INTEGRATED COMMUNICATIONS MANAGEMENT OF BROADBAND NETWORKS

Crete University Press, Heraklio, Greece
ISBN 960 524 006 8

Edited by David Griffin

First published in 1996 by
CRETE UNIVERSITY PRESS
Foundation for Research and Technology
P.O. Box 1527, Heraklio, Crete, Greece 711 10
Tel: +30 81 394235, Fax: +30 81 394236
email: pek@iesl.forth.gr

The ICM consortium consists of the following companies:

Alcatel ISR, France
Alpha SAI, Greece
Ascom Monetel, France
Ascom Tech, Switzerland
Centro de Estudos de Telecommunicações, Portugal
Cray Communications Ltd., United Kingdom (Prime contractor)
Danish Electronics, Light & Acoustics, Denmark
De Nouvelles Architectures pour les Communications, France
Foundation for Research and Technology - Hellas, Institute of Computer Science, Greece
GN Nettest AS, Denmark
National Technical University of Athens, Greece
Nokia Corporation, Finland
Queen Mary and Westfield College, United Kingdom
Unipro Ltd., United Kingdom
University College London, United Kingdom
University of Durham, United Kingdom
VTT - Technical Research Centre of Finland

Chapter 5

VPC and Routing management

Editor: David Griffin
Authors: David Griffin, Panos Georgatsos, Andy Carr, Dennis Nyong, Bruno Rossi, Pravin Patel, Peter Baxendale, Michel Besson

This chapter presents a VPC (Virtual Path Connection) and Routing Management Service for multi-class ATM networks. The objective of the VPC and Routing Management Service is to guarantee network availability whilst guaranteeing that the network meets the performance requirements of the different service classes the network supports. This Management Service is beneficial to the network operator since it ensures that the network resources are used as efficiently as possible in conditions of dynamically changing traffic patterns.

The proposed system offers the generic functions of performance monitoring, load monitoring and configuration management in ATM networks. In addition, it provides specific functions for routing and bandwidth management in a hierarchical structure. The components in the hierarchy differ in terms of the level of abstraction, complexity and timescale over which they operate.

The main contribution of this chapter is the answer to the question: what are the components required for the VPC and Routing Management Service and how do they interoperate?

5.1 Introduction

There is significant activity in the research and development of new services to be deployed on broadband networks. In particular multi-media, entertainment, mobile, cooperative working, tele-working and VPN services are currently being deployed. The new services must coexist with the traditional voice and data traffic and Asynchronous Transfer Mode (ATM) [5.27] is an ideal technology for supporting this integration. But careful planning and management must be involved to ensure that the many different services can exist on the same infrastructure as efficiently as possible without compromising their very different performance and bandwidth requirements.

The efficient operation of the network depends on a number of design parameters, one of the most influential being routing. Routing in ATM is based on Virtual Path Connections (VPCs), a route is defined as a concatenation of VPCs. It has been widely accepted that VPCs offer valuable features that enable the construction of economical and efficient ATM networks, the most important being management flexibility. Because VPCs are defined by configurable parameters, these parameters and subsequently the routes based on them can be configured and re-configured on-line by a management system according to network conditions.

The overall objective of a routing policy is to increase the network throughput in terms of call admissions, while guaranteeing the performance of the network within specified levels. The design of an efficient routing policy is of enormous complexity, since it depends on a number of variable and sometimes uncertain parameters. This complexity is increased by the diversity of bandwidth and performance requirements of different connection types in a multi-class network environment. Furthermore, the routing policy should be adaptive to cater for changes in the network: topological changes due to faults or equipment being taken in and out of service; and changing traffic patterns.

As traffic patterns change over time, network performance may deteriorate when the bandwidth allocated to VPCs, and therefore to routes, is not in accordance with the quantity of traffic that is required to be routed over them. To combat this, the VPC topology, the routes, and the bandwidth allocated to VPCs must be dynamically re-configured to meet the demands of the traffic. A VPC and Routing management system is therefore required to take advantage of the features of VPCs while ensuring that the performance of the network is as high as possible during conditions of changing traffic.

To exemplify the above, consider the simple example of Figure 5.1. Six nodes are connected by six unidirectional links of equal capacity. During the day-time there is equal traffic A→C, C→E and A→E, but in the evening there is an increase in traffic A→E (more than can be supported by the capacity of the links) and a reduction in traffic A→C and C→E. By defining the VPC topology as shown in the figure it is possible to assign the full link bandwidth to VPCs AC, AE and CE and zero bandwidth to ACE during the day-time. In the evening, when the traffic pattern changes, the bandwidth allocated to VPC ACE is increased from zero and the bandwidth allocated to VPCs AC and CE is reduced. Now traffic A→E can be routed over either of the direct VPCs AE or ACE.

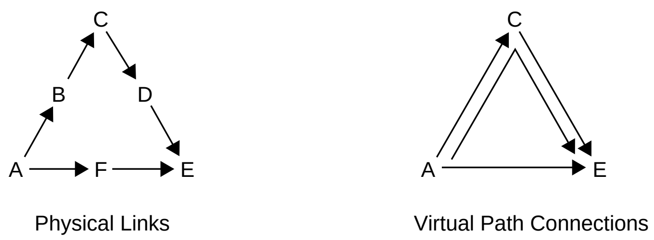

Figure 5.1 Physical links and VPCs

Note that in this example the re-configuration of the network is achieved simply by altering the parameters in the CAC (Connection Admission Control) which define the bandwidth allocated to the VPCs and the routing tables do not have to be modified.

