical network, we can compute, for each state, the optimal flows in the switched network. The only other values needed to compute the objective function are the B_u^c which we know how to calculate using the blocking model. The call routing subproblem (6-8) can be solved with a linearization algorithm such as the Frank-Wolfe method also known as flow deviation (FD) method [18]. This method widely used in data networks is modified for multi-rate circuit-switching. We can notice that the solution is only a local optimum since the objective of the problem is not convex.

The Frank-Wolfe method is an iterative descent algorithm where we first calculate a direction of descent and then the step size for a move in that direction. In the present case, the direction calculation reduces to the following shortest path problem

$$\min_{Y} \sum_{u} \sum_{c} \frac{\partial Z(f)}{\partial X_u^c} Y_u^c \tag{9}$$

subject to constraints (7–8) where $X_u^c = \sum_p \sum_\ell f_\ell^{p,c} I_{u,(p,c,\ell)}$ is the current flow of class c offered to arc u and Y_u^c is the multi-commodity flow on the link that defines the direction of the next move. The link metric used to find the shortest paths is

$$\begin{aligned} \frac{\partial Z(f)}{\partial X_u^c} &= \sum_{v} \sum_{j} \frac{\partial \left[w_v^j X_v^j B_v^j \right]}{\partial X_u^c} \\ &= w_u^c B_u^c + \sum_{j} w_u^j X_u^j \frac{\partial B_u^j}{\partial X_u^c} \ . \end{aligned} \tag{10}$$

The linear program that defines the best direction is then

$$\min_{Y} \sum_{u} \sum_{c} \left[w_u^c B_u^c + \sum_{j} w_u^j X_u^j \frac{\partial B_u^j}{\partial X_u^c} \right] Y_u^c \tag{11}$$

subject to constraints (7–8) .

The step size, also known as the flow deviation from the current flow $\mathbf{X}$, is computed by solving this one-dimensional minimization

$$\min_{0 \le v \le 1} \ Z(v\mathbf{Y} + (1-v)\mathbf{X}). \tag{12}$$

3.2 Dimensioning Problem (DP)

The dimensioning problem is simply given by the solution N_u of the Kuhn-Tucker first order conditions. Note that this set of equations is separable by link if the traffic flows are known.

$$\begin{aligned} -K_u \;=\; & \sum_{k=0}^{|\mathcal{M}|} \Pr(S_k) \sum_{c} \left(\sum_{p} \alpha_k^{p,c} \left(\sum_{\ell} f_\ell^{p,c}(S_k) I_{u,(p,c,\ell)}(S_k) \right) \right) \\ & \left\{ \left. \frac{\partial w_u^c(S_k)}{\partial N_u} \right|_{N_u = N_u(S_k)} B_u^c(S_k) \ + w_u^c(S_k) \left. \frac{\partial B_u^c(S_k)}{\partial N_u} \right|_{N_u = N_u(S_k)} \right\} . \end{aligned} \tag{13}$$

Figure 3: Network used for numerical results

For multi-rate circuit-switched networks, the bandwidth requirement is constant and therefore, the term $\partial w_u^c(S_k)/\partial N_u$ always equals to zero. This is not the case in ATM where the effective bandwidth is a function of the server rate. This is an added complexity that we will examine further in the presentation of numerical results.

3.3 Penalty Calculation

Because there are no GoS constraints in the formulation (2–5), the solution will not meet the constraints unless the penalties α are given the correct value. If we suppose that these penalties have the same function as Lagrangian multipliers, we can update them for each iteration i of the global algorithm using a sub-gradient-like method [22]. For an OD pair p, a state k and a class c, we have

$$\alpha_k^{p,c}(i+1) = \left[1 + (L_k^{p,c} - \overline{L}_k^{p,c})\theta\right] \alpha_k^{p,c}(i) \tag{14}$$

with θ being a constant value which indicates how much we penalize the gap between the call blocking probability and its bound.

4 Results

In this section, we provide results for the application of the MF model to the synthesis of both multi-rate and ATM networks. The study of multi-rate networks will allow us to compare the computation time, accuracy of the solutions and the effect of circuit routing against the results obtained from another integrated design method. The study of ATM will give some insight on the effect of effective bandwidth on the MF algorithm and on the difference it produces on the optimal network structure. Finally, within the ATM study, we will assess how our reliability-integrated method compares with VN methods currently used by practitioners.

The results presented here were obtained for the Sprint network shown in Fig. 3. This network, also used in [4, 8, 11], is composed of 10 switching nodes, 18 physical nodes and 27 transmission links. The unit trunk cost K_u, evaluated from the data provided in [8], is

Table 1: Class of service parameters for ATM networks

Class	1	2	3
PCR	1	6	24
MCR	0.5	3	12
CLP	0.001	0.001	0.001

the sum of a termination cost of 300$ per switch and an airline distance cost of 0.05$ per mile. The traffic demand matrix is also taken from [8] for the first class of service. Since we consider three classes, we generate the other demands randomly. In all cases, we assume that the GoS requirements are 1% in the normal state and 20% in any of the failure states. In the case of ATM, the peak cell rate (PCR), the mean cell rate (MCR) and the cell loss probability (CLP) are given in Table 1 for each class considered. The PCR and the MCR are provided as multiples of the BBU. These values were chosen so that there would be a significant difference between the effective bandwidth and the peak rate of connections making the multi-rate and ATM networks different as well.

