ON DYNAMIC CAPACITY MANAGEMENT IN VOICE/FRAME RELAYING NETWORKS

Åke Arvidsson*
Teletraffic Research Centre
The University of Adelaide
G.P.O. Box 498
Adelaide, South Australia 5001
Australia

Abstract

We consider call admission and capacity management in a dynamically reconfigurable network that carries a multitude of traffics. Our approach is a heuristic capacity assignment algorithm designed to maximise an arbitrary profit function. The algorithm is based on the concept of equivalent circuits, i.e. a given relationship between transmission capacities and call handling capabilities expressed as effective numbers of circuits. Such relationships are examined for the two traffics of our application, voice and frame relaying. Methods and results are independent of the transfer mode used.

1 Statistical Multiplexing of Voice/Frame Relaying

1.1 Preliminaries

Statistical multiplexing has become an increasingly complex issue with the integration of new services into telecommunications networks. Traditionally, information sources such as voice have been considered "smooth" which means that they produce information at an even, fixed rate. Most new services, such as data carried over frame relaying [13, 28, 29, 38, 39, 40], are, however, the opposite of smooth, namely "bursty". It is well known that statistical multiplexing of bursty sources poses particular problems as compared to non-bursty sources. The approach taken here is a layered view of statistical multiplexing, as in e.g. [15, 26, 27, 42] and many others. We distinguish between two levels of multiplexing:

- Subscriber multiplexing: Statistical multiplexing of subscribers means exploiting the fact not all subscribers are likely to use their equipment at the same time.
- Connection multiplexing: Statistical multiplexing of connections takes advantage of the fact not all bursty connections are likely to transmit information at the same time.

Statistical gain through multiplexing is always achieved at a certain cost: Multiplexing subscribers will cause rejection of some call attempts and multiplexing connections will result in some information being lost. It is the task of network engineering and network management to supply sufficient capacity and to maintain reasonable call admission policies so that acceptable loss performance is achieved at both levels respectively.

1.2 Service Classes

It is often assumed that a call can be classified as belonging to a certain *class* according to the characteristics of its capacity requirements. Table 1, extracted from [11], gives

*On leave from Department of Communication Systems, Lund Institute of Technology, Lund, Sweden.

Class	$\overline{r}$	r	Burstiness	Activity
Voice	64 kb/s	64 kb/s	1	1.000
Bulk data	0.1 Mb/s	1.5 Mb/s	20	0.067
Transaction data	0.5 kb/s	0.1 Mb/s	200	0.005

Table 1: *Suggested traffic characteristics for voice and data services.*

typical values in terms of average $\overline{r}$ and peak bit rate rate r for three such classes: voice, bulk data and transaction data. We call the peak-to-average ratio $r/\overline{r}$ the *burstiness* of the class while the term degree of *activity* refers to its inverse $\overline{r}/r$.

1.3 Subscriber Multiplexing

Statistical multiplexing of subscribers relies on the fact that only a fraction of the subscribers will use their equipment at the same time. The Erlang B-formula gives the call loss probability under the assumption of incoming call attempts forming a Poisson process

$$E_N(A) = \frac{A^N}{N!} \Bigg/ \sum_{n=0}^{N} \frac{A^n}{n!} \tag{1}$$

for a trunk with N circuits to which a traffic of A Erlang is offered. This is the only relevant multiplexing for voice unless some kind of speech compression such as TASI is used.

1.4 Connection Multiplexing

1.4.1 Modelling

Connection multiplexing operates on calls in progress rather than users and exploits the fact that only a fraction of all (bursty) calls will actually have anything to transmit at the same time. In our model it is assumed that the peak rate r and the average rate $\overline{r}$ are sufficient traffic descriptors and that any two calls are independent.

Consider a transmission channel with capacity C and let N be an upper limit to the number of calls in progress on the channel. Since bursty callers alter between an active and a passive state, the input rate to the channel will vary stochastically between 0 and Nr as the number of active calls n varies between 0 and N. If users' input nr at any time exceeds C, the excess information $nr - C$ will be lost. It is hence noted that information loss will occur only when the number of active callers is in the range $n = \lceil C/r \rceil, \ldots, N$ and that even then only a fraction $nr - C$ of the offered information will be lost.

The equivalent number of circuits associated to a channel is the maximum number of calls N that it can support in order that the expected fraction of lost information is less than some acceptable level ϵ. N depends on the transmission capacity C of the channel, users' activity $p_{\mathrm{a}} = \overline{r}/r$ and peak rate r and on the acceptable loss level ϵ. It is convenient to normalise user access rate r to one and express transmission capacities in units of r and we write

$$N = f(C/r, p_{\mathrm{a}}, \epsilon) \tag{2}$$

The loss can be further reduced, and hence the number of equivalent circuits increased, by supplying a buffer of capacity K in which temporary overloads can be accomodated and stored for later transmission as callers' input falls below $n = \lceil C/r \rceil$. Introducing a buffer means that the number of equivalent circuits is also a function of K and we get

$$N = f(C/r, p_{\mathrm{a}}, K, \epsilon) \tag{3}$$

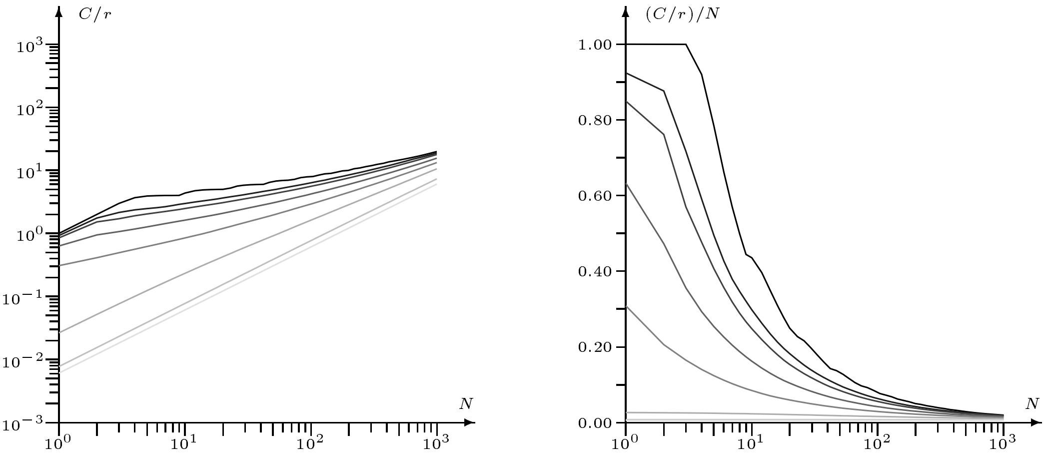

Figure 1: *Capacity requirements (left) and average bandwidths (right) assigned per call for* $K = 0$, *1, 2, 5, 10, 20, 50, 100,* $p_\mathrm{a} = 0.5\%$ *and* $\epsilon = 10^{-8}$.

