Traffic Management for IBC networks

by David Griffin (GPT Ltd) and Pravin Patel (Dowty Communications Ltd)

1 Introduction

This paper concentrates on the traffic management functions for a multi-service IBC network implemented using ATM technology. The main requirement of a traffic management system is to optimise the use of the network so as to ensure that the network resources are used as efficiently as possible whilst allowing the customers (users) to enjoy the quality of service (QoS) they have been guaranteed. In effect it should ensure fairness and efficient use of resources.

Because the IBC network is intended to transport many different services it must be able to cope with a multitude of connection types with differing requirements on connection establishment bandwidth attributes, Quality of Service and user behaviour. Many of these parameters are unknown today and will be unknown or changing during the operation of such network. In addition ATM technology provides a very flexible way of containing connections onto routes formed by virtual paths. This means that control functions of the network can not be fixed and be efficient for all time and so new traffic management functions are required to adapt the operating parameters of the network to changing requirements.

The functions of traffic management will be to manage control function strategies. These strategies must be very simple since they will be very close to the hardware which will execute traffic management actions based on these strategies.

The main function of these strategies is to attempt to do proactive / preventive management rather than reactive management. However reactive management does have a role in traffic management especially under network element failure conditions.

These management functions could be provided by manual / established means, but due to the amount of information and the complexity of the management task. It is foreseen that these functions will need to be implemented by using advanced tools to assist human operators or by closing the loop completely by using expert systems, machine learning techniques, etc... to remove the need for the majority of human intervention.

This paper shows how management techniques can be used for:

- Call Acceptance when the attributes of the calls are not well known.
- Management of virtual paths and routing where user behaviour changes throughout the day and over time (eg per year).

The traffic management strategies will attempt to influence the following parameters:

- The strategy employed to accept calls in the network at the nodes. The algorithm used at the call acceptance function can be altered by the traffic management.

- The bandwidth allocated to a connection in the network. Some algorithm will be used to allocate the bandwidth to virtual paths and to the connections on the virtual paths during the call establishment phase. Again these algorithms can be managed from the traffic manager.
- Routing tables and the strategies of using the routing tables can be managed by the traffic manager. This is particularly useful for congestion control and load balancing as well as re-routing during network element failures.

2 Background

2.1 Integrated broadband communications

Integrated broadband communications describes a network that is capable of supporting many different services with different requirements. The technology chosen within RACE and CCITT for implementing such a network is ATM.[10]

2.2 ATM

ATM cells consist of a 5 octet header and a 48 octet information field. Cells are identified on a link by their VCI (virtual channel identifier) and VPI (virtual path identifier) which are part of the header. Cells belonging to a particular connection are allocated to a virtual channel and groups of virtual channels are allocated to a virtual path which in turn are grouped onto a link (Figure 1).[2][10]

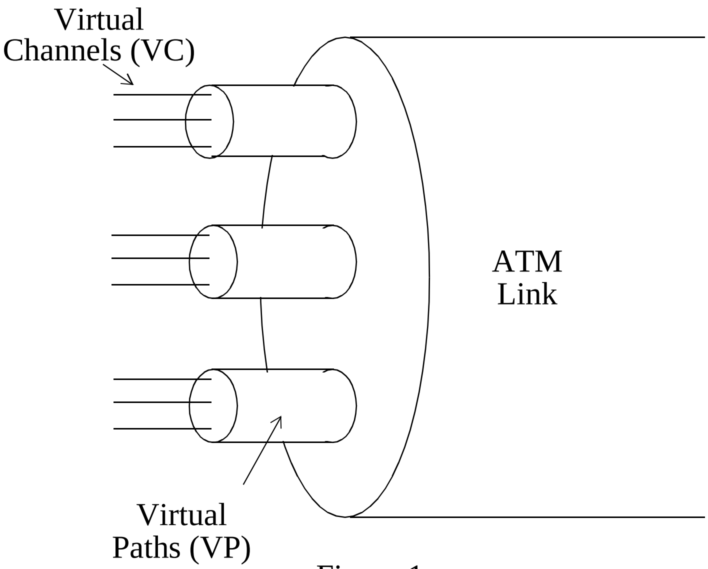
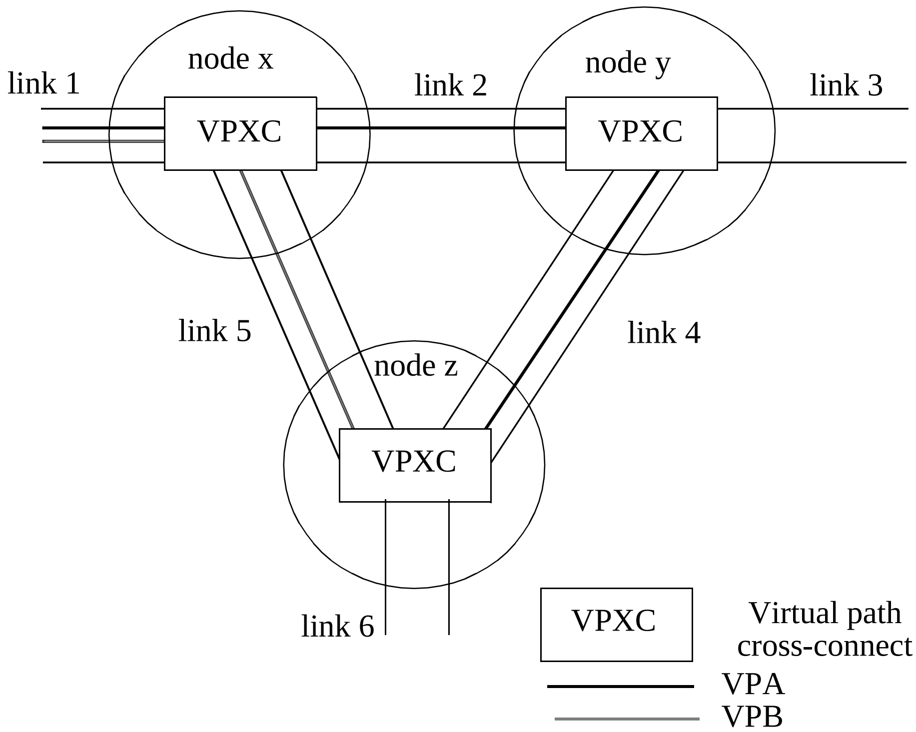