Although there is a significant research interest in the area of performance management on ATM, particularly in routing [5.1] [5.2] [5.3] [5.4] [5.5] [5.9] [5.19] [5.21] [5.22], bandwidth assignment [5.7] [5.8] [5.13] [5.16] [5.17] [5.18] [5.20] and VPC management [5.6] [5.10] [5.12], the problem of VPC and route management remains largely open. The majority of management systems deployed today are concerned with network configuration and network monitoring and the management intelligence is provided by the human users of the management systems. There is a trend [5.11] [5.14] [5.15] [5.24] to increase the intelligence of the management functions to encapsulate human management intelligence in decision making TMN components to move towards the automation of the monitoring, decision making and configuration management loop.

Within the above framework, this chapter concentrates on VPC and routing management for ATM based B-ISDN networks, and proposes a management system for implementation using TMN and OSI systems management principles.

Considering the requirements the issues are analysed in order to decompose the system into a number of distinct but cooperating functional components which are then mapped to the TMN architecture. The identified components are described and their operational dependencies and information exchange are analysed in the context of the overall system operation. The decomposition and subsequent synthesis leads to a hierarchical system of physical TMN components (OSs) which operate at different levels of abstraction and at different time intervals. An extended version of the ITU M.3020 interface specification methodology (see Chapter 3) has been followed covering the phases of Management Service specifications and decomposition to service and functional components.

The emphasis of the work is on the development of a suitable architecture for a complex management system implementing the VPC and Routing Management Service, rather than on the detailed development and experimentation of management algorithms, e.g. load balancing, VPC topology design. However, in order for the architecture to be validated, it was necessary to enhance existing, or develop new, management algorithms in the context of the architectural components and to experiment with the resulting system.

The interactions required between the control and management planes of ATM networks have been studied, thereby specifying the boundaries of responsibility of the VPC and Routing Management Service. The overall management system interacts with the underlying network mainly for the purpose of configuring control plane operational parameters. The management system compliments rather than replaces or duplicates control plane functionality, having a less stringent requirement on real-time response and computational overhead. Specifically the following interactions have been identified: monitoring traffic conditions; configuring VPC topologies; managing the CAC algorithms in terms of maximum cell loss rates and the bandwidth allocated to VPCs; managing the route selection algorithms in the network by configuring the routes available to connection types between source-destination pairs; and dynamically managing route selection priorities.

The remainder of this chapter is organised as follows: Section 5.2 discusses the specific environmental assumptions made by the designers of the VPC and Routing Management Service in addition to those introduced in Chapter 2. Section 5.3 defines the Management Service and presents the issues behind its functional decomposition. In Section 5.4, the VPC and Routing Management Service is mapped to the TMN architecture and the role and functionality associated with each of the system building blocks is introduced. The next sections - Section 5.5 to Section 5.13 - address each of the system components in detail, discussing the theoretical issues behind their design, the algorithms deployed and the information models they present to the rest of the system. Finally Section 5.14 presents the conclusions and identifies future work.

5.2 Background

5.2.1 The environment

The section describes the network environment from the perspectives of the VPC and Routing Management Service.

The ATM equipment being deployed today is targeted for use in LAN environments, utilising (semi) permanent connections. Research, standardisation and development work is progressing on the interconnection of ATM switches for WAN implementation for both public and private networks supporting switched connections.

It is assumed that the network being managed is a public network offering switched, on-demand services. The network supports a full range of services[1], ranging from simple telephony and file transfers to multi-media conferences.

5.2.2 Assumptions on the network services

It is assumed that network services are composed of a number of unidirectional connections. Each connection is of a particular connection type or class. The term class of

1. Provided that calls are able to be decomposed into a number of unidirectional connections. Distributive services requiring point-to-multipoint connections, are not explicitly considered.

service (CoS) is used to denote a particular connection type. The CoSs are the bearer services provided by the network.

A large range of CoSs may be defined by the network operator according to its business policy and perceived user requirements. Users (or the applications they are using) make requests for connections of a specific CoS. For example, telephony services may be supported by a number of CoSs which offer a range of qualities - different transfer qualities or call blocking probabilities.

The CoS definition characterises the connection type in terms of bandwidth and performance requirements.

It is assumed that the bandwidth requirements can be characterised by mean and peak parameters. This is because the CAC algorithms deployed in the switches are based on these parameters. Alternative bandwidth parameters may be used according to the specific CAC algorithms deployed in the Network Elements.

It is assumed that the performance requirements are characterised by:

- cell loss probability
- cell delay
- cell delay variation
- connection blocking probability (or availability)

Other performance parameters may be included in the CoS definition, connection release delay for example, but these may not be influenced by this Management Service and are not considered further here.

An important point here is the relationship between the classes of the bearer services provided by the network and the four AAL classes recommended by the ITU-T [5.28]. The AAL provides a limited range of services, e.g. connection-oriented versus connectionless, error recovery, re-transmissions with the assumption of a given performance of the underlying bearer service.

In our view there needs to be a range of bearer services of different qualities and costs to support the AAL services. This will allow decisions to be made on whether to use a comprehensive AAL with a cheap, low performance bearer service or a lightweight AAL with a higher performance bearer service (e.g. smaller cell loss ratio). This view is in accordance with the ATM Forum [5.23] which explicitly recommends the augmentation of the AAL service classes with a range of quality of service classes. The AAL exists in the user terminals whilst the underlying bearer service is provisioned by the network operators.

This work described in this chapter concentrates on the management of the bearer services from the viewpoint of the network operator. Although AAL issues are considered from the perspective of the requirements they impose on the underlying bearer services the end-to-end management issues of layer 4 and above are not the focus of the work.

Another important point is the role of connection-oriented services with a predefined bandwidth and performance compared to that of best effort (no performance requirements), unspecified bitrate (UBR) or available bitrate (ABR) services. We recognise the requirement for all types of services but our work concentrates on the management needs of the services with a predefined bandwidth and performance. Best effort and ABR services are controlled via the signalling protocols. If they are to coexist with defined quality services on the same network, the network resources must be properly partitioned. There is scope for the management plane to interact with the control plane to dynamically allocate sufficient resources for unguaranteed connections, however this is an issue for future work.