The equivalent number of circuits may hence determined by studying a queuing system with N users that generate information at peak rate 1 and average rate p_a, one server that processes information at rate C/r and a buffer of size K and calculate the fraction of lost information for successively larger N and given C/r, K and p_a until the loss reaches ϵ. We have used a model developed by Anick et al. [5], in which negative exponentially distributed on and off periods are assumed. In short, the authors derive expressions for the full distribution of the amount of information in an infinite buffer and define the information loss for a finite buffer of length K as the probability that the length of the infinite buffer exceeds K.

1.4.2 Numerical Results

We have investigated a wide range of parameters: p_a = (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20, 0.50, 1.00), K = (0, 1, 2, 5, 10, 20, 50, 100)[1] and $\epsilon = (10^{-6}, 10^{-8}, 10^{-10})$. The case depicted in figure 1 refers to $p_\mathrm{a} = 0.005$, $\epsilon = 10^{-8}$ and successively brighter curves refer to the different buffer sizes. The diagram to the left shows the transmission capacity C/r required to provide a certain number N of equivalent circuits and the diagram to the right shows the allocated capacity per circuit $C/(rN)$ when N calls are multiplexed.

In conclusion, the numerical investigation shows that

- Small N and no buffer mean that no multiplexing gain is achieved (slope equal to one in the diagram to the left; i.e. the number of calls N increase at the same rate as the allocated capacity C/r) and the equivalent bandwidth is equal to users' peak rate (vertical line in the diagram to the right; i.e. calls effectively occupy their entire peak rate)

- Increasing the number of calls N in parallel or providing larger buffers K improves multiplexing gain (slope less than one in the diagram to the left; i.e. the number of calls N increase faster than allocated capacity C/r), which results in lower equivalent bandwidth (falling line in the diagram to the right; i.e. calls effectively

[1] Note that K is normalised as in [5]: One unit of buffer space is the average amount of information one user generates during one, continuous period of activity.

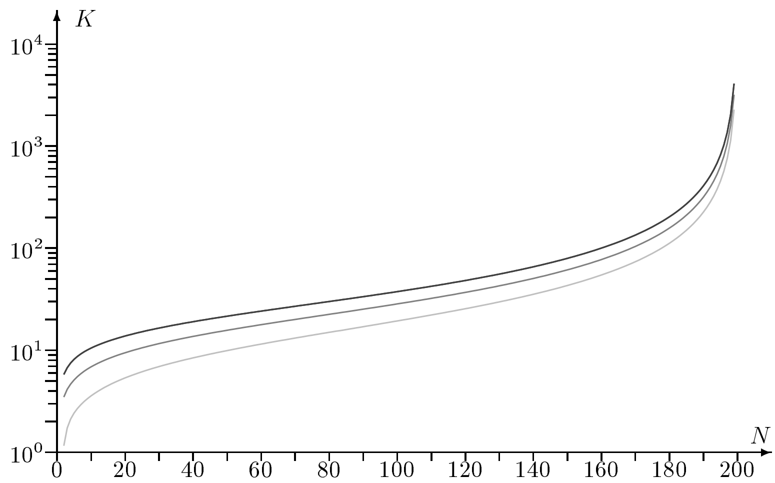

Figure 2: *K as a function of N for $C/r = 1$, $\epsilon = 10^{-11}$, 10^{-9}, 10^{-7} and 10^{-5} and $p_{\mathrm{a}} = 0.5\%$.*

occupy less than their peak rate and tend towards their average rate as $N \to \infty$ and $K \to \infty$).

- Curves for buffer sizes larger than $K = 100$ tend to be very close to the one for $K = 100$, i.e. the gain obtained by increasing buffers beyond 100 units appears to be relatively small.
- The importance of a buffer lessens the more calls we have in parallel, i.e. larger N (the curves approach each other in figure 1.
- The positive impact of the buffer decreases as the activity p_{a} increases and, as expected, vanishes completely for $p_{\mathrm{a}} = 1$.
- Increasing the acceptable loss ϵ results in improved statistical gain for small N, while the loss value has less impact on multiplexing gain for larger N.
- Lower activity means that fewer calls in parallel are required to obtain significant statistical gain but that more calls must be multiplexed before full gain is achieved.

Another view of the results is presented in figure 2 which shows buffer space K as a function of the number of virtual circuits N for activity $p_{\mathrm{a}} = 0.005$, capacity $C/r = 1$ and, by successively brighter curves, for different loss probabilities, $\epsilon = (10^{-11}, 10^{-9}, 10^{-7}, 10^{-5})$. The figure indicates that

- The growth in buffer demand between a utilisation of about $N = 20/200 = 10\%$ an $N = 180/200 = 90\%$ is relatively small but that
- to obtain a utilisation higher than 90% is very costly in terms of the buffer size required.

1.4.3 Comparison to Simulations

To get an idea of how applicable the results above are to frame relaying in particular, a scenario with a frame relaying network and N independent users was formed and implemented as a simulation programme. Our model is limited to N users with individual

Parameter	Value(s)	
N	100	
r	256,000	
ρ	≈ 0.9	
Z	$m = 0.300$	$C = 0.111$
η	1:1	
D_{U}	$m = 11.4$	$C = 1.6$
T_{U}	$m = 18.3$	$C = 1.2$
D_{S}	$m = 1,200$	$C = 3.2$
T_{S}	$m = 0.42$	$C = 8.1$
f	1	
D_{F}	$m = 20,000$	$C = 10$
D_{C}	$m = 12$	$C = 0$
T_{R}	$m = 0.042$	$C = 8.1$
α_N	0.95	
α_R	0.50	

Table 2: *Scenario parameters. Data volumes are expressed in octets and times in seconds.*

access lines of capacity r and a transmitter with capacity C, C chosen to normalise the load ρ, supported by an infinite buffer. The rest of the network was, partly due to difficulties in making relevant assumptions, simplified to a normally distributed end-to-end delay with mean m_Z and standard deviation σ_Z.

Lack of sufficient, relevant data prohibited discrimination of users into distinct business categories, although examples show that different classes of business users may generate quite different traffics [37, 43]. Instead, users' activities were classified as interactive sessions or file transfer sessions, both of infinite duration in order to quickly achieve and maintain a steady state. The ratio of interactive users to file transfer users is denoted by η.