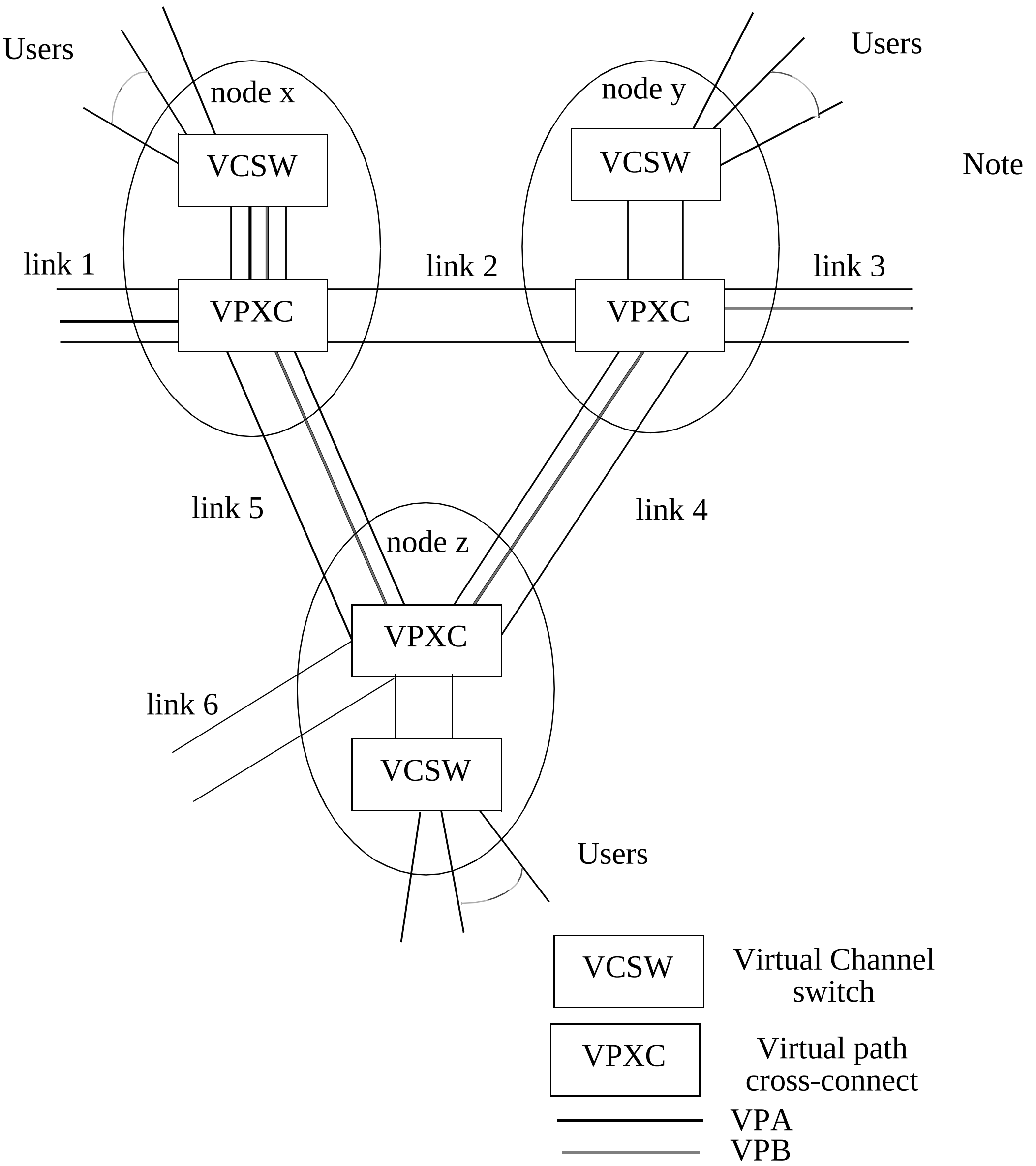
Figure 1

Virtual paths will usually span more than one link and cells belonging to a particular virtual path will be routed between links at virtual path cross-connect equipment (Figure 2). Cells belonging to particular virtual channels can be switched between virtual paths at virtual channel switches.



VPs routed via cross-connects
Figure 2

Figure 3 shows VP A terminating at node X and VP B starting at node X and routed over links 5, 4 and 3. Cells required to traverse VPs A and B will be switched at the VC switch at node X.



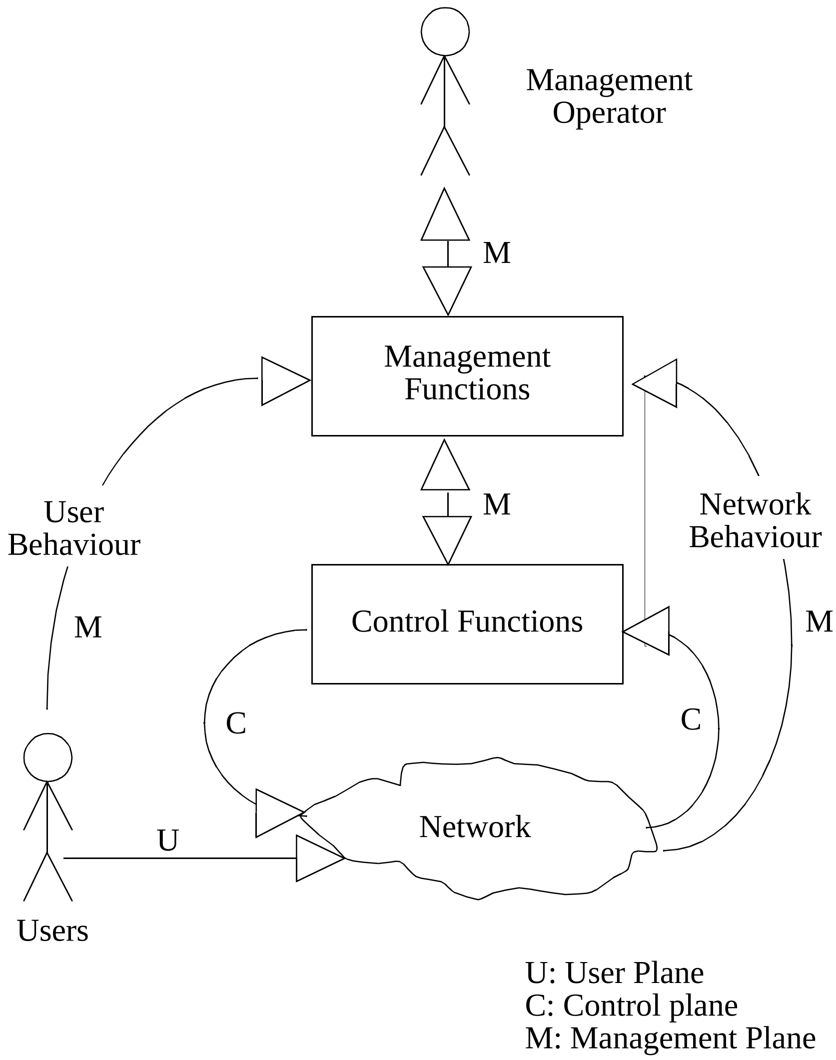
VCs Switched between VPA and VPB at node x.
Figure 3

that although there are two distinct routing/switching functions it may be the case that, in a particular implementation, virtual path cross-connects and virtual channel switches are physically co-located in a generic ATM switch.

The VCI remains constant within a VP and the VPI remains constant only over a single link, neither have end-to-end significance. It is the responsibility of the routing function in the network to alter the VPI and VCI fields in the cell header according to routing table entries that were created at connection set-up time.

2.3 User, control and management planes

Network functions can be generally split into those contained in the user, control and management planes.[8][9]. The functions associated with the control plane are those that are necessary for the normal operation of the network, whereas those contained in the management plane are concerned with the efficient use of the network resources and are not essential for the functioning of the network. This paper discusses the management plane and its interactions with the control plane for the purposes of traffic management (Figure 4).



The User, Control and Management plane views
Figure 4

3 Traffic management

Traffic management functions should not be involved in on-line connection set-up decisions these are delegated down to the control plane of the network as relatively simple local decisions that can operate quickly. Traffic management does however determine the parameters within which the call acceptance function operates. It continually monitors the performance of the network and modifies the behaviour of connection acceptance so that the network is efficiently used and so that users do not perceive degraded quality of service.