5.2.3 Assumptions on user behaviour

User behaviour changes over time, in two ways: the type of user and the number of each type of user. There are potentially many different types of users characterised by the types of service they use and also by their usage patterns. The behaviour of individual users changes over time with respect to the services they use and the way that they use their services.

We assume that estimates of aggregate user behaviour can be made and trends can be identified in the short term (e.g. business vs. domestic traffic throughout the working day) and the medium to long term (e.g. seasonal variations, new service introduction, competition). This is supported by the law of large numbers.

5.2.4 Management and control plane interactions

Chapter 2 describes the activity of the control plane in setting up new connections, switching cells, etc. The management plane should not be involved with on-line connection set-up decisions. These are delegated down to the control plane as relatively simple local decisions. However, management plane functionality does determine the parameters, based on which, the control plane operates. It continually monitors the performance of the network and modifies the behaviour of the control plane algorithms so that the network is efficiently used and so that users do not perceive degraded Quality of Service (QoS).

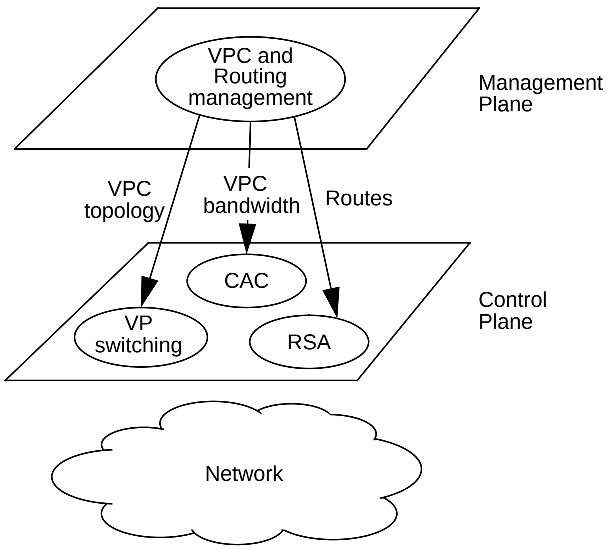

Figure 5.2 Management influence on the control plane

Before connections may be set-up, VPCs must have previously been installed. VPC definition and creation is the responsibility of the management plane. During the connection set-up phase, at each switching point, the Route Selection Algorithm (RSA) must select an outgoing VPC over which the new connection will be routed; and the CAC algorithm must determine whether the connection can be accommodated on the selected VPC. VPC selection is made from a subset of all outgoing VPCs, and it is the responsibility of the management plane to identify a subset of VPCs which are suitable for a particular CoS. The CAC algorithm determines whether there is sufficient capacity on the selected VPC based on the bandwidth allocated to it and on the portion of that bandwidth which is being consumed by existing connections. The management plane provides the CAC algorithms with the bandwidth allocation for each VPC.

5.2.5 Management requirements on network elements

A management interface is required on the network nodes (VC switches and VP cross-connects) to allow interactions with the TMN system. It should be possible to monitor the following information:

- The node at the source of a VPC should identify the used bandwidth on a VPC.
- When a connection is accepted, released or rejected, the source node of the connection should provide an indication detailing the source and destination node ids and the CoS of the accepted, released or rejected connection.
- On request, a source node should provide the number of currently active connections of the specified CoS to a specified destination node.

It should be possible to configure and subsequently modify the following:

- Bandwidth allocated to VPCs,
- Topology of VPCs,
- CAC cell loss target,
- Route selection entries,
- Route selection parameters.

5.3 The VPC and Routing Management Service

5.3.1 Objectives

The objective of the VPC and Routing Management Service is to guarantee network availability whilst guaranteeing the performance requirements of current and future connections as specified in their CoS definitions.

The Management Service achieves its goals through managing VPCs and the associated routing tables to accommodate changes in user behaviour and traffic patterns. In particular, the Management Service influences:

- the routes VPCs take through the network by configuring the VP routing tables in the VP cross-connects,
- the bandwidth allocated to VPCs by changing the parameters in the CAC part of the VC switches at VP entry points. This will be done at two levels:

    - a longer term strategy to deal with predicted traffic levels, e.g. on an hourly basis,
    - a shorter term strategy to tailor the bandwidth allocated to VPCs to follow the actual usage of the VPCs,
- the route selection tables in the Call Control (CC) part of the VC switches to alter the choice of VPCs the CC makes during connection set-up. This is particularly useful for congestion control and load balancing as well as re-routing during network element failures.

The task of the VPC and Routing Management Service is to design a VPC network and a routing plan[1] to meet predicted demand, furthermore its task is to dynamically manage the VPC network and the routing plan to cater for variations in use. It has both static and dynamic aspects. The static aspect is related to the design of a VPC network and a routing plan to meet predicted demand. In fact the static aspect is of quasi-static form in the sense that is invoked whenever the predictions change significantly. The dynamic aspect manages the VPC network and the routing plan to cater for unpredictable user behaviour within the epoch of the traffic predictions.

This Management Service belongs to the performance and configuration management functional areas and specifically covers traffic management while its static aspects are related to the network planning functions.

Figure 5.3 shows the relationship of VPC and Routing Management with the network, human managers (TMN users), other management functions, network customers and other network operators.

By following the methodology described in Chapter 3 the following section decomposes the Management Service into its constituent Management Service Components (MSC) and Management Functional Components (MFC).

5.3.2 Decomposition

Connection rejection is affected by two factors: the number of alternative routes and the available capacity on the VPCs comprising the routes. These two factors cannot be treated in isolation. The VPC and Routing management system must ensure that there are a sufficient number of routes and that their is sufficient capacity on the VPCs forming the routes, to guarantee network performance and availability.