Interactive A-side users issue commands, wait for response, evaluate responses for some time and then issue new commands and so on. B-side users receive commands, compute responses, transmit them and await more commands. The length of a command is denoted by D_{U} and the length of a response by D_{S} while evaluation and computation times are denoted by T_{U} and T_{S} respectively.

A-side file transfer users transmit f frames of file data, evaluate received acknowledgements etc. after which new sets of f frames are normally transmitted. B-side users transmit acknowledgements, evaluate received data etc. after which new acknowledgements are normally transmitted. If, however, evaluation at any side results in an error, a restoration phase is entered in which users send a control message, await responses, evaluate responses and either send new control messages or return to the normal mode. File volumes are denoted by D_{F}, control message volumes by D_{C} and the response time is denoted by T_{R} for both ends respectively. Protocol evaluation is assumed to be random: The probability of remaining in normal mode is α_N and the probability of sending one more restoration message is α_R.

Numerical values are summarised in table 2 and were taken from in particular Pawlita [37] but also from [8, 9, 22, 24, 28, 31, 36, 45], refer to [6] for a detailed discussion. Normal distributions were used for all stochastic data volumes and times. Some parameters were varied, table 3 summarises the modifications.

Buffer utilisation was sampled in all simulations and compared to analytical results for the same activities p_{a} as recorded in the simulations. Figure 3 refers to run 1 in table

Run	Modified parameter	Comment
1	—	Standard run
2	$N = (52 + 52)$	50% traffic via common gateway (rate r)
3	$m(\mathrm{Z}) = 0.150$, $C(\mathrm{Z}) = 0.1133$	Calls over long land/submarine cable
4	$m(\mathrm{Z}) = 0.015$, $C(\mathrm{Z}) = 0.1133$	Calls over short land/submarine cable
5	$N = 24$	Decreased number of calls
6	$N = 400$	Increased number of calls
7	$\eta = 1 : 4$	Increased fraction of file transfers
8	$\eta = 4 : 1$	Increased fraction of interactive users
9	$f = 4$	Multiple frame forward file transfer
10	$\alpha_N = 0.99$	Improved file transfer security
11	$\alpha_R = 0.1$	Simplified file transfer recovery

Table 3: *Variations of scenario parameters. Data volumes are expressed in octets and times in seconds.*

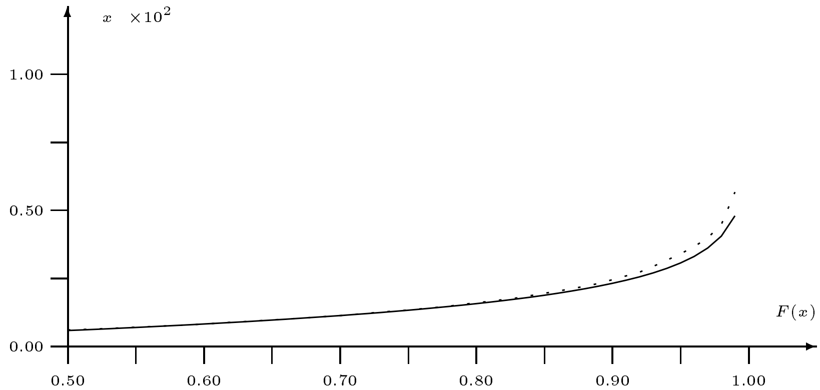

Figure 3: *Probability distributions obtained for the standard scenario.*

3 and shows the buffer level k as a function of the distribution function $F(k)$, i.e. the diagram can be interpreted as the buffer length k required to reach a certain probability of acceptance $F(k)$. The fact that the curve obtained through simulation (dotted line) is above the one obtained though analysis (solid line) means that the analytical model underestimates buffer utilisation.

Similar graphs were obtained for all runs in table 3 with overestimations in runs 3, 4, 6, 10 and 11 and underestimations in runs 1, 2, 5, 7, 8 and 9. It is concluded that

- The analytical model is qualitatively valid under the circumstances of the scenario, but that
- its quantitative applicability depends on details of users' activities which the analytical model does not provide parameters to capture.
- The assumptions on file transfer protocols were the most critical ones in table 3: While the analytical model slightly underestimated the buffer occupancy in run 1, it exhibited considerable overestimations in runs 10 and 11.
- Dispersion and skewness for active periods were both pretty close to unity, which suggests that active periods are relatively accurate modelled by a negative exponential distribution, but

- the large values noted for both dispersion and skewness of passive periods indicate that the distribution of these is far from negative exponential.

The last item hints at a point of further research. However, although analytical models may be made to match simulations, the problems of predicting user behaviour, such as volumes and protocol details, remain. A better approach may therefore be to take negotiated parameters [28] into account, although this is reportedly not done in present frame relaying networks in the United States [35, 44].

1.5 Discussion

We have taken a layered view of statistical multiplexing: Subscriber multiplexing and connection multiplexing. The bridge is the concept of *equivalent circuits*, also known as the *equivalent trunk* approach [33]. Comparing it to the relevant, alternative approach, *equivalent bandwidth* [21, 25, 33], we note the following advantages:

- The rich theory in the area of circuit switched networks is immediately applicable also for networks that employ a mixture of switching modes through the concept of equivalent circuits.
- The simple acceptance procedure obtained: A new call bid is accepted if less than N calls currently are in progress, otherwise the bid is rejected.
- Service integrity is improved since violations in one class, such as excessive input, does not lead to service degradations in other classes.
- Multiplexing homogeneous, bursty users, i.e. users of the same class only, permits detailed queuing models that can be designed to reproduce particular class characteristics.
- Dimensionality and complexity of such models are likely to be such that the number of equivalent circuits can be calculated in real time and/or stored in tables.

The disadvantages of the approach are best illustrated by considering a number of transmission channels between two points, each of which is dedicated for a particular class. In a situation when all equivalent circuits in class t are engaged will t-calls be rejected, despite possible idle capacity between the two points in question which is reserved for other classes.

On the other hand, modelling heterogeneous sources necessitates rougher queuing models and, in turn, increased safety margins that largely will be unused. We believe that the drop in utilisation caused by our approach need not be greater, provided appropriate channel dimensioning and updating are employed.