4.1 Multi-Rate Networks

The first set of results that we present is for multi-rate circuit-switched networks. In that case, we have numerical results obtained from an accurate modeling of an adaptive routing method [10] against which to compare the results obtained with the multi-commodity flow model. First we briefly describe the reference model, then show the comparison in terms of computation times and finally in terms of the accuracy of the MF technique.

4.1.1 Reference Method A model has been proposed in [9] for the solution of the complete integrated synthesis where both the switched network synthesis *and* the optimal circuit routing are computed together and a fast numerical algorithm has been described in [10]. In what follows, this is called the SG model and the solutions are also denoted in that way. The routing of connections in the switched network is an optimal load sharing between all the possible one- and two-link paths in the network.

We have used four sets of results to compare the CPU times, cost and GoS. The first one is the SG network obtained by solving the integrated design with the real-time adaptive routing method. This yields both the optimal switched network and the optimal circuit routing in the transmission layer. We can then use the MF algorithm with this optimal circuit routing to construct an optimal switched network: the result is called the MF network. We can also use the MF algorithm to construct an optimal switched network with different circuit routing. Here, we have used two rules: Equal sharing between the first and second link-disjoint shortest paths (MF 50-50) and 2/3 on the shortest and 1/3 on the second shortest path (MF 67-33).

4.1.2 Calculation Times We can see from Table 2 that the computation times for the MF algorithm are substantially smaller than for the SG method. This can be explained mostly by the fact that the SG algorithm needs to solve the Erlang system during the

Table 2: Costs and performance for multi-rate networks

Networks	MF Netw.	SG Netw.	MF 50-50	MF 67-33
Cost	16.66	16.98	17.52	22.31
CPU Time (secs)	148	654	120	138
Max Loss Class 3 (%)	0.07	0.003	0.008	0.0007

Netw. = Network

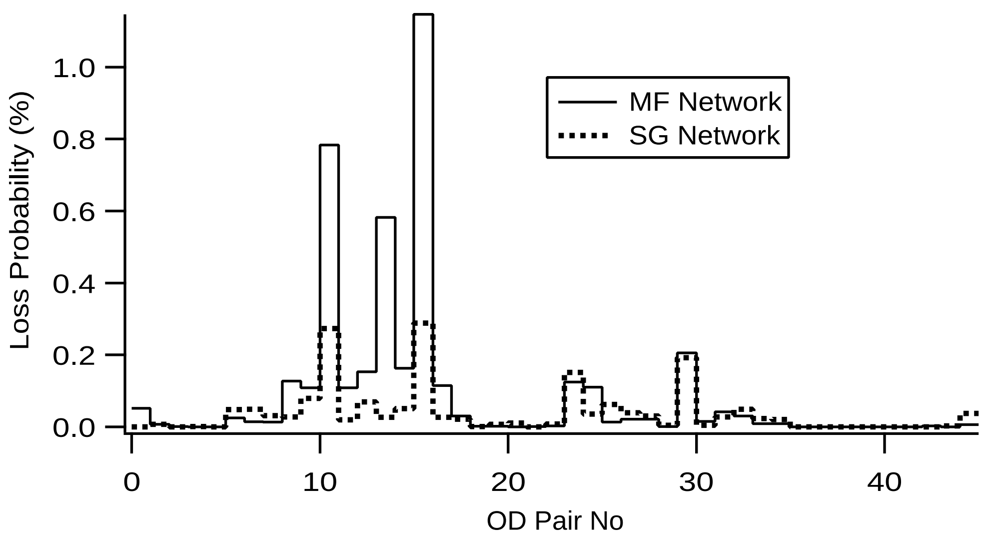


Figure 4: GoS for failure of transmission link 2-14

optimization of the switched network which is the largest component of the overall CPU times, a calculation not required by the MF method. In addition, it solves the optimal circuit routing subproblem but this is not a very large part of the CPU time. We can thus say that the MF method can yield significant reductions in CPU times for a network of this size and that we expect that the savings would be even larger as the network size grows.

4.1.3 Accuracy of the Solutions Table 2 also shows that the cost of the optimal solutions for the switched network obtained by the MF and the SG methods are not very different when the circuit routing is the same in both cases (first two columns of the table). Nevertheless, the value of the cost is not the only element by which to judge the quality of a solution. As we said, the MF model can only give an average allocation of calls on the paths between each o-d pair; it does not specify any real-time routing algorithm. In practice, networks do operate with some kind of real-time routing and there is a question of the GoS that is actually obtained in a network that has been designed by the MF method.