As mentioned previously the Management Service should provide adaptivity to changing traffic conditions. There are two levels at which the traffic can change: cell level variations within the scope of a single connection; and connection level variations as users establish and release calls. The former is dealt with by the CAC and UPC functions of the control plane. Connections can never exceed the bandwidth parameters defined for a CoS due to the role of the UPC functions. If connections do not consume the full bandwidth the shortfall cannot be used by other connections because of the concept of pre-defined bandwidth reservation at connection set-up time which is paid for by the users. For this reason cell level variations are of no concern to this Management Service and the management of connection level variations is the main focus.

1. The term routing plan denotes the sets of routes and selection criteria for each source-destination pair and CoS.

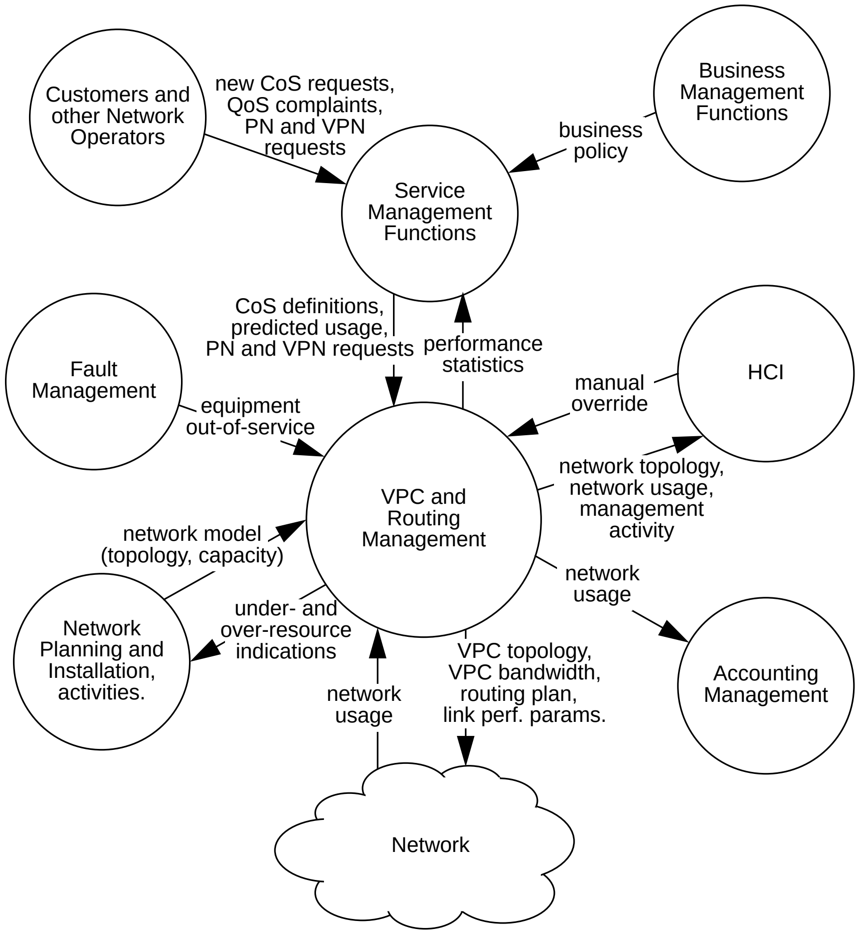

Figure 5.3 Enterprise view of VPC and Routing management

The following views of the network are useful for offering different levels of abstraction to assist the task of formulating the problem faced by the VPC and Routing Management Service.

- *The physical network* consisting of the network nodes and the transmission links.
- *The VPC network* consisting of the VC switches interconnected by VPCs.
- *The ClassRoute networks*. For each CoS, the ClassRoute network is the sub-network of the VPC network which consists only of the VPCs that belong to routes of that CoS.
- *The SDClassRoute networks*. For each CoS and a given source-destination (s-d) pair, the SDClassRoute network is the sub-network of the ClassRoute network

consisting only of the VPCs that belong to the routes interconnecting the given (s-d) pair.

In the case of alternative routing more than one path appears in a SDClassRoute network. Different ClassRoute networks may have common links since a VPC may carry more than one CoS. Different SDClassRoute networks may also have links in common since the paths between two s-d pairs (for the same or different CoSs) may have common parts.

Having introduced the above network views the goal of the VPC and Routing Management Service can be formulated as follows:

- Given the physical network and the traffic predictions per s-d and CoS, define VPC and SDClassRoute networks so that the traffic demands are met and the performance levels specified per CoS are guaranteed.

The solution requires answers to the following questions:

- How is the VPC network constructed and how frequently will it change?
- How are the ClassRoute networks constructed and how frequently will they change?
- According to what criteria will routing be achieved in the ClassRoute networks? i.e. Given the VPC and ClassRoute networks how are the route selection parameters assigned and how frequently will they change?

The definition of the VPC and ClassRoute networks is an iterative procedure. The two tasks are not independent because routes are defined in terms of VPCs and the VPCs are defined in order to support routing.

The VPC and the ClassRoute networks are constructed using, as input, estimates of the network traffic per s-d pair and CoS. The construction of these two networks is related to the network planning activities in the sense that these activities define the physical network topology and its capacity based on long term network traffic predictions.

As well as accommodating changes in the predictions the VPC and Routing management system should cater for inaccuracies in the predictions.

Whenever the traffic predictions change, the VPC and ClassRoute networks need to be reconstructed. The level of reconstruction depends on the significance of the changes. As a result, new values for VPC bandwidth may be given, or the topology of the VPC network may change (by creating and deleting VPCs) or the topology of the ClassRoute networks may change (by creating and deleting routes). Each of these reconfigurations deals with a different level of abstraction according to the network views described above. Moreover they may be performed within different time scales and they require different levels of complexity and hence computational effort. We envisage that an efficient way to deal with such reconfigurations is through a hierarchical system.