2 Dynamic Capacity Management

2.1 Virtual Path Networks

We consider a telecommunication network consisting of nodes and links. Nodes are switching systems, local switches or transit switches, while the links are digital transmission channels such as coaxial cables, micro wave links, satellite links or optical fibres. This network is referred to as the physical network. For a number of reasons it may be desirable to enhance a basic, physical network by creating a higher layer of (sub)networks which are logically independent of each other and of the underlaying physical network. Such networks are formed by reserving certain amounts of transmission capacity on links and cross connecting the reserved channels through possible, intermediate transit nodes. The combination of reserved capacities on links and possible cross connections is called a virtual path (VP).

Two nodes being interconnected by a VP is, from the point of view of acceptance, routing and establishment of calls, essentially identical to having a single, direct route between them to which they have exclusive access. The process of creating and/or rearranging a network of VPs between a number of end nodes is called capacity management [11], bandwidth management [23] or bandwidth switching [1] and is performed at a network management centre.

Since the optimality of a certain VP configuration is dependent on both link capacities and currently offered traffics, VP assignments must be reevaluated in response to changes, typically on the order of a few call holding times. The act of continuously updating of VPs is referred to as dynamic capacity management (DCM). To cope with traffic fluctuations faster than the order of a mean holding time, overflows and alternate routing can be applied in the logical VP-network. The proper selection of overflow routes can be handled by conventional methods for circuit switched networks.

Some of the motives behind VP-networks and DCM are

Network Costs: In a network with direct VPs between end nodes, switching functions such as acceptance, routing and establishment of calls are carried out directly between these nodes and the functions of transit nodes are reduced to one: Switching. Reducing the functionality of a transit switch, in turn, opens up possibilities of considerable simplifications and hence cutting costs [2, 10, 11, 34].

Clearly, possible savings are partly offset by reduced link utilisation, but studies have shown that substantial cost savings can be made [10, 11, 12, 34]. Moreover, the above cost tendencies are expected to continue, whence a network structure with fewer nodes and a sparse grid of high speed transmission links is likely to prove an increasingly attractive alternative in the future.

Simplified Multiplexing: By creating separate VPs for different classes is multiplexing simplified in that problems related to multiplexing sources with different characteristics are eliminated. Moreover, since calls in VP networks are carried end-to-end on reserved capacity, the decision of whether to accept a call request or not only invokes the originating node.

Faster Call Handling: The exclusion of all transit nodes reduces the total processing time for a call bid and the simplified multiplexing shortens the remaining work. Signaling will be performed directly end-to-end and lessen signal transfer and processing times while the resulting, reduced signal load further contributes to speeding up call handling.

Improved Network Management Capabilities: VPs provides a quick means to reroute traffics in case of temporary overloads or equipment failures in that only the contents of transit node routing tables need to be modified.

A Means for Providing New Services: VPs can easily be set up between a number of access points of a customer to form a closed, private subnetwork under control of the user.

In ATM networks VPs are implemented by splitting the address part of the cell header into two parts: a virtual path identifier (VPI) and a virtual circuit identifier (VCI) [2, 27, 33, 41]. Though transit switching still must be carried out on cell-by-cell basis, transit nodes are relieved from all functions but switching and smaller addresses mean smaller and cheaper switching networks. The VP-concept can also be implemented directly in the synchronous digital hierarchy (SDH) [7, 17]. Transit nodes for SDH, digital cross connects and add/drop multiplexers, are currently being developed or are already available [3, 4].

2.2 Algorithms for Dynamic Capacity Management

2.2.1 Existing Algorithms

We have found DCM algorithms or algorithms closely related to DCM published by Gopal et al. [19, 20], Herzberg [23], Evans [14], Gersht et al. [18] and Mase et al. [32]. Algorithms are also outlined by Hui et al. [27]. Finally, Mase et al. [33] discuss such algorithms in terms similar to those in [14, 18, 27] but without going into any detail. Summing up on these it is found that

- most algorithms explicitly or implicitly presume linear equivalent bandwidths; the diagram to the right in figure 1 clearly shows that the equivalent bandwidth not always linear in the range 0–1000 calls,
- most algorithms explicitly or implicitly presume the existence of predefined paths for all VPs; selecting and storing predefined paths for every possible situation of overload and/or equipment failures is far from trivial, and
- some algorithms produce real valued solutions; VPs in PDH- [16] or SDH-based networks must be determined as an integer number of circuits or virtual containers respectively.

As none of the algorithms appears immediately useful for our problem, we have developed one capacity and that does not require predefined paths. The algorithm is a heuristic and as such it does not guarantee that the final solution is a global optimum. On the other hand, the "optimality" guaranteed by some of the algorithms above is not entierly global but only applies only under the condition of a fixed set of paths.

2.2.2 Our Algorithm

Our algorithm is referred to as the automatic greedy algorithm[2] and exhibits some similarities to algorithm developed by Gopal et al. [19, 20], though we permit several service classes between OD-pairs, non-linear equivalent bandwidths and search dynamically for routes. Its main steps can be described as follows:

1. Read the tables that provide the relationships between capacities and circuits.
2. Read the number of nodes N, link capacities $C_{o,d}$ (expressed as units of capacity) and offered traffics $A_{t,o,d}$ (expressed as Erlangs).
3. Assign high, initially acceptable call loss levels $\alpha(t)$ for all service classes t.
4. Find the shortest paths available to each OD-pair o, d and each service class t.
5. Compute the gain achieved for each traffic t, o, d if one unit of capacity is added to its shortest path.

[2] "Automatic" refers to that the algorithm finds VP routes itself and "greedy" to that every step performed aims at achieving the highest immediate profit without taking long term consequences of a particular decision into account.

6. Compute the loss paid for each traffic t, o, d if one unit of capacity is added to its shortest path.

7. Find the OD-pair and the service class $t_{\max}, o_{\max}, d_{\max}$ that yields the highest gain/loss ratio.

8. If the highest gain/loss ratio is 0 then go to step 11, else proceed to step 9.

9. Assign one unit of capacity to the OD-pair and service class $t_{\max}, o_{\max}, d_{\max}$ that provides the highest gain/loss ratio.

10. Go to step 4.

11. If low, ultimately acceptable loss levels $\alpha(t)$ have been reached for all t or all capacity has been assigned then stop, else proceed to step 12.

12. Reduce currently acceptable loss levels $\alpha(t)$ for all t.

13. Go to step 4.

It is observed that the algorithm tries to successively distribute the available capacity to VPs in such a way that a minimum amount of capacity is used and that maximum value is obtained for each unit of capacity, by using shortest paths and by chosing the VP with the highest gain/loss ratio respectively. The successive reductions of acceptable loss serve to ensure that reasonable fairness in grade of service is maintained among VPs and to make sure that assigned capacity will actually carry a reasonable amount of traffic. The algorithm will terminate when for every VP either (i) a final, predetermined, desirable loss level has been reached or (ii) no more capacity is available for VPs still suffering from unacceptable losses.