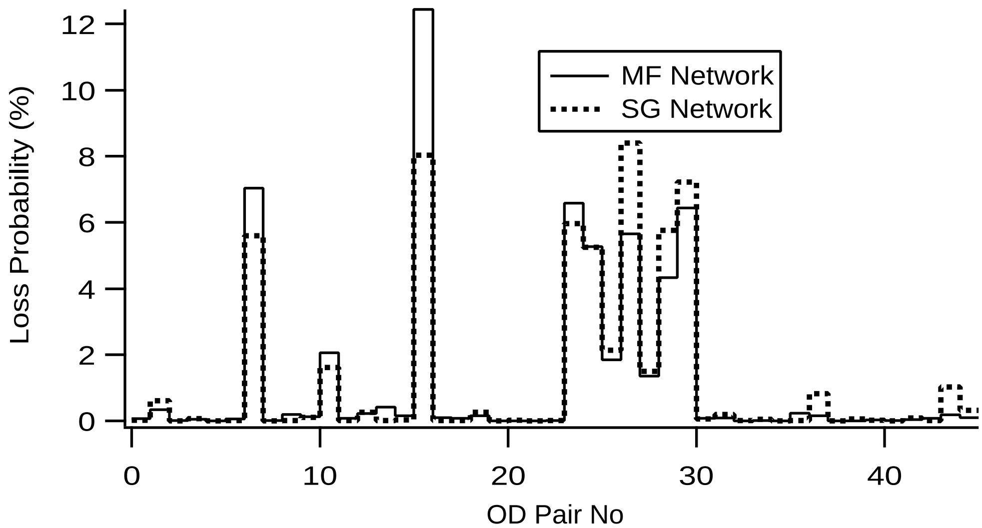


Figure 5: GoS for failure of transmission link 4-7

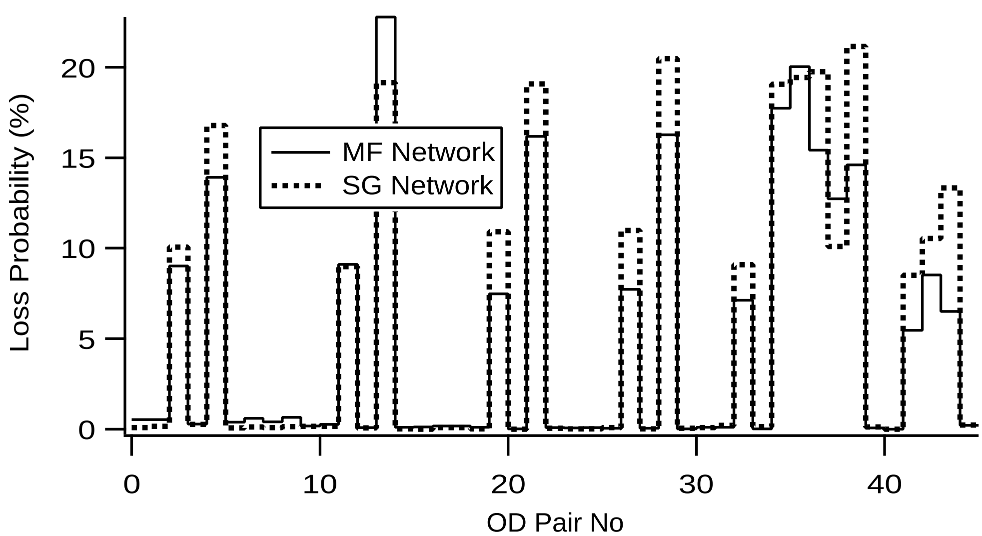


Figure 6: GoS for failure of transmission link 6-8

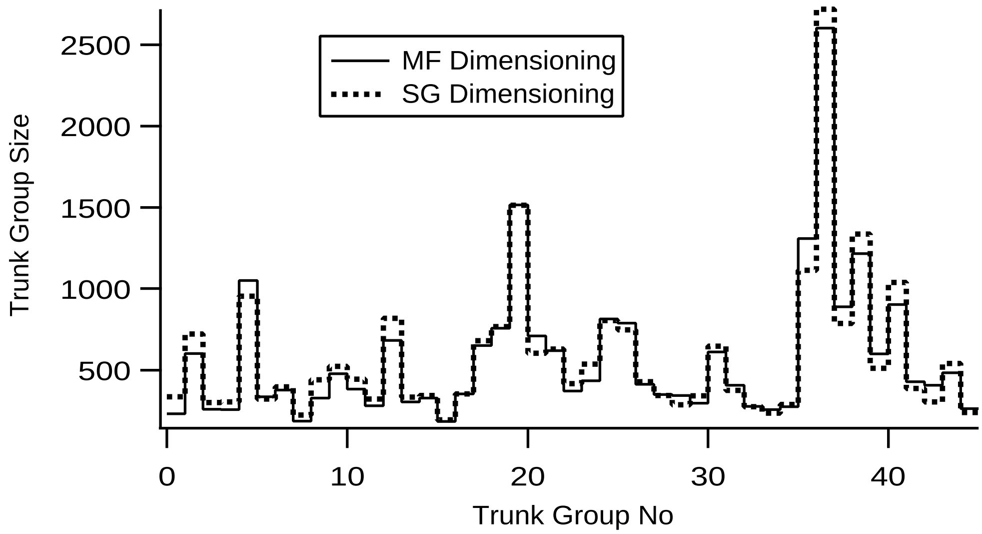


Figure 7: Dimensioning of trunk groups: SG circuit routing

Three typical examples can be found in Fig. 4, 5 and 6 that are representative of what we have obtained for the other states. The figures show the call loss probability of each o-d pair for class 3 traffic, the one experiencing the largest loss, for various failure states. These results were obtained by first dimensioning the network with the MF algorithm using the SG circuit routing. The GoS was then computed using the SG real-time call routing but on the MF network. In many cases, we see that the GoS that some o-d pairs actually experience when a real-time routing algorithm is used can be quite different from the values calculated by the MF model. We also note that the MF solution is neither optimistic nor conservative across all failure states. Nevertheless, the important point is that the actual GoS is never much above the maximum permitted value and that the overall network cost produced by the MF algorithm is not significantly higher than the cost produced by the more accurate SG method.