The essence of the hierarchy we propose is as follows. First the VPC bandwidth is reconfigured within the existing SDClassRoute networks. If it is not possible to accommodate the traffic predictions within the SDClassRoute networks they are reconfigured within the existing VPC network. If it is found that the VPC topology is insufficient for the predicted traffic then the VPC network is reconfigured. Ultimately it may be discovered that the physical network is unable to cope with predicted traffic and the network planning functions are informed to request that additional physical resources are deployed. Another possible result of this would be for network level management to trigger service level management which may then decide to migrate some services in order to bypass the bottlenecks in the network, which may not be circumvented otherwise. An example of this is a video on demand service relocating its video server. This is an interesting example of the network management level functions triggering service level management, rather than vice versa.

This analysis indicates the need for three management components: Bandwidth Allocation (for VPC bandwidth updates given SDClassRoute networks), Route Planning (for route updates given the VPC network) and VPC Topology (for defining VPC topologies given the physical network topology).

The above assumes that the traffic predictions are accurate, but as mentioned previously, this cannot be taken for granted. For this reason we introduce a lower level into the hierarchy which refines the initial estimates by taking into account the actual usage of the network. The lower level functionality operates within the SDClassRoute networks and redefines VPC bandwidth and route selection parameters according to measured network load. Redefinition of SDClassRoute networks and VPC topology is not done at this level since the computational load must be as light as possible. However this level will provide triggers to the higher level when it is proved that the initial estimates (from the higher level) under or over estimate the actual situation and this cannot be resolved within the existing SDClassRoute networks. Even if the predictions are accurate there is still a case for lightweight lower level functions to cater for traffic fluctuations around the predicted values within the timeframe of the predictions.

This indicates the need for two components in the lower level: Bandwidth Distribution (for updating VPC bandwidth) and Load Balancing (for updating route selection parameters).

This hierarchical approach to the problem exhibits fair management behaviour whereby initial management decisions taken with a future perspective are continuously refined in the light of current developments. Apart from its fairness, this behaviour provides a desirable level of adaptivity to network conditions.

For making traffic predictions more accurate, the actual usage statistics are fed back into the Predicted Usage Model to provide it with the necessary data for refining its view of traffic patterns so that future predictions are more accurate as experience is gained. This also allows new trends in usage patterns to be captured.

5.3.3 Synthesis

This section presents an overall description of the functionality involved in the VPC and Routing Management Service based on the components identified in the previous section. This synthesis will lead to the identification of any supporting functionality required for the operation of the main functional components presented in the previous section. The following sections present the resulting system from two viewpoints: static and dynamic.

5.3.3.1 Static part

The static part is achieved by an initial invocation of the Route Design functions (comprised of the VPC Topology, the Routing Plan and the Bandwidth Allocation components). They are treated as a single function here (see Section 5.4.2)). Figure 5.4 shows the information boundaries of the Route Design functions, specifying its input and output information.

The task of the Route Design functions is to identify a network of VPCs and a routing plan which satisfies the predicted usage requirements given the constraints imposed by the network model and the performance targets of the CoSs. This process involves defining the link performance parameters, the topology of the VPC network, the routing plan and the bandwidth reserved on each VPC.

First of all "highways" are identified, these involves the allocation of performance targets (e.g. the cell loss ratio) to the links and the construction of logical highways by concatenating links with the same performance target. Mapping the predicted traffic to highways results in the identification of routes which are used to generate the VPC topology. The routing plan, i.e. the definition of route selection tables and the selection priorities, is created and finally the initial VPC bandwidth allocation is decided.

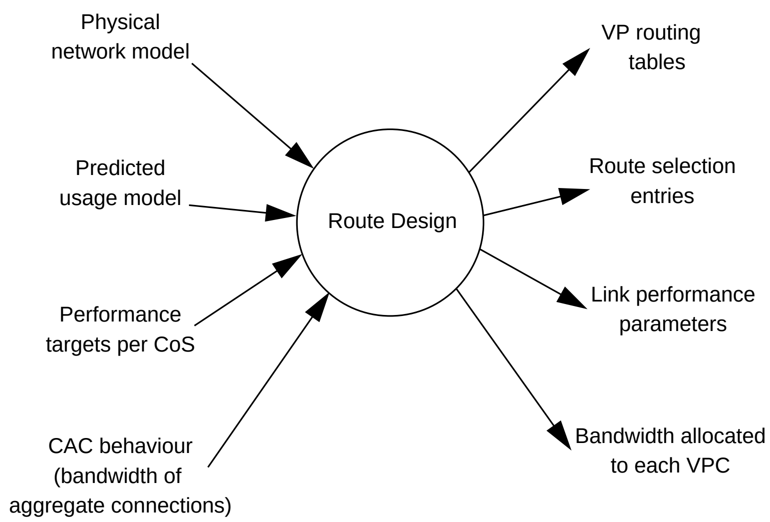

Figure 5.4 VPC and routing management - static aspect

All these tasks are part of an iterative procedure corresponding to complex optimisation problems, aiming at providing cost effective design solutions. The optimisation targets reflect the business policy objectives of the network operator.

Part of this process will be to identify all the possible routes between sources and destinations and to choose a certain subset. The subset chosen will depend on many factors:

- The maximum delay and delay jitter targets for each CoS. Some routes may be unsuitable for some CoSs because they are routed through too many switches and link buffers.
- The cell loss probability requirements for each CoS. If there are sufficient numbers of connections with different requirements on cell loss probability, it may be possible to route them differently. Links, and the VPCs traversing them, could be labelled as high loss rate or low loss rate for example. The cell loss probability targets should be defined for each link and forwarded to the CAC functions responsible for them.
- The connection acceptance probabilities associated with each CoS. A subset of VPCs can be designed for CoSs requiring a high acceptance probability (priority) and additional bandwidth can be assigned to these VPCs at the expense of the VPCs supporting CoSs requiring a lower acceptance probability.
- Which CoSs can be multiplexed together on the same VPCs or links based on the compatibility of their performance parameters as described above.
- The number of logically and physically diverse routes required to ensure the availability of the network in fault and damage conditions.
- 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.