The tables referred to in step 1, one for each traffic t, give the number equivalent of circuits $N(t, i)$ for class t and $C = i$ units of capacity, $i = 1, 2, \ldots$. (The quantity "unit of capacity" is the integer unit in which capacity is expressed: For PDH an integer number of 64 kb/s channels, for SDH the smallest virtual container used and for pure ATM networks a rate in cells/second large enough to carry a call of any class.)

Our implementation uses an initial setting of loss levels at step 3 of 50% and reduces it in step 11 to the ultimately acceptable level of 0.05% through two intermediate levels of 5% and 0.5% respectively.

Shortest paths in step 4 are determined using Floyd's algorithm [30] with the length associated to link o, d designed to find the shortest path in number of links traversed with preferential treatment to paths having more spare capacity left than other paths of equal length

$$\text{Length}(o, d) = \begin{cases} 1 + \dfrac{1}{{C'_{o,d}}^4} & C'_{o,d} > 0 \\ \infty & C'_{o,d} = 0 \end{cases} \tag{4}$$

where $C'_{o,d}$ denotes the remaining, not yet assigned capacity on link o, d.

In step 5, the gain $G(t, o, d)$ for t, o, d is the additional t-traffic that would be carried from o to d if one unit of capacity was added to its currently shortest path

$$G(t, o, d) = A_{t,o,d}[E_{N'_{t,o,d}}(A_{t,o,d}) - E_{N_{t,o,d}}(A_{t,o,d})] \tag{5}$$

where $E_N(A)$ is the Erlang loss formula (1), $N_{t,o,d}$ is the total number of circuits currently available to t, o, d and $N'_{t,o,d}$ is the total number of circuits that would be available to t, o, d if one unit of capacity was added to its shortest path. Observing that circuits cannot be formed by joining capacities from several distinct routes, both N and N' are determined for each route individually from the tables read in step 1 and then summed. Moreover, we define $G(t,o,d) \equiv 0$ if the capacity already assigned to t,o,d is enough to ensure that a certain degree of service $\alpha(t)$ has been achieved, that is if $E_{N_{t,o,d}}(A_{t,o,d}) < \alpha(t)$. We also set $G(t,o,d) \equiv 0$ if no more capacity is available between o and d.

The loss $L(t,o,d)$ associated to t,o,d addressed in step 6 is the sum of all gains that can be achieved at the same point and that require some of the capacity also requested by t,o,d:

$$L(t,o,d) = \sum_{t'=1}^{T} \sum_{o'=1}^{N-1} \sum_{d'=o'+1}^{N} I(\mathcal{L}_{t,o,d} \cap \mathcal{L}_{t',o',d'} \neq \emptyset)\, G(t',o',d') \tag{6}$$

where $\mathcal{L}_{t,o,d}$ is the set of links traversed by the shortest path for t,o,d and $I(\cdot)$ is an indicator function taking the value of 1 if its argument is true, otherwise 0.

2.2.3 Discussion

An obvious advantage with the proposed algorithm is its robustness. That is, unlike methods based on mathematical programming, it will remain stable and converge at the same speed for all types of non-linearities and discontinuities in gain and loss functions and irrespective of the ways in which routes for VPs are chosen. These properties leave full freedom to modify and extend the algorithm to meet particular demands such as

- Biased selection of routes: If, for technical or economical reasons, a biased selection of routes may be desirable this can readily be incorporated into the cost function (4) as penalties associated with certain links for some or all service classes. Such penalties may be dependent on the capacity already seized on that link.

- A limitation to the number of distinct physical paths can be imposed by not allowing more than a prescribed number of distinct routes $R_{\max}$ for each traffic. For OD-pairs that already has $R_{\max}$ distinct routes, the shortest path is defined as the shortest of these.

- Predetermined routes: If, for some reason, predefined routes are preferred and available, automatic selection of routes can easily be inhibited by a modification similar to the one above.

- Profit maximisation: The gain $G(t,o,d)$ need not be just carried traffic but can, for example, be scaled both according to traffic class, origin and destination, by multiplying (5) by a weight $w = w(t,o,d)$ based on e.g. revenues.

- Traffic concentration: By defining a minimum allowable gain $G_{\min}$ can small and potentially uneconomical traffics be eliminated from a primary run. Such traffics can then be taken care of in a secondary run in which the VPs already established in the first run are seen as potential carriers, i.e. as virtual links. By thus carrying smaller traffics on top of larger ones can both be carried at a lower cost due to the non-linearities of the equivalent bandwidth (connection level) and Erlang's B-formula (subscriber level).

2.3 Numerical Results

The proposed algorithm was applied to voice and frame relaying networks under the assumptions of table 4. All networks were analysed for two different capacity resolutions (two interpretations of "unit of capacity"): 1.920 Mb/s and 0.640 Mb/s.

The parameters for voice can be considered as standard values and the values for frame relaying were selected taking into account CCITT recommendation I.233 on frame

Class	r	$\overline{r}$
Voice	64.00 kb/s	64.00 kb/s
Frame relaying	2,048 kb/s	20.48 kb/s

Table 4: *Assumed characteristics of voice and frame relaying.*

Capacity Units	Voice Circuits (1,920)	Voice Circuits (640)	Frame relaying Circuits (1,920)	Frame relaying Circuits (640)
1	30	10	77	25
2	60	20	154	51
3	90	30	232	77
4	120	40	309	102
...	...	...	...	...
100	3000	1000	8025	2621
...	...	...	...	...

Table 5: *Relations between units of capacity and equivalent circuits.*

mode bearer services and relevant application data reviewed in section 1.4.3. Finally, assuming PDH/STM for voice and $r = 2.048$ Mb/s, $K = 100$ and $\epsilon = 10^{-8}$ for frame relaying we obtain the relations between transmission capacities and circuits depicted in table 5 for the two resolutions respectively.

It is observed from table 5 that the first unit of capacity will carry about 2.5 as many frame relaying calls as voice calls, which suggests a charging policy for frame relaying calls of about $1/2.5 = 0.4$ compared to voice calls. Allowing for paying off the extra equipment needed for frame relaying, we have increased this number and assumed a profit factor $w(t, o, d)$, section 2.2.3, of 0.5 for $t =$ frame relaying.

In total 100 networks with varying number of nodes M and inter connectivity S were studied. An example of such a network, for which $M = 10$ and $S =$ Sparse, and an abbreviated form of the output of the algorithm is given in appendix A.