3.1 Call control at connection acceptance

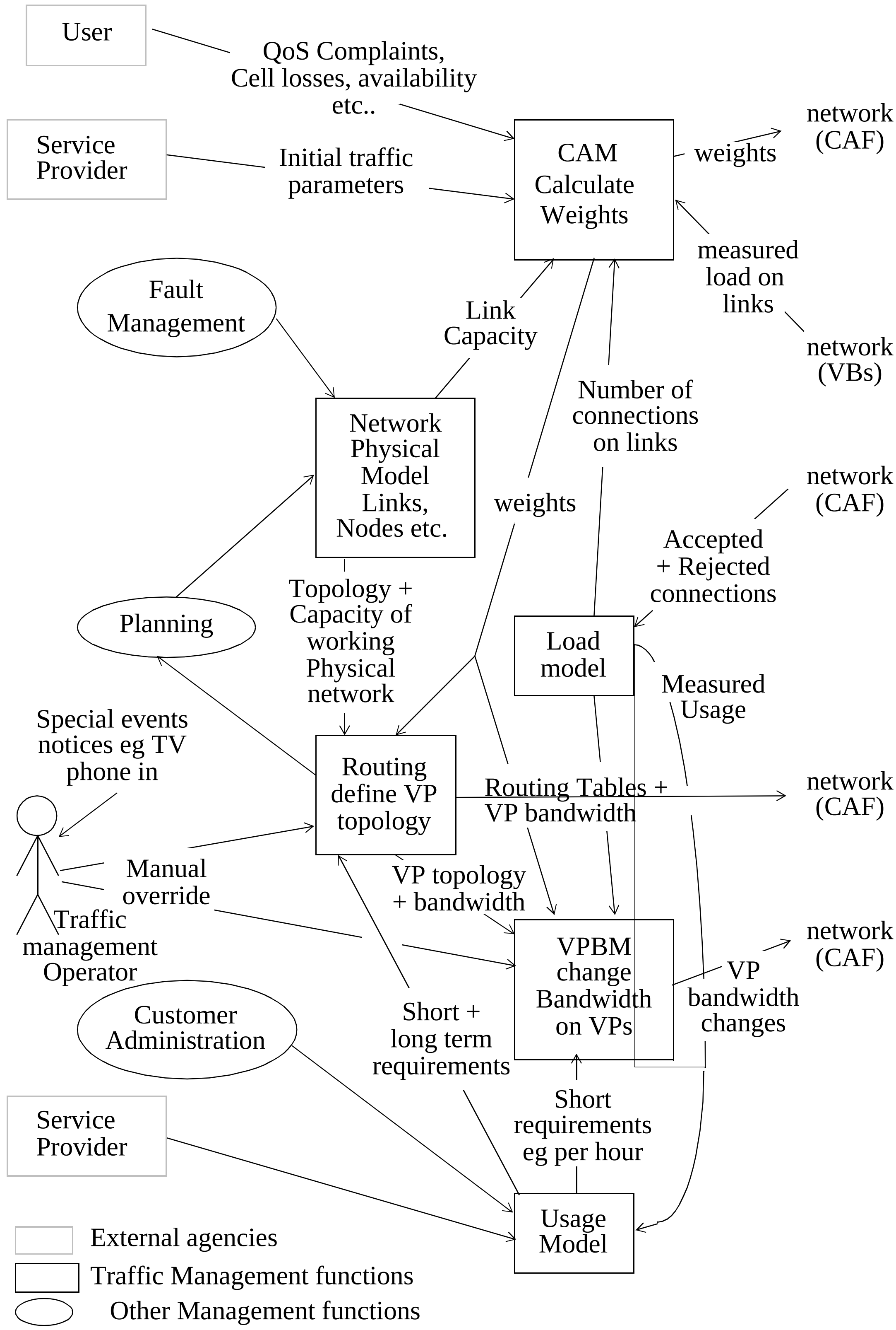
During normal operation, connection requests will be made to the network identifying the connection type and its destination. At the access switch, the routing function will search its routing table for entries satisfying the destination for the requested connection type. This search will usually result in a list of suitable VPs which will be ordered according to a priority field. The VPs will be investigated in order to find the VP with the highest priority that has capacity to accept the connection. If one is found the local database is updated with the new connection information and the connection request will be forwarded to the next switching point where the process is repeated until the destination switch is reached. If at any point along the route a VP cannot be found with adequate spare capacity the connection will be rejected by that switch and the previous switch will be notified so that it can try an alternative route with a lesser priority. If no routes can be found the connection will be rejected. The fact that connections have been accepted or rejected needs to be identified to the management system for use in its decision making activities.