4.1.4 Dimensioning and Circuit Routing Because the MF model does not optimize circuit routing, we can use the routing produced by the SG method as one possible instance of routing, along with other simple rules, to check the effect of routing on the solutions. First, consider Fig. 7 showing the size in BBU of all the trunk groups. We see on that figure that there is little difference between the trunk group dimensioning produced by the SG and MF algorithms when using the optimal circuit routing obtained from the SG algorithm. A more substantial difference can be seen on Fig. 8 where we see the effect of using the 50-50 and 67-33 routings. The last case deviates substantially from the optimal trunk group dimensioning, which means that the choice of circuit routing is relatively important on the structure of the switched network. The effect on the network cost is also significant, as we saw on Table 2.

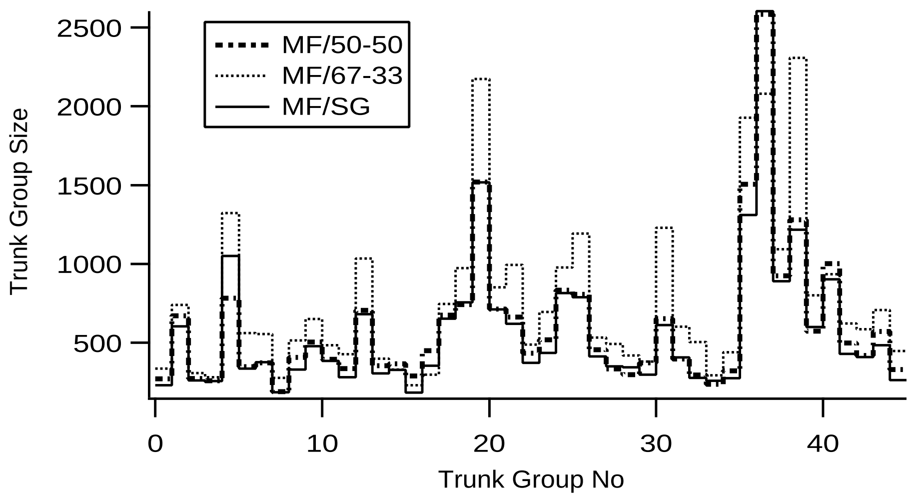


Figure 8: Effect of circuit routing on trunk group dimensioning: other circuit routing

4.1.5 Summary This case study indicates that the MF algorithm can substantially reduce the CPU time required for the design of the switched network as compared with a more accurate model of adaptive routing. The accuracy of the actual GoS experienced when using such a network with a real-time routing is sufficient for design purposes to the extent that it yields solutions that meet the GoS constraints in failure states and that the cost of the resulting network is not significantly different from the optimal cost.

This conclusion is nevertheless subject to one important condition. The circuit routing in the transmission layer should not be too different from the optimal routing. This shows once again the importance of integrated design and also the usefulness of the MF model that could easily replace the SG algorithm in the integrated design algorithm.

4.2 ATM Networks

We now examine the results obtained with the MF method for ATM networks with switched VCs and effective bandwidth. The main difference with the multi-rate case is the dependence of the effective bandwidth on the trunk group size which changes the optimal dimensioning equations. The first set of results has to do with the relationship between the GoS and dimensioning on the one hand and circuit routing on the other. The second looks at the use of service-specific Virtual Networks and how they can lead to a poor GoS in case of failure.

4.2.1 Effect of Effective Bandwidth on Dimensioning The first concern of the effect of effective bandwidth on dimensioning is relating to the convergence of the method. We have found that despite the complexity introduced in the calculation of directions of descent, the overall algorithm is robust and gives satisfactory results. We shall describe these results in details in the next subsection.

Table 3: Costs and performance of ATM networks

Circuit Routing		**SG**	**50-50**	**67-33**
Cost (M $)		12.86	13.37	16.37
GoS$_{max}$ for failure states (%)	1	0.8	0.8	1
	2	5	6	7
	3	29	36	35
Resulting GoS$_{max}$ for no failure state (%)	1	0.002	0.001	0.00005
	2	0.01	0.009	0.0004
	3	0.09	0.07	0.004
CPU Time (s)		183	162	151

Before going further in presenting the results, let us have a look at the costs of the multi-rate and ATM networks in Tables 2 and 3 respectively. Comparing those costs, we note that depending on the circuit routing used, the introduction of effective bandwidth in ATM networks produces a saving from 23% to 27% as compared with the multi-rate networks. This gain is foreseeable since the effective bandwidth for an incoming call on a given trunk group is generally smaller to the peak rate of that call (See Appendix B).

Nevertheless, this cost saving comes with an additionnal complexity regarding the optimal dimensioning equations (13). We can see that for our set of results the CPU times for ATM networks is increased by as much as 35% over what is needed for multi-rate networks. This added complexity is significant but still affordable considering that the maximum CPU time obtained for ATM networks is not much greater than 3 minutes.