Table 5.1 shows the entities to be specified or managed to ensure the performance of each CoS.

Performance Parameter	Management Solution
Cell loss	Set CAC cell loss target
Delay	Route definition
Delay jitter	Route definition
Call blocking	Bandwidth reservation schemes

Table 5.1 Managing performance parameters

The overall task is a planning and design exercise, VPC and Routing management must define the VPCs and routes based on the predicted traffic and the physical constraints of the network.

The physical constraints are related to the topology of the network and the capacity of the links (assuming that the switches are non blocking). This information is maintained in a network model. It is assumed that maintenance and fault management functions will update the network model as resources are taken in and out of service.

An estimate of the quantities of traffic that will be offered to the network is required. This data can be predicted by a combination of information from the service providers, customer administrations and by historical trend information. A Predicted

Usage Model is therefore required to maintain the anticipated number of connections of each CoS between all source-destination pairs. The model identifies how traffic patterns change over time and will reflect for example the differences between business and domestic traffic over a working day, or how traffic patterns change over the course of a week.

In order that this model can be translated into bandwidth requirements, knowledge is required of the bandwidth characteristics of each connection type, and how these characteristics sum for the purpose of statistical multiplexing. This information depends on the CAC algorithm deployed in the Network Elements and the parameters it uses to characterise the different connection types. The bandwidth and performance characteristics of each connection type are stored in a Connection Type Model. A CAC Manager is needed to model the behaviour of the CAC algorithms in the network in order to calculate the bandwidth needed by aggregate connections.

The route design process will generate the initial configurations of:

- the set of VP routing tables to be downloaded to the VP cross-connects,
- the set of route selection tables to be downloaded to the routing functions in CC in the VC switches,
- the bandwidth allocated to each VPC,
- the link performance parameters to be downloaded to the CAC functions in the VC switches.

5.3.3.2 Dynamic part

There are three dimensions to the dynamic part, distinguished by function but also by the timescale over which they operate:

1. *VPC Bandwidth Management*

   This is further decomposed into two levels, again distinguished by function and by timescale:

   - *Bandwidth Allocation*

     The Bandwidth Allocation component is invoked whenever the predicted traffic changes significantly. Based on the predicted usage, the source-destination predictions are mapped to VPCs according to the current routing plan, and the minimum bandwidth required by each VPC in order to meet the predicted demand is identified.

     If it is impossible to allocate sufficient bandwidth for the predicted traffic within the constraints of the current routing plan and the link capacities, the Route Planning component is notified, see below.

     If the predicted demand does not consume the full link capacity the Bandwidth Distribution component, below, will distribute the remaining capacity (an implied “common pool” bandwidth) among the VPCs.

   - *Bandwidth Distribution*

     Taking the current load into account, the Bandwidth Distribution component, implements the allocation of bandwidth to VPCs as requested by the Bandwidth Allocation component. The current load must be considered to avoid situations where the predicted required bandwidth (from Bandwidth Allocation) is lower than the current measured load and implementing the

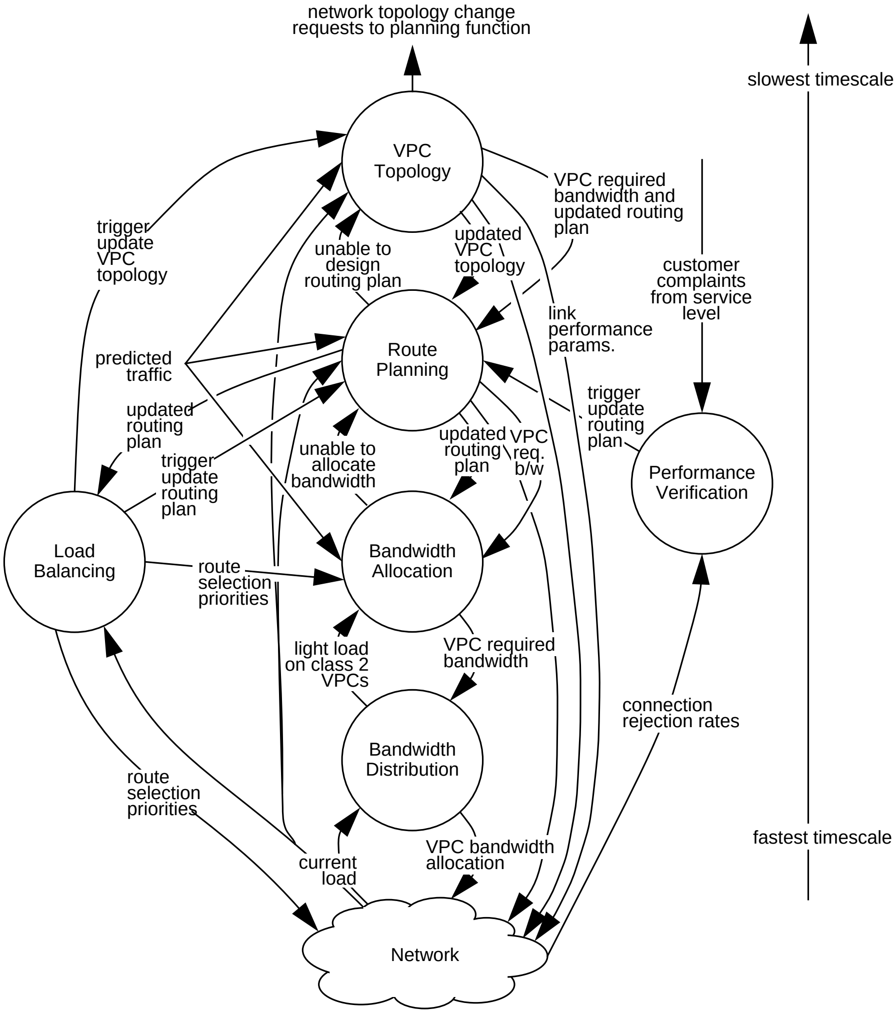

Figure 5.5 VPC and routing management - dynamic aspect

new bandwidth allocation would reduce the VPC bandwidth allocation below the existing bandwidth consumption and violate the assumptions made by the CAC algorithms and possibly cause excessive cell losses.