3.2 The role of traffic management in connection acceptance

During the connection acceptance phase it is necessary to identify the spare capacity on VPs by knowing the bandwidth allocated to that VP and its current load. It is further necessary to identify whether the spare capacity is adequate for the requested connection type to ensure that VPs are not loaded above their allocated capacity so that in turn the links are not overloaded to the extent that excessive cell losses are experienced.

Virtual path bandwidth management is responsible for creating and modifying the routing tables and the bandwidth allocated to virtual paths. Call acceptance management is responsible for identifying the bandwidth required by connection types and for identifying the bandwidth of multiplexed connections i.e. the level of statistical multiplexing that can be achieved without deteriorating QoS. The CAM will forward this information to the CAF in the form of parameters that can be manipulated to identify whether the new configuration of connection types on a VP will exceed its capacity.

VPBM and CAM are the heart of the traffic management system (Figure 5).

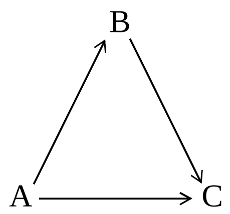


VPBM: Virtual path bandwidth Management
CAM: Call Acceptance Management
VB: Virtual Buffer
CAF: Call Acceptance function

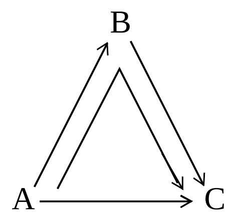
Functional diagram of traffic Management
Figure 5

4 Virtual path bandwidth management (VPBM)

ATM is based on the two level hierarchy of virtual channels and virtual paths. Because virtual paths are not directly fixed to physical resources it is possible to create a flexible routing infrastructure by virtue of the cross-connects.[1][2][4][5]



(i) Physical Links



(ii) Virtual Paths

Physical Links and Virtual Paths
Figure 6

Consider the simple example in figure 6. Three nodes are connected by three links of equal capacity. During the day-time there is equal traffic A->B, B->C and A->C, but in the evening there is a lot of traffic A->C (more than the capacity of the link will support) and a low level of traffic B->C and A->B. By defining the virtual path topology as shown in the figure it is possible to assign the full link bandwidth to VPs AB, AC and BC and zero bandwidth to ABC during the day-time and introduce VP ABC in the evening by increasing its bandwidth from zero and by reducing the bandwidth assigned to VPs AB and BC.

The task of VPBM and routing is to design a virtual path network to cater for variable traffic demands. As an ongoing task it has to continually modify VP bandwidth to follow the changes in traffic patterns. To do this task the VPBM and routing functions need to take into account the underlying network and the predicted usage of the network.

4.1 Physical constraints

The ATM network will have physical constraints, this paper and the NEMESYS work makes the assumption that the switches and cross connects are non-blocking (ref. to non blocking switches) and that the capacity of the network is entirely dependent on the capacity of the ATM links. Therefore the physical constraints are those related to the links including the physical locations they route between and their bandwidth capacity. Note that the use of the term link in this context refers to the capability of a link input buffer with a defined service rate and a fixed length, i.e. cells are discarded when the buffer is full.

The fault management function contributes to the definitions of the physical network by informing the traffic management function of any network resources that are out of service.

4.2 Usage model

The other major input required for the definition of VPs and routing information is an estimate of the traffic the network has to transport in the form of a usage model. Obviously this data is difficult to predict, but a combination of information from the service provider and from customer administration will provide a first approximation. Historical trend information after the network has been operational for some time will improve the accuracy of the model.

A usage model is required which details the number of connections of each type required between all source-destination pairs. This model should be capable of showing the temporal variations in usage, e.g. variations over the working day, week, year.

In order that this model can be translated into bandwidth requirements some knowledge is required of the bandwidth characteristics of each connection type, and how these characteristics sum for the purpose of statistical multiplexing. This information will be provided by the call acceptance manager and will be continually modified during the operation of the network.

4.3 Routing function

It is the task of the routing function to identify how to satisfy the usage requirements given the constraints imposed by the physical resources. Virtual paths will be defined so as to provide a set of routes with the required total bandwidth between source-destination pairs.

Part of this process will be to identify all the possible routes between sources and destinations and to choose a certain subset of these.

There will be many considerations in choosing this subset including:

- The number of logically and physically diverse routes required to ensure the availability of the network in fault and damage conditions.
- Some services may be more sensitive to delay than others and so the connection type supporting those services should contain a more stringent requirements on delay. The usage model will indicate the relative proportions of different connection types and their geographical distribution. Therefore some routes may be unsuitable for some connection types because they are routed through too many switches and link buffers.
- Similarly some connection types may have different requirements on cell loss rates. If there are sufficient numbers of connections with different requirements on cell loss probability it is possible to route these connections differently. Links could be labelled as high loss rate or low loss rate for example.
- Considerations such as load spreading should be taken into account. This concept involves ensuring that traffic is spread as evenly as possible over the network to minimise disruption due to fault conditions and to maximise availability over the whole network. In essence it attempts to remove local differences in load.