4.2.2 Circuit Routing and ATM Design The effect of circuit routing on the design of the ATM switched layer can be seen from Table 3. The first line of the table is the cost of the switched network obtained with our method for the three different circuit routings considered. As expected, the cost seems to increase when we move farther away from the SG routing since the routing rules are suboptimal. Figure 9, that depicts the trunk group sizes, shows how this cost increase is obtained. In fact, while the SG and 50-50 routing produce comparable trunk group sizes, the more imbalanced routing (67-33) tends to add and concentrate more capacity on certain trunks, thus explaining the larger increase in costs.

Recall that in our method, there can be different GoS for different services both for the no-failure and failure states. In the second line of the table, we present, for each type of circuit routing, the maximum blocking experienced for each class by any OD pair in any failure state. It is clear that class 3, which is the class with the larger bandwidth requirement, also experiences the biggest blocking in the event of failure. This is due to the nature of the performance model used to approximate the blocking [16]. Note that the circuit routing does not seem to affect much the maximum GoS values obtained for the failure states. We also underline that, for class 3, even though the maximum values for a single OD are over 30%, most ODs will experience a maximum blocking well below

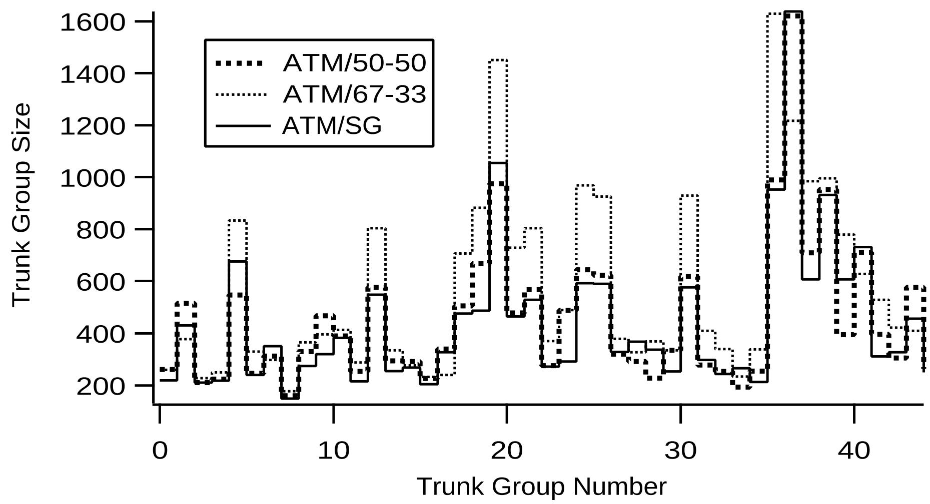


Figure 9: Trunk group dimensioning

20%. A typical case is depicted in Fig. 10 for which only a single OD goes beyond 30%. Note that in this figure, the extreme routing also produces more extreme solutions with OD pairs experiencing almost zero blocking whereas others go beyond 20%.

The next line in the table shows the resulting maximum values for the normal state. We get much less than 1% blocking in the normal state and it is clear that we have overdimensioned the network in order to maintain a reasonable GoS in the event of failures. That is the price to pay for the required robustness. However, the procedure allow us to do it in an intelligent way, thus avoiding the costs of gross redundancy. Note that as we move closer to a more imbalanced circuit routing, the GoS obtained in the no-failure state is greater even though the maximum GoS for the failure states is almost the same. This confirms two facts: that it is beneficial to spread the circuit routing and that blind overdimensioning does not necessarily improves the worst GoS situations.

The last line of the table corresponds to the CPU times (in seconds) needed to compute the synthesis. The whole procedure takes about 3 minutes on a SUN Ultra-1/167 with 64M of RAM. This time is acceptable for a synthesis problem in particular if we consider that each iteration requires the resolution of a routing problem for every failure state. To improve the solution time, this part of the procedure could be run in parallel. Another observation relates to the decrease of the CPU time as we move away from the SG routing. This is exactly the opposite trend seen with the network cost. This result confirms what is expected: as we use circuit routing policies closer to optimality, we have to deal with problems that are more difficult to solve.

4.2.3 Virtual Networks Virtual networks can be built to carry a single class of traffic; in that context, one speaks of service-specific VN. In these networks, there is no VC switching: a direct VP for each service is established between each o-d pair. The network