In addition to implementing the policies of the Bandwidth Allocation component, Bandwidth Distribution attempts to compensate for inaccuracies in the traffic predictions generated by the Predicted Usage Model by distributing any unallocated link bandwidth (the “common pool”) among the VPCs.

Additionally, unused bandwidth (allocated bandwidth minus current load) in each VPC is redistributed among the VPCs sharing the same routes to avoid situations where some VPCs are heavily utilised (and consequently there is little bandwidth available for new connections) whilst other VPCs on the same links are lightly utilised. Unused bandwidth is distributed as evenly as possible within certain constraints. For example, VPCs can be assigned a class or priority attribute to indicate which VPCs should gain unused bandwidth at the expense of lower priority VPCs. VPCs used for private leased lines or for CoSs with low blocking probabilities will be assigned higher priorities. This facility will control how seriously the Bandwidth Distribution component should take the predictions in the form of bandwidth requirements specified by Route Design. If high priority VPCs (class 2 - see Section 5.6.1) are found to be lightly used, the Bandwidth Allocation component is notified to request a reduction in the required bandwidth.

The Bandwidth Distribution component is invoked whenever there is an update in required VPC bandwidth from the Bandwidth Allocation component. Furthermore it periodically polls the Current Load Model to adapt VPC bandwidth to current usage. This latter activity will be performed relatively frequently, every 15 minutes for example. By varying the averaging interval the sensitivity of the Bandwidth Distribution function can be controlled.

2. *Routing Management*

This is further decomposed into two levels, again distinguished by function and by timescale:

- *Reconfiguration of the routing plan* (Route Planning component)

  On receipt of an indication from the Bandwidth Allocation component that it is unable to meet the demands of the predicted traffic, the Route Planning component first of all attempts to redesign the routing plan on the existing VPC network, to remove bottlenecks for example. It tries to increase the number of alternative routes, using the current VPC topology. This process also identifies the new bandwidth requirements on the VPCs - but it must take into account the current load (existing connections must not be disturbed).

  In order to enhance alternative routing and to compensate for inaccuracies in the routing estimates, Route Planning may assign a set of ‘back-up’ routes to each CoS in addition to the primary set of routes. For a given CoS, the set of ‘back-up’ routes may only be chosen from the set of routes allocated to the higher quality CoSs.

  The output will be the updated routing plan and new required bandwidth for each VPC. The defined routes per CoS are given to the Load Balancing component to identify route priorities according to current traffic patterns. The bandwidth requirements are passed down to the Bandwidth Allocation component who simply passes them to Bandwidth Distribution without any additional processing.

  If the Route Planning component cannot design a new routing plan to accommodate the predicted traffic due to limitations in the existing VPC network topology an indication is sent to the VPC Topology component.

- *Management of route selection parameters* (Load Balancing component)

  Load Balancing takes a network-wide view and tries to influence the routing decisions so that new connections use the routes with the highest availability. The routes at a node are prioritised according to certain criteria. Since ATM networks are connection-oriented, the selection criteria are related to the availability of the route to accommodate new connections (spare capacity). This way the network load distribution may be regulated (hence the name Load Balancing).

  Moreover, the Load Balancing component, based on actual usage records, quantifies network availability for accepting new connections not only at the access, but also at the transit nodes. Based on these measurements, Load Balancing notifies the Bandwidth Allocation component of undesirable trends in network availability with the purpose of either redefining VPC bandwidth or creating new routes (via the Route Planning or VPC Topology components above) as appropriate.

3. *Reconfiguration of VPC network topology* (VPC Topology component)

   The VPC Topology component redesigns the VPC network topology to meet the new requirements. This must take into account the current topology and the fact that there are existing connections which must not be disturbed.

   New VPCs may be created to coexist with the current ones and a new routing plan will be defined so that the new VPC topology may be introduced gradually for new connections (by using the priority field in the route selection tables). The bandwidth requirements for the VPCs in the final VPC topology are identified and passed down to the lower level components who do no further processing on them until they arrive at the Bandwidth Distribution component. In this latter function, the bandwidth allocated to the old VPCs is reduced when old connections are released and the bandwidth allocated to the new VPCs is increased to carry new connections as link bandwidth becomes available.

   If the VPC Topology component is unable to design a VPC network to satisfy the traffic demand because of limitations in the underlying physical network, e.g. not enough links, it will notify the network planning functions to deploy additional network resources. Conversely if the VPC Topology component finds that the physical network is over-resourced for the predicted traffic, it will notify the planning function that resources may be taken out of service, or the service management functions so that appropriate service migration actions may be taken.

5.3.4 Aspects of hierarchical management

This section describes how VPC bandwidth and routing management is achieved in the framework of the proposed hierarchical system.

5.3.4.1 VPC bandwidth management

Based on the usage of the network VPC and Routing management will attempt to modify the bandwidth allocated to VPCs to accommodate changes before and as they happen. This is an on-going task that will be performed throughout the operation of the network.