The output of this routing process is a set of routing tables which is down-loaded to the call control functions located at the switches and cross connects in the network. The tables at the switches contain entries per connection type for each destination identifying the virtual paths that the connection could be routed over next. A priority field is included in the tables to dictate the order in which VPs should be attempted during the connection set up phase. The routing tables at the cross-connects map VPIs on incoming links to VPIs on outgoing links to identify part of the overall topology of a particular VP.

4.4 VPBM functions

The usage model identifies how traffic patterns change over time and will reflect for example the differences between business and domestic traffic over a working day. This model is an important input to the VPBM who will attempt to modify the bandwidth allocated to VPs in the network to accommodate the predicted changes before and as they happen.

The usage model periodically sends usage predictions to the VPBM so that the VPBM can modify VP bandwidth to accommodate the predicted traffic in the next time period. The frequency of these transactions will depend on the rate of change in user behaviour. Depending on the amount by which usage patterns change it may be necessary to remove some VPs from use or introduce new VPs and occasionally it may be necessary to completely rethink the topology of VPs. Sometimes there will be significant changes in traffic patterns due to equipment failures or special events such as a TV phone-in and these special events will be catered for by a human operator entering the detail of the foreseen conditions into the VPBM or routing components.

As an ongoing process the VPBM needs to track the actual usage of VPs to identify whether it needs to modify their allocated bandwidth. It can do this by getting the number of connections of each type on a VP from the load model. By using the parameters from the CAM it will translate the list of connection types into bandwidth and compare this figure to the bandwidth the VP has allocated to it. When discrepancies are noticed it will tune the bandwidth allocated to VPs to the measured usage by increasing or decreasing the allocated bandwidth but ensuring that there is enough spare bandwidth to allow additional connections to be set up. At the same time the usage model should be modified to reflect the differences between actual and predicted traffic patterns so that as time progresses the usage model will become more and more accurate.

As well as tracking the utilisation of individual VPs the VPBM will take a global view of the distribution of load over the whole network. It will attempt to eliminate local deviations from the network-wide utilisation figure by modifying the VP network and the routing parameters. If the VPBM finds that a particular portion of the network is continually over-utilised and that it is difficult to remove this condition due to the limitations of the underlying physical resources it will forward this information to the network planning functions so that additional resources can be introduced to meet the demand in this area. Conversely the VPBM may find that portions of the network are always under-utilised and similarly the network planning function may decide that some of the resources could be re-deployed elsewhere.

5 Call Acceptance Management (CAM)

Call acceptance management is responsible for determining the strategy for accepting/ clearing down calls on the network. However the calls to the network are accepted/ cleared by the control part of the network. The strategy for the control part is determined by the CAM.[4][5] Control of the calls is required to:

1. Ensure that the network is not overloaded with calls and hence cause congestion in the network.
2. Ensure that network resources are not under-utilised by excessive control of calls accepted onto the network.
3. Ensure that once the calls are accepted to set up the threshold of the resources that the call can use, essentially source policing, e.g. bandwidth requested, delay throughput, connection path and QoS parameters such as sensitivity to cell losses etc.
4. Ensure that the connections of the call and the resources are released in an orderly manner when the call is cleared down. The timeliness of cleardown and the release of the resources may be different for different types of calls. The strategy being determined by CAM. eg if a customer is using traditional "leased line" type services on the ATM network one would not want to release the resources used by the call.

Parameters which the CAM can alter/set-up/determine are:

1. The types of calls that are accepted on the link. The types of call being characterised by:
   - i) Bandwidth requirements of the connections.
   - ii) QoS parameters required for the calls such as:
     - delay throughput
     - delay variation (jitter)
     - data loss sensitivity
     - data error
     - data insertion
     - premature termination
   - iii) Contractual terms such as
     - semi-permanent connections
     - permanent connections
     - transient connections
     - source policing parameters
2. The number of connections accepted on the links.
3. The algorithms used in accepting the calls.

5.1 Role of CAF

The Call Acceptance Function (CAF) is part of the control plane and is located in the ATM switches because it needs to operate quickly during the connection set-up phase. The operation of this function is described above in section 3.1. It is the job of CAM to determine the parameters by which the CAF and VPBM determine the bandwidth of connection types and the way they multiplex together.