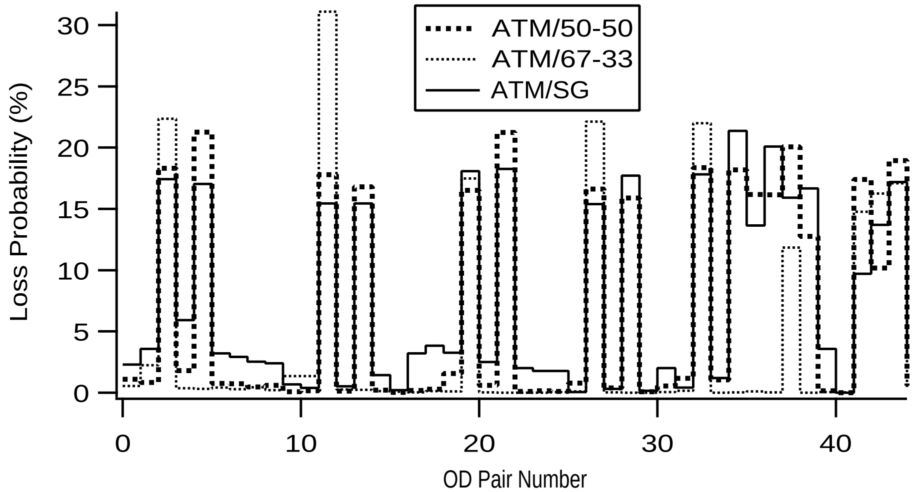


Figure 10: Switched VC: GoS for class 3 for state having link (1,8) failed

Table 4: Costs and GoS for ATM

Routing		SG		50-50		67-33	
Network		SVC	VN	SVC	VN	SVC	VN
Cost		12.86	14.44	13.37	14.60	16.37	16.30
Max Loss (%)	1	0.002		0.001		5×10^{-5}	
	2	0.01		0.009		4×10^{-4}	
	3	0.09		0.07		.004	
CPU secs		183	1.6	162	1.6	151	1.7

dimensioning then becomes much simpler since it is only necessary to dimension the individual VPs, which amounts to the inversion of the Erlang B function for each service and each o-d pair. Reliability is not generally taken into account in the design of the VN. However, to improve network robustness, the dimensioned VP is usually routed in the physical network using some rule (such as the 50-50 rule) so that the capacity is split into two disjoint paths. One of the questions that we want to answer in this section is whether such a scheme, widely used by practitioners, is indeed effective in the event of failures. Secondly, we want to see how the results compare with the SVC results presented in the previous section.

The first comparison we want to present is depicted in Table 4 which was obtained as follows. For a given VP routing, the switched network was dimensioned using the MF algorithm with VC switching (SVC Network). This yielded among other things the actual GoS obtained for the normal state as well as for each failure state. Using the GoS obtained for the normal state as bounds, the Virtual Network was then dimensioned so that the GoS

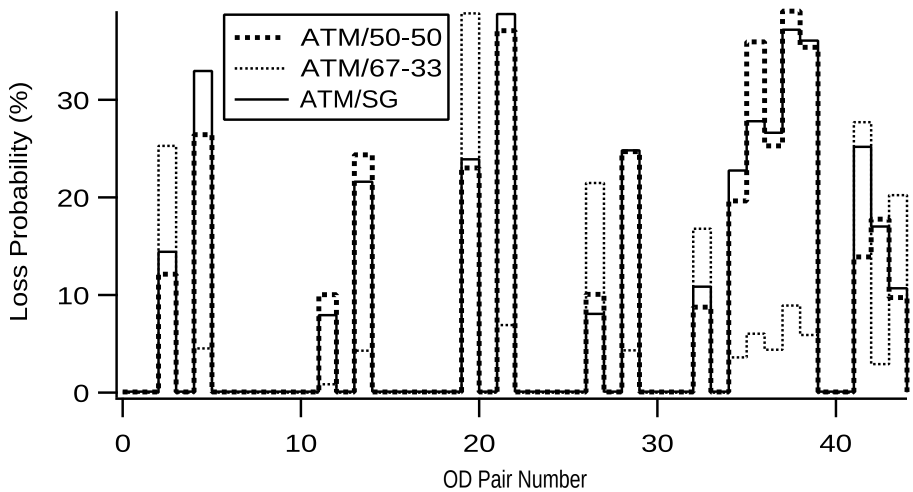


Figure 11: Virtual Network: GoS for class 3 for state having link (1,8) failed

would be no worse than the GoS obtained with the MF algorithm. Comparing the VN network cost with the SVC network cost would then indicate the cost increase (or decrease, as the case may be) caused by provisioning for failures while using service-specific virtual networks.

It is quite clear from the results of Table 4 that there is a substantial economic advantage in using SVCs. The VN network solution is 12% more expensive when the VP routing is the SG solution. Note however that this advantage decreases as the VP routing starts to differ from the SG solution. In fact, for the 67-33 VP routing, the switched VC network is slightly more expensive than the service-specific virtual networks.

The second advantage of the SVC method consists of creating a more robust network, that is a network that adapts better to the effect of failures. On Figure 11, the actual GoS for each o-d pair is plotted for the SG, the 50-50 and the 67-33 VP routings. When we compare the results of that figure with those previously obtained for the SVC case (shown by Fig. 10) we can see that in general, the grade of service is not as good. In fact, in the VN case, we can find many o-d pairs that present GoS close to 40% blocking in the event of failures whereas in the SVC case the 20% loss constraint was respected in almost all cases. Note that these poor results for the VN networks were obtained despite the fact that splitting was used and that the VPs were dimensioned so that the no-failure GoS was as tight as for the SVC method. This points out to the importance of integrating failures in the design.

5 Conclusions and Further Work

In this paper we have presented a model for the synthesis of multi-rate and ATM networks with SVC connections under reliability constraints. The general framework consists of enumerating the most important states of the network, solve a call routing subproblem taking into account failure propagation and design a dimensioning procedure that makes a trade-off between cost and expected losses. In order to adapt the method to the SVC case, we have used effective bandwidths and an approximation to evaluate the end-to-end blocking probabilities. The model not only integrates the notions of GoS and QoS but also accounts for cost-effective robustness by considering failure propagation linked to circuit routing and different GoS for the failed states.

The method provides a tool to put the capacity where it is required, thus avoiding gross overdimensioning. Moreover, since the method is not based on a particular call routing procedure, it is flexible and can apply to many call routing techniques. Because of its efficiency, it is a good candidate to replace the switched network synthesis algorithm in the SG integrated design algorithm, hopefully providing further reductions in computation times.