Results are summarised in table 6. For each of the two traffics we give the overall average loss E_{tot} and the average loss per OD-pair $E_{\mathrm{OD-pair}}$ in % and for links we give the levels of utilisation U_{tot} expressed as carried traffic per 2 Mb/s link.

As expected, the table shows that finer granularity leads to a generally better designed network: lower loss levels, less capacity required and higher utilisation. On the other hand as the execution time of the algorithm is approximately proportional to the granularity,

		Voice				Frame relaying				Links	
N	S	E_{tot}		$E_{\mathrm{OD-pair}}$		E_{tot}		$E_{\mathrm{OD-pair}}$		U_{tot}	
		(1.920)	(0.640)	(1.920)	(0.640)	(1.920)	(0.640)	(1.920)	(0.640)	(1.920)	(0.640)
10	Sparse	0.4	0.2	0.4	0.2	0.3	0.1	0.2	0.1	15.2	16.0
10	Dense	0.3	0.1	0.2	0.1	0.1	0.1	0.1	0.1	17.7	18.7
20	Sparse	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.0	9.2	9.7
20	Dense	0.3	0.1	0.2	0.1	0.1	0.1	0.1	0.0	13.5	14.4

Table 6: *Results for DCM with two different degrees of resolution over 100 randomly generated networks.*

the gain obtained from additional precision may be lost in fewer and slower updates in a dynamic environment. The impact of network dynamics to network performance under different DCM updating regimes is currently in progress.

3 Conclusions

The properties of frame relaying traffic with respect to statistical multiplexing have been investigated and a method to mathematically relate given traffic characteristics, transmission capacity, buffer space and acceptable loss level to a maximum permissible number of parallel connections (the equivalent number of circuits) has been proposed end evaluated by means of simulations. The concept of equivalent circuits has been applied in a new, flexible capacity allocation algorithm for multi-service networks based on virtual paths, for example reconfigurable networks.

Current work is concentrated on the relationship between network dynamics and the need to update virtual path designs. Further studies of frame relaying with respect to statistical multiplexing based on negotiated parameters are also being planned.

Acknowledgement

The permission of Australian and Overseas Telecommunications Corporation Limited to publish this work is gratefully acknowledged. The views and opinions expressed in this work are those of the author, and do not necessarily imply AOTC policy or future service offerings.

References

[1] Addie, R. & Warfield, R.: *Bandwidth Switching and New Network Architectures*, paper no. 2.3iiA.1 in Proc. 12th Int. Teletraffic Cong., Torino, Italy (1988).

[2] Addie, R., Burgin, J. & Sutherland, S.: *Information Transfer Protocols for the Broadband ISDN*, paper no. 22.6.1 in Proc. IEEE Globecom '88, Hollywood, Florida, U.S.A. (1988).

[3] Ahmed, H., Brégi, P., Haag, J.-P. & Wiest, B.: *Synchronous Transmission Systems and Cross Connects for Broadband Networks*, Electr. Commun., vol. 64, no. 2/3, pp. 194–203 (1990).

[4] Andersson, J.-O., Improta, S. & Evangelisti, G.: *Digital Cross-Connect Systems — A Tool for Managing the Transport Network*, paper no. C6.5 in Proc. XIIIth Int. Switching Symp., Stockholm, Sweden (1990).

[5] Anick, D., Mitra, D. & Sondhi, M.: *Stochastic Theory of a Data-Handling System with Multiple Sources*, Bell Syst. Tech. J., vol. 61, no. 8, pp. 1871–1894 (1982).

[6] Arvidsson, Å.: *A Study on Statistical Multiplexing and Dynamic Capacity Management in Voice/Frame Relaying Networks*, Report no. 7/92, Teletraffic Research Centre, The University of Adelaide, Adelaide, South Australia, Australia (1992).

[7] Astani, K., Harrison, R. & Ballart, R.: *CCITT Standardization of Network Node Interface of Synchronous Digital Hierarchy*, IEEE Commun. Mag., vol. 28, no. 8, pp. 15–20 (1990).

[8] Boggs, D., Mogul, J. & Kent, C.: *Measured Capacity of an Ethernet: Myths and Reality*, Comp. Commun. Rev. vol. 18, no. 4, pp. 222–234 (1988).

[9] Bottomley, L., Nilsson, A. & Blatecky, A.: *Traffic Measurements on a Working Wide Area Network*, pp. 767–772 in Proc. 13th Int. Teletraffic Cong., Copenhagen, Denmark (1991).

[10] Brungard, D., Grotjohann, H. & Kallenberg, P.: *Impact of New Transmission Technologies on the Network Architecture, Network Management and Economics*, paper no. A1.2 in Proc. XIIIth Int. Switching Symp. Stockholm, Sweden (1990).

[11] Burgin, J.: *Management of Capacity and Control in Broadband ISDN*, Int. J. of Dig. and Ana. Cabled Sys., vol. 2, pp. 155–165 (1989).

[12] Burgin, J.: *Broadband ISDN Resource Management*, Computer Netw. and ISDN Syst. vol. 20, no. 1–5, pp. 323-331 (1990).

[13] IEEE Communications Magazine: *Special Issue on Frame Relaying*, vol. 30, no. 3 (1992).

[14] Evans, S.: *A Mathematical Model and Related Problems of Optimal Management and Design in a Broadband Integrated Services Network*, J. Australian Math. Soc., ser. B, vol. 31, part 2, pp. 150–175 (1989).

[15] Filipiak, J.: *M-Architecture: A Structural Model of Traffic Management and Control in Broadband ISDNs*, IEEE Comm. Mag., vol. 25, no. 5, pp. 25–31 (1989).

[16] CCITT Recommendations G.702–G.704: *Digital Hierarchies*, CCITT, ITU, Geneva, Switzerland.

[17] CCITT Recommendations G.707–G.709: *Synchronous Digital Hierarchy*, CCITT, ITU, Geneva, Switzerland.

[18] Gersht, A. & Shulman, A.: *Optimal Routing in Circuit Switched Networks*, IEEE Trans. on Commun., vol. 37, no. 11, pp. 1203–1211 (1989).

[19] Gopal, G., Kim, C.-K. & Weinrib, A.: *Dynamic Network Configuration Management*, paper no. 302.2 in IEEE Int. Conf. on Commun., Atlanta, Georgia, U.S.A. (1991).

[20] Gopal, G., Kim, C.-K. & Weinrib, A.: *Algorithms for Reconfigurable Networks*, pp. 341–347 in Proc. 13th Int. Teletraffic Cong., Copenhagen, Denmark (1991).