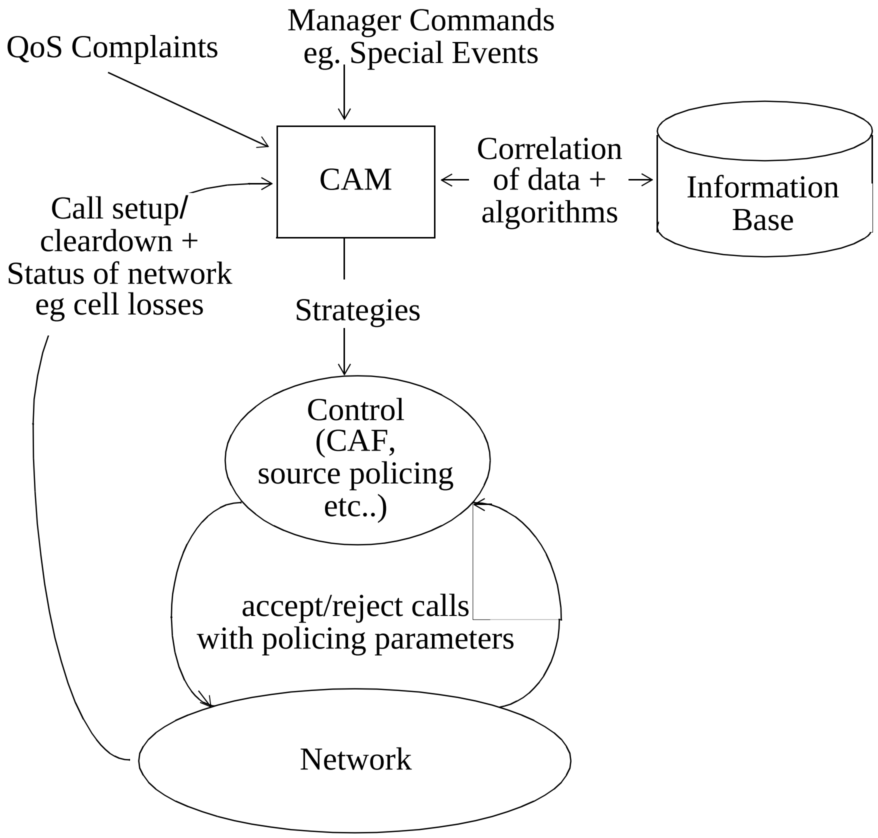
The constraints on accepting the next new call are based on:

1. The current traffic on the link and network.
2. The resource requirements/characteristics of the call, e.g. routing requirements, bandwidth requirements etc.

What is required in the CAF is a relatively simple way of deciding whether a new connection of the specified type can be accommodated on a particular VP given the bandwidth allocated to it and its current load. If Variable Bit Rate (VBR) connections are accepted according to their mean bandwidth a link can be 100% utilised but there will be significant cell losses, on the other hand, if connections are accepted according to their peak bandwidth there will never be any cell losses but the links will be under-utilised for much, if not all, of the time. The goal is to accept connections somewhere between these two extremes so that utilisation is improved from the acceptance-on-peak situation but so that the cell loss probability is less than or equal to the acceptable limit for the connection types on that link. In this paper a technique known as the linear weights method is described, but other techniques are being investigated in the Nemesys experimental work. [5] viz:

1. The linear weights method: In this method the resource requirements of the new calls are added to the current utilisation of the links. If that figure exceeds a threshold then the call would be rejected.
2. The neural network(NN) approach: In this method a NN is trained by the CAM so that it can find out what the best mix of traffic that can be carried by a link. Once the NN has learned the mix types, the learned NN can be sent to the control function. Whenever a new call is tested to see if it is acceptable the NN will come up with the solution of acceptance or rejection. The big advantage of this method is that it is very flexible and the NN can be trained for different situations, e.g. cup-final day etc. Also once the NN has been trained it executes very fast. The big disadvantage is that while training the NN, the users on the network will experience QoS problems.
3. Rule based/AI system: If the call satisfies some rules then the call would be accepted. The rules being developed by experts and downloaded by the CAM to the CAF.

Figure 7 shows graphically the relationship between the CAM, CAF and the network. It shows that the events from the network and the users are processed by the CAM using an information base to store/retrieve data. It then computes strategies and algorithms which are used by the control to accept/reject calls on the network.



Relationship between CAM, CAF and Network
Figure 7

5.2 Linear weights method

Calls are split into their constituent connections and each connection is classified according to its resources requirements. These classifications are called connection types.

Each connection type is assigned a weight between 0 and 1 which describes the proportion of total link bandwidth consumed by that connection type, for a constant bit rate (CBR) connection types this will be representative of the mean bit rate but for variable bit rate (VBR) connection types the weight will represent a bandwidth somewhere between the mean and peak bit rate. Each VP is allocated a certain proportion of link bandwidth and will be assigned a weight-total to reflect this. The CAF simply has to sum the weights of the existing connections on a VP together with the weight of the proposed connection. If the sum is less than the weight-total of the VP the connection can be accepted.

Note: This technique is known to be sub-optimal because it assumes that the weights can be summed linearly. This is not the case for optimal statistical multiplexing. It is described here because of its simplicity and the small number of parameters needed to transfer information from CAM to CAF and CAM to VPBM.

5.2.1 CAM - CAF relationship

It is the task of CAM to determine what the weights ought to be to achieve the highest level of statistical multiplexing on VPs, and hence links, to achieve the QoS targets (this can be roughly translated to mean the number of cell losses). The weights are downloaded to the CAF and forwarded to the VPBM.