Both in the case of multi-rate and ATM networks, we analyzed the results for different circuit-routing strategies: optimized routing, 50-50 and 67-33. The results confirm the importance of traffic spreading and routing optimization. Nevertheless, the widely used 50-50 circuit routing rule seems to be a good alternative to a full optimization.

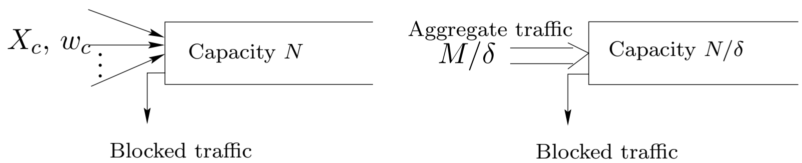
Concerning ATM networks, our method was compared with a popular scheme of VN dimensioning. We found that the VN scheme produces, as expected, a more expensive network. Such overdimensioning does not translate, however, into better GoS in the event of failures. This emphasizes the importance of integrating reliability in the network design.

A The Labourdette and Hart approximation

To use the approximation, we assume the following hypothesis for a multi-service link model (see Fig. 12(a)):

1. there are s classes of traffic with different bandwidth requirements w_c such as $w_c \ll N$ ($s = |\mathcal{C}|$). For ATM networks, w_c refers to the effective bandwidth of class c on the link;
2. blocked calls are lost;
3. call arrivals are independent Poisson processes (X_c is the offered traffic to the link). The connection times are exponential distributed (yet [14] considers any distribution);
4. the call admission policy is a complete sharing policy, that is, there is no trunk reservation. Each new call of class c competes with the others for the available capacity on the link.

In [16], the authors present an equivalent system (Fig. 12(b)) where the s classes of traffic are superposed to create a single traffic. This approximation strikes a compromises between complexity and accuracy. The complexity is dominated by the calculation of an Erlang-B function and the solution of a polynomial equation. It tends to be exact under an *asymptotic regime* when the link capacity and the traffic tend to infinity with a constant ratio $N \to \infty$, $X_c \to \infty$, with X_c/N constant. The blocking probability for class c is given



(a) Multi-rate loss model (b) Equivalent system

Figure 12: Loss model

by

$$B^c \simeq \begin{cases} \dfrac{1-\gamma^{w_c}}{1-\dfrac{N}{M}} E\left(\dfrac{M}{\delta}, \dfrac{N}{\delta}\right) & \text{if } \gamma \neq 1 \\ \dfrac{w_c M}{V} E\left(\dfrac{M}{\delta}, \dfrac{N}{\delta}\right) & \text{if } \gamma = 1 \end{cases} \tag{15}$$

where γ is the unique positive root of the following equation with z unknown:

$$\sum_{c=1}^{s} w_c X_c z^{w_c} = N \tag{16}$$

and

$$\delta = \frac{\log \dfrac{N}{M}}{\log \gamma}, \tag{17}$$

$$M = \sum_{c=1}^{s} w_c X_c, \tag{18}$$

$$V = \sum_{c=1}^{s} {w_c}^2 X_c. \tag{19}$$

M is the total offered traffic (18) of the s classes of services superposed on the link, V the variance (19) and δ the bandwidth requirement (17) of the equivalent traffic.

B Effective Bandwidth

In multi-rate networks, the bandwidth allocated to each service is fixed to the peak rate. In ATM networks, however, traffic sources have different characteristics and requirements in terms of cell quality of service (QoS). QoS is generally expressed as performance measures with respect to the cell level of the ATM connection: CLP (Cell Loss Probability), CTD (Cell Transfer Delay), CDV (Cell Delay Variation), etc. With statistical multiplexing, it is not efficient to perform a peak rate bandwidth allocation to the incoming calls as it is done in circuit-switching. Effective bandwidth is a statistical measure which lies between the mean rate and the peak rate of a call. It represents the bandwidth required by the call in order to satisfy its QoS on a link. This concept allows the ATM network to be operated as a circuit-switched network. For an incoming call, an effective bandwidth is computed for each VP of the path on which it is routed.

There are several models to compute the effective bandwidth [23, 24, 25, 17, 26, 27, 28, 29]. The approach chosen here is based on the total cell loss rate [30]. It is a generalization of the effective bandwidth given by [29]. It depends on the mean cell rate (MCR) m and the variance σ^2 of the source rate and also the total cell loss rate r and the link capacity N. For each service class, if h is the peak cell rate (PCR) and with $\sigma^2 = m(h-m)$, the effective bandwidth w is given by

$$w = \begin{cases} am\left(1+3z\left(1-x\right)\right) & \text{if } 3z \leq \min\left(3, \dfrac{h}{m}\right) \\ am\left(1+3z^2\left(1-x\right)\right) & \text{if } 3 < 3z^2 \leq \dfrac{h}{m} \\ ah & \text{otherwise,} \end{cases} \tag{20}$$

with

$$x = m/h\,, \tag{21}$$

$$a = 1 - \frac{\log r}{50}\,, \tag{22}$$

$$z = \frac{-2\log r}{N/h}\,. \tag{23}$$

References

[1] M. Minoux. Network synthesis and optimum network design problems: models, solution methods and applications. *Networks*, 19:313–360, 1989.