[21] Guérin, R., Ahmadi, H. & Naghshineh, M.: *Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks*, IEEE J. Sel. Areas in Commun., vol. 9, no. 7, pp. 968–981 (1991).

[22] Gusella, R.: *Characterizing the Variability of Arrival Processes with Indexes of Dispersion*, IEEE J. Sel. Areas in Commun., vol. 9, no. 2, pp. 203–211 (1991).

[23] Herzberg, M.: *Network Bandwidth Management — A New Direction in Network Management*, pp. 218–225 in Proc. 6th Australian Teletraffic Sem., Wollongong, New South Wales, Australia (1991).

[24] Hicks, P.: *Performance Measurements of Packet Switched Data Networks*, paper 8.2 in Proc. 1st Australian Teletraffic Sem., Melbourne, Victoria, Australia (1986).

[25] Hong, D. & Suda, T.: *Congestion Control and Prevention in ATM Networks*, IEEE Netw. Mag., vol. 5, no. 4, pp. 10–16 (1991).

[26] Hui, J.: *Resource Allocation for Broadband Networks*, IEEE J. Sel. Areas in Commun., vol. 6, no. 9, pp. 1598–1608 (1988).

[27] Hui, J., Gursoy, M., Moayeri, N. & Yates, R.: *A Layered Broadband Switching Architecture with Physical or Virtual Path Configurations*, IEEE J. Sel. Areas in Commun., vol. 9, no. 9, pp. 1416–1426 (1991).

[28] CCITT Recommendation I.233: *Frame Mode Bearer Services*, CCITT, ITU, Geneva, Switzerland.

[29] Keough, L.: *Special Report: Frame Relay*, Data Communications, vol. 21, no. 7, pp. 67–86 (1992).

[30] Kleinrock, L.: *Queueing Systems, Volume II: Computer Applications*, John Wiley & Sons, New York (1976).

[31] Marshall, W. & Morgan, P.: *Statistics of Mixed Data Traffic on a Local Area Network*, paper 4.1A–1 in Proc. 11th Int. Teletraffic Cong., Kyoto, Japan (1985).

[32] Mase, K. & Imase, M.: *An Adaptive Capacity Allocation Scheme in Telephone Networks*, IEEE Trans. on Commun., vol. 30, no. 2, pp. 354–359 (1982).

[33] Mase, K. & Shioda, S.: *Real-Time Network Management for ATM Networks*, pp. 133–140 in Proc. 13th Int. Teletraffic Cong., Copenhagen, Denmark (1991).

[34] Moondra, S.: *Impact of Emerging Switching-Transmission Cost Tradeoffs on Future Telecommunications Network Architectures*, IEEE J. Sel. Areas in Commun., vol. 7, no. 8, pp. 1207–1218 (1989).

[35] Newnham, S.: *Personal communication*, Telecom Australia Research Laboratories, Melbourne, Victoria, Australia (May, 1992).

[36] Okada, H.: *Performance Evaluation of Medium-Access Control Protocols in Local Area Networks*, paper 4.1A–4 in Proc. 11th Int. Teletraffic Cong., Kyoto, Japan (1985).

[37] Pawlita, P.: *Two Decades of Data Traffic Measurements: A Survey of Published Results, Experiences and Applicability*, paper 5.2A.5 in Proc. 12th Int. Teletraffic Cong., Torino, Italy (1988).

[38] CCITT Recommendation Q.922: *ISDN Data Link Layer for Frame Mode Bearer Services*, CCITT, ITU, Geneva, Switzerland.

[39] Rahnema, M.: *Frame Relaying and the Fast Packet Switching Concept and Issues*, IEEE Netw. Mag., vol. 5, no. 4, pp. 18–23 (1991).

[40] Roy, A: *Frame Relay and its Transportation Dilemma*, pp. 393–400 in Proc. Singapore Int. Conf. on Netw., SICON '91 (1991).

[41] Sato, K.-I., Ohta, S.. & Tokizawa, I.: *Broad-Band ATM Network Architecture Based on Virtual Paths*, IEEE Trans. on Commun., vol. 38, no. 8, pp. 1212–1222 (1990).

[42] Schoute, F.: *Simple Decision Rules for Acceptance of Mixed Traffic Streams*, paper no. 4.2A.5 in Proc. 12th Int. Teletraffic Cong., Torino, Italy (1988).

[43] Trierscheid, R. & Nüssler, K.: *Investigations on Traffic Characteristics and Network Behaviour in Datex-P*, paper 5.4A–3 in Proc. 11th Int. Teletraffic Cong., Kyoto, Japan (1985).

[44] Warfield, R.: *Personal communication*, Telecom Australia Research Laboratories, Melbourne, Victoria, Australia (May, 1992).

[45] CCITT Recommendation X.25: *Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Terminals Operating in the Packet Mode and Connected to Public Data Networks by Dedicated Circuit*, CCITT, ITU, Geneva, Switzerland.

A A Sample Network

This appendix documents one of the 100 networks considered in table 6: Figure 4 shows the layout of the physical network, the transmission capacities on each link (in units of 1.920 Mb/s) is given by table 7 and the offered traffics are given (in Erlangs) by tables 8 and 9 for voice and frame relaying respectively.

Finally, the last two pages give abbreviated printouts from the algorithm after finishing the VP design for each of the two resolutions respectively. For each class are given

- total offered and carried traffics over the network as a whole,
- the mean, standard deviation, maximum and minimum loss per OD-pair,
- the mean, standard deviation, maximum and minimum of the number of different physical paths used and considered but not used per OD-pair and
- the mean, standard deviation, maximum and minimum of the length of the physical paths used and considered but not used per OD-pair.

For the network we also give

- The total number of capacity units provided and used over the network as a whole and
- the mean, standard deviation, minimum and minimum of the utilisation per physical link.