The modifications to VPC bandwidth occur at two levels (Figure 5.6), a coarse level based on the predictions of the Predicted Usage Model and a fine adjustment to tailor the allocated bandwidth to that actually consumed by the connections on the VPCs. These levels are also distinguished by the time periods over which they work. The coarse level makes adjustments at relatively long intervals, e.g. every few hours, whilst the fine tuning is done much more frequently, e.g. several times per hour.

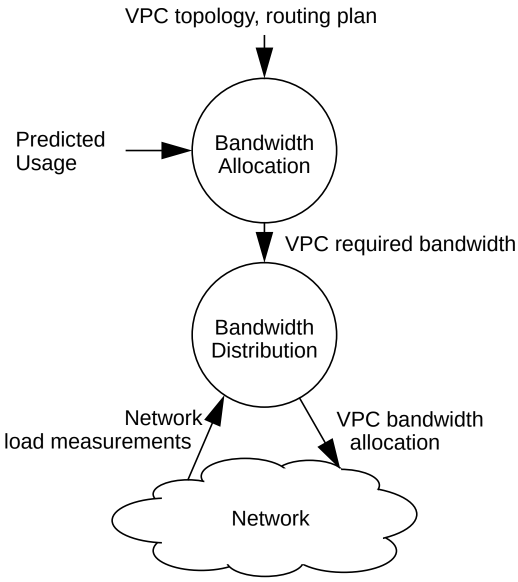

Figure 5.6 Bandwidth management - a hierarchical approach

At the coarse level, realised by the Bandwidth Allocation component, the Predicted Usage Model periodically sends traffic predictions so that VPC bandwidth may be modified to accommodate the predicted traffic for the next time period within the defined set of routes.

At the lower level, the Bandwidth Distribution component tracks the number of connections on the VPCs to identify whether their allocated bandwidth needs to be modified to match measured usage requirements. A VPC could have too much bandwidth and hence other VPCs are deprived of capacity, or a VPC could have too little bandwidth and it may be possible to increase it’s allocation by redistributing bandwidth from other VPCs.

When discrepancies are noticed Bandwidth Distribution will tune the allocated bandwidth to the measured value by increasing or decreasing the allocated bandwidth but ensuring that there is enough spare bandwidth to allow additional connections to be set up.

5.3.4.2 Routing management

As for VPC bandwidth management, routing management is also achieved through a two level hierarchy (Figure 5.7).

The higher level of the routing management hierarchy - the Route Planning component - operates at epochs where network usage predictions change, producing new sets of routes per CoS, based on the current set of VPCs.

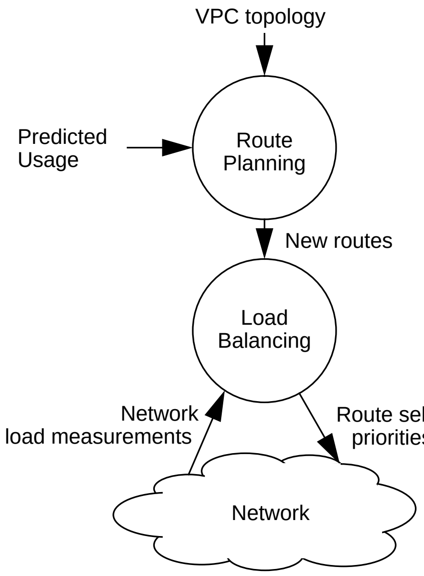

Figure 5.7 Routing management - a hierarchical approach

The lower level - the Load Balancing component - operates within the time-frame of network usage predictions and within the defined set of VPCs and routes from the higher level. Load Balancing influences route selection parameters based on actual network usage. The lower management level is introduced to compensate for inaccuracies in network usage predictions and short-term fluctuations of the load around the predictions. The Load Balancing component aims at making efficient use of the network resources defined for routing (VPCs). Taking into account the multi-class network environment and the fact that the routes of all CoSs share the same VPC infrastructure, efficient use of VPC resources should be achieved not only at the level of a single CoS, but also at the overall CoS level. Moreover the Load Balancing component warns the higher level component of undesirable trends in network availability based on actual usage measurements.

5.4 Mapping to the TMN architecture

5.4.1 MSCs and MFCs

The previous section indicates the following decomposition of the VPC and Routing Management Service into MSCs:

- *A Management of VPC Topology MSC* which is placed in a VPC Topology MFC
- *A Management of VPC Bandwidth MSC* which is further decomposed into:
  - a VPC Bandwidth Allocation MFC
  - a VPC Bandwidth Distribution MFC
- *A Management of Routes MSC* which is further decomposed into:
  - a Route Planning MFC
  - a Load Balancing MFC
- *A Performance Verification MSC* which is placed in a Performance Verification MFC

Apart from the Load Balancing MFC and the Performance Verification MFC, the above MFCs have both static and dynamic aspects. The static aspect is related to the network planning activity and is used to initialise the network. The dynamic aspect reconfigures the VPC network and the routing plan to adapt to changing conditions during the operation of the network.

Additionally, the following support MFCs are required:

- a Configuration Management MFC which includes the network model
- a Current Load Model MFC for providing the required network statistics
- a CAC Manager MFC for the TMN to reproduce the CAC behaviour for dimensioning purposes
- a Predicted Usage Model MFC
- a Connection Type Model MFC

5.4.2 Mapping to the TMN functional architecture

Figure 5.8 shows the allocation of MFCs to OSFs and also places the OSFs into the architectural layers.

By adopting a hierarchical TMN architecture, the merits of centralised and distributed management approaches are combined. Frequently used, lightweight, management functions have been pushed as close as possible to the network elements to avoid the management communications overhead inherent in centralised systems while comprehensive management applications exhibiting complex functionality are located in higher levels of the hierarchy.

5.4.3 Description of the architectural components

This section briefly describes the basic functionality and operation of the main components in terms of their input, objectives, constraints and results. The description of their functionality has been presented in Section 5.3.3.2, subsequent sections deal with some of these components in more detail.