[2] M.O. Ball. Computing network reliability. *Operations Research*, 27:823–838, July 1979.

[3] S. Chamberland and B. Sansò. Sequential and parallel approaches to incorporate reliability in the synthesis of computer networks. *Journal of Network and Systems Management*, 5(2):131–157, 1997.

[4] A. Girard and B. Sansò. Multicommodity flow models, failure propagation and reliable network design. *IEEE/ACM Transactions on Networking*, 6(1):82–93, February 1998.

[5] B. Sansò. *Fiabilité et routage dans un réseau de télécommunications.* PhD thesis, Université de Montréal, Centre de recherche sur les transports, Université de Montréal, C.P. 6128, Succursale A, Montréal Canada H3C 3J7, 1988.

[6] B. Sansò and F. Soumis. Communications and transportation networks reliability using routing models. *IEEE Transactions on Reliability*, pages 29–38, April 1991.

[7] B. Sansò, F. Soumis, and M. Gendreau. On the evaluation of telecommunications network reliability using routing models. *IEEE Transactions on Communications*, 39(10):1494–1501, October 1991.

[8] D. Medhi. A unified approach to network survivability for teletraffic networks: Models, algorithms and analysis. *IEEE Transactions on Communications*, 42:534–548, 1994.

[9] A. Girard and S. Savignac. Integrated models for reliable network design: numerical results. Technical Report 96-08, INRS-Télécommunications, Place Bonaventure 900 de la Gauchetière Ouest, Niveau C C.P. 644 Montréal, Québec, H5A 1C6, Canada, September 1995.

[10] X.C. Sun and A. Girard. A fast numerical algorithm for ATM network synthesis. *Telecommunications Systems*, 10:355–388, 1998.

[11] D. Medhi and S. Guptan. Network dimensioning and performance of multi-service, multi-rate loss networks with dynamic routing. *IEEE/ACM Trans. on Networking*, 5(6):944–957, December 1997.

[12] B. Sansò, M. Gendreau, and F. Soumis. An algorithm for network dimensioning under reliability considerations. *Annals of operations research*, 36:263–274, 1992.

[13] M. Plante and B. Sansò. A typology for multi-technology, multi-service network synthesis methods. Les Cahiers du GERAD 98-51, GERAD, Pavillon André-Aisenstadt 2920, Chemin de la Tour 4e étage, bureau 4482 Montréal (Québec) Canada H3T 1J4, 1998.

[14] J.S. Kaufman. Blocking in a shared resource environment. *IEEE Transactions on Communications*, 29(10):1474–1481, October 1981.

[15] J.W. Roberts. A service system with heterogeneous user requirements: Application to multi-services telecommunications systems. In G. Pujolle, editor, *Performance of Data Communication Systems and their Applications*, pages 423–431. North-Holland Publishing Co., 1981.

[16] J.F.P. Labourdette and G.W. Hart. Blocking probabilities in multitraffic loss systems: insensitivity, asymptotic behavior, and approximations. *IEEE Transactions on Communications*, 40(8):1355–1366, August 1992.

[17] J.H. Hui. Resource allocation for broadband networks. *IEEE Journal on Selected Areas in Communications*, 6(9):1598–1608, December 1988.

[18] L. Fratta, M. Gerla, and L. Kleinrock. The flow deviation method: An approach to store-and-forward communication network design. *Networks*, 3:97–133, 1973.

[19] B. Gavish and I. Neuman. Capacity and flow assignment in large computer networks. In *Proceedings of IEEE INFOCOM'86*, pages 275–284, April 1986.

[20] B. Gavish and I. Neuman. System for routing and capacity assignment in computer communication networks. *IEEE Transactions on Communications*, 37(4):360–366, 1989.

[21] K. Murakami and H.S. Kim. Optimal capacity and flow assignment for self-healing ATM networks based on line and end-to-end restoration. *IEEE/ACM Transactions on Networking*, 6(2):207–221, 1998.

[22] F. Mobiot. Modèles de flot pour le dimensionnement et la fiabilité de réseaux multi-débits et ATM. Master's thesis, École Polytechnique de Montréal, 1999.

[23] A. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high-speed networks. *IEEE/ACM Transactions on Networking*, 1(3):329–343, 1993.

[24] R.J. Gibbens and P.J. Hunt. Effective bandwidth for multi-type UAS channels. *Queueing Systems*, 9:17–28, 1991.

[25] H. Ahmadi R. Guérin and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications*, 9(7):968–981, 1991.

[26] F.P. Kelly. Effective bandwidths at multi-class queues. *Queueing Systems*, 9:5–16, 1991.

[27] F.P. Kelly. *Notes on effective bandwidths*, pages 141–168. Oxford University Press, 1996.

[28] J. Walrand G. Kesidis and C-S. Chang. Effective bandwidths for multiclass markov fluids and other ATM sources. *IEEE/ACM Transactions on Networking*, 1(4):424–428, 1993.

[29] J.W. Roberts, editor. *Performance evaluation and design of multiservice networks: COST 224 Final Report.* October 1991.

[30] M. Ritter and P. Tran-Gia, editors. *Multi-rate models for dimensioning and performance evaluation of ATM networks: COST 242 Interim Report.* University of Würzburg, June 1994.