OD	C	OD	C	OD	C	OD	C	OD	C	OD	C
1– 8	37	1– 9	29	2– 5	70	2– 9	83	3– 7	59	3–10	8
4– 5	61	4– 6	96	6– 7	95	7– 8	102	7–10	34	8– 9	80
8–10	39										

Table 7: *Network capacities.*

OD	A	OD	A	OD	A	OD	A	OD	A
1– 2	56.2	1– 3	50.6	1– 4	166.9	1– 5	90.9	1– 6	105.3
1– 7	189.5	1– 8	24.9	1– 9	205.7	1–10	11.8	2– 3	74.2
2– 4	201.6	2– 5	201.7	2– 6	189.5	2– 7	106.2	2– 8	70.8
2– 9	178.4	2–10	98.6	3– 4	244.7	3– 5	12.1	3– 6	36.6
3– 7	110.6	3– 8	161.5	3– 9	101.2	3–10	113.4	4– 5	38.8
4– 6	379.8	4– 7	209.2	4– 8	281.8	4– 9	118.3	4–10	17.1
5– 6	184.2	5– 7	56.5	5– 8	175.6	5– 9	88.8	5–10	78.6
6– 7	98.0	6– 8	68.7	6– 9	65.8	6–10	206.6	7– 8	110.8
7– 9	147.4	7–10	197.8	8– 9	26.6	8–10	203.1	9–10	178.2

Table 8: *Offered voice traffics.*

OD	A	OD	A	OD	A	OD	A	OD	A
1– 2	43.0	1– 3	90.3	1– 4	47.2	1– 5	133.4	1– 6	51.3
1– 7	115.2	1– 8	150.6	1– 9	128.2	1–10	92.7	2– 3	77.6
2– 4	100.7	2– 5	44.5	2– 6	99.4	2– 7	195.3	2– 8	77.3
2– 9	96.1	2–10	124.3	3– 4	55.8	3– 5	94.5	3– 6	100.8
3– 7	123.2	3– 8	83.6	3– 9	143.8	3–10	77.7	4– 5	158.5
4– 6	246.4	4– 7	248.8	4– 8	51.6	4– 9	108.3	4–10	99.2
5– 6	186.5	5– 7	101.0	5– 8	39.7	5– 9	145.5	5–10	98.2
6– 7	178.2	6– 8	58.1	6– 9	121.2	6–10	120.8	7– 8	175.2
7– 9	99.4	7–10	158.8	8– 9	89.2	8–10	80.4	9–10	33.4

Table 9: *Offered frame relaying traffics.*

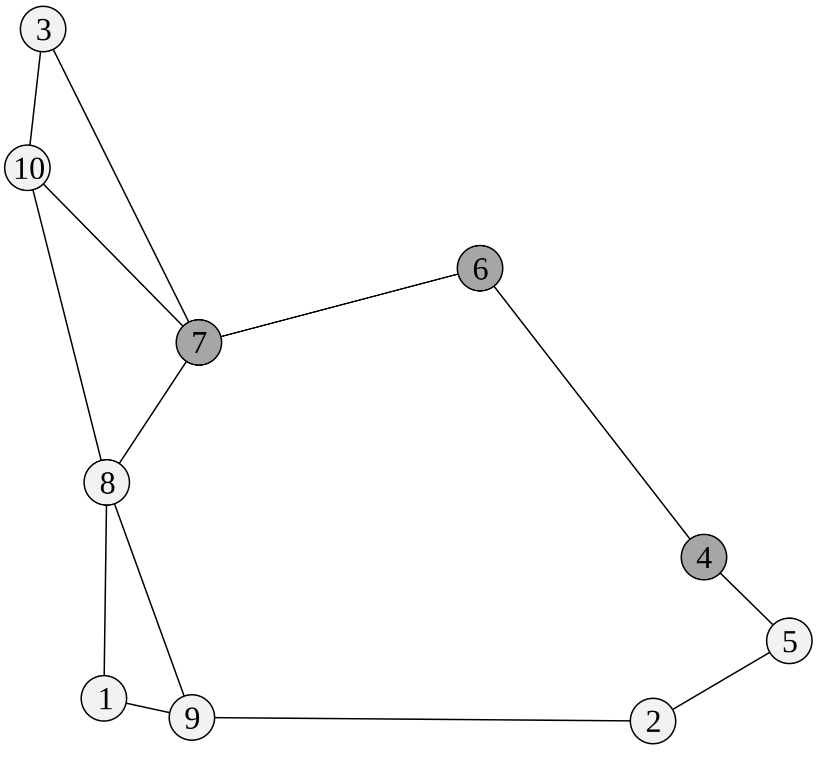

Figure 4: *Network layout.*

```
RESULT FOR NETWORK 1 USING AUTOMATIC GREEDY ALGORITHM:


*** Voice traffic ***

Total offered traffic = 5734.7
Total carried traffic = 5734.3

Mean loss    Sdev loss     Max loss     Min loss
  0.00008      0.00022      0.00139      0.00000

Mean path    Sdev path     Max path     Min path
     1.02         0.15            2            1   (Used)
     1.20         0.40            2            1   (Cons)

Mean lgth    Sdev lgth     Max lgth     Min lgth
     2.24         0.98            4            1   (Used)
     2.59         1.27            6            1   (Cons)

Reachability: Full.


*** Frame relaying traffic ***

Total offered traffic = 4944.9
Total carried traffic = 4944.6

Mean loss    Sdev loss     Max loss     Min loss
  0.00007      0.00021      0.00127      0.00000

Mean path    Sdev path     Max path     Min path
     1.02         0.15            2            1   (Used)
     1.20         0.40            2            1   (Cons)

Mean lgth    Sdev lgth     Max lgth     Min lgth
     2.24         0.98            4            1   (Used)
     2.59         1.27            6            1   (Cons)

Reachability: Full.


*** Link utilisation ***

Total provided network capacity =   793
Total assigned network capacity =   760

Mean util    Sdev util     Max util     Min util
  0.93643      0.06386      1.00000      0.76471
```

```
RESULT FOR NETWORK 1 USING AUTOMATIC GREEDY ALGORITHM:


*** Voice traffic ***

Total offered traffic = 5734.7
Total carried traffic = 5733.4

Mean loss    Sdev loss     Max loss     Min loss
  0.00017      0.00015      0.00046      0.00000

Mean path    Sdev path     Max path     Min path
     1.04         0.21            2            1   (Used)
     1.04         0.21            2            1   (Cons)

Mean lgth    Sdev lgth     Max lgth     Min lgth
     2.28         1.00            4            1   (Used)
     2.28         1.00            4            1   (Cons)

Reachability: Full.


*** Frame relaying traffic ***

Total offered traffic = 4944.9
Total carried traffic = 4944.6

Mean loss    Sdev loss     Max loss     Min loss
  0.00005      0.00011      0.00049      0.00000

Mean path    Sdev path     Max path     Min path
     1.02         0.15            2            1   (Used)
     1.04         0.21            2            1   (Cons)

Mean lgth    Sdev lgth     Max lgth     Min lgth
     2.24         0.98            4            1   (Used)
     2.28         1.00            4            1   (Cons)

Reachability: Full.


*** Link utilisation ***

Total provided network capacity =  2379
Total assigned network capacity =  2130

Mean util    Sdev util     Max util     Min util
  0.87387      0.05935      0.95833      0.73529
```