5.2.2 Determining the weights

When the network is first created or whenever a new connection type is introduced, the service provider who requested the connection type will be able to give some information regarding its traffic characteristics. This information is used to create the first approximation for the weight for that connection type. As time goes by the CAM will learn more about the connection type and how it multiplexes with other traffic and will modify the weight to reflect this increased knowledge.

There will be several indicators that the weights need to be modified including:

- An unacceptable number of cell losses measured at a particular link buffer.
- The CAF is rejecting connections when it is known that the network is under-utilised.
- QoS complaints from users, eg. excessive cell losses or availability is too low.

When these situations are brought to the attention of the CAM, it needs 7zto identify the offending combinations of connections on which links.

This is basically done by comparing the actual load with the theoretical load, i.e. the load calculated by summing the weights of the individual connection types. This is sensibly measured on a link basis. The actual load on a link can be estimated by the use of one or more variable service rate virtual buffers attached to the link (ref. NEMESYS work or elsewhere). The theoretical load can be calculated by interrogating the load model which can identify active connections on VPs and relate this to the VPs associated with a particular link. If there are discrepancies between these values it can be assumed that one or more of the weights are inaccurate. If the measured load is greater than the theoretical load one or more of the weights need to be increased and if the measured load is less than the theoretical load one or more of the weights need to be decreased.

By looking at the combinations of connection types together with the measured and theoretical loads for a number of links it will be possible to get some clues as to which of the weights need to be modified.

6 Experimental work

The Project Nemesys is developing a series of prototypes to test some of the ideas described in this paper. The first prototype developed the experimental framework, the second prototype developed the simulators and the framework for the management and the third and final prototype is developing the management ideas described in this paper.[5] The following components have been developed:

1. Network simulator
2. Service simulator
3. User simulator
4. Experiment infrastructure

5. Traffic Manager
6. Service Manager

The CAM and the VPBM are part of the traffic manager and are responsible for the implementation of the strategies for the control modules which are developed in the Network simulator.

7 References

1. K Mase and S Shioda, "Real-Time Network Management for ATM Networks", in Teletraffic and Datatraffic in a period of change, Proc. of 13th International Teletraffic Congress, Copenhagen, 1991.
2. KI Sato, S Ohta and I Tokizawa, "Broadband ATM Network Architecture based on Virtual Paths", in IEEE trans. on Communications, vol. 38, no. 8, Aug. 1990.
3. C Rasmussen and K Kvols, "Flexibility and Optimisation in Broadband ISDNs", in Teletraffic and Datatraffic in a period of change, Proc. of 13th International Teletraffic Congress, Copenhagen, 1991.
4. NEMESYS consortium, "Specification of the NIKESH Experimental System", RACE deliverable 05/GPT/BNM/DS/X/009/b1, May 1990.
5. NEMESYS consortium, "NIKESH - experiment 2 in NEMESYS", RACE deliverable 05/KT/AS/DS/B/017/b1, June 1991
6. P Lebouc and PE Stern, "Distributed Problem Solving in Broadband Telecommunications Network Management", in Proc. AVIGNON’91 Conf. on Expert Systems and Applications, Avignon, May 1991.
7. W Kleinoder, P Lebouc, PE Stern and M Tielemann, "Constraint Logic Programming for a Virtual Path Bandwidth Manager", in Proc. Fourth RACE TMN Conf., Dublin, Nov. 1990.
8. CCITT Rec. I.320, "ISDN Protocol Reference Model", in Blue book, Vol. III, Fascicle III.8, Geneva 1989.
9. CCITT draft Rec. I.321, "B-ISDN Protocol Reference Model and its Application", Study Group XVIII, Geneva 1990.
10. CCITT draft Rec. I.150, "B-ISDN ATM Functional Characteristics", Study Group XVIII, Geneva 1990.

8 Acronyms

IBC:	Integrated Broadband Communications
ATM:	Asynchronous Transfer Mode
VP:	Virtual Path
CAM:	Call Acceptance Management.
VPBM:	Virtual Path Bandwidth Management
AI:	Artificial Intelligence

AIP:	Advanced Information Processing
ATM:	Asynchronous Transfer Mode
CAF:	Call Acceptance Function
CBR:	Constant Bit Rate
B-ISDN:	Broadband Integrated Services Digital Network
NN:	Neural Network.
QoS:	Quality of Service
TMN:	Telecommunications Management Network
VB:	Virtual Buffer
VBR:	Variable Bit Rate
VC:	Virtual Channel
VCI:	Virtual Channel Identifier
VP:	Virtual Path
VPBM:	Virtual Path Bandwidth Manager
VPI:	Virtual Path Identifier

9 Acknowledgements

The work described in this paper was supported by the Commission of European Community under the RACE programme in the project NEMESYS R1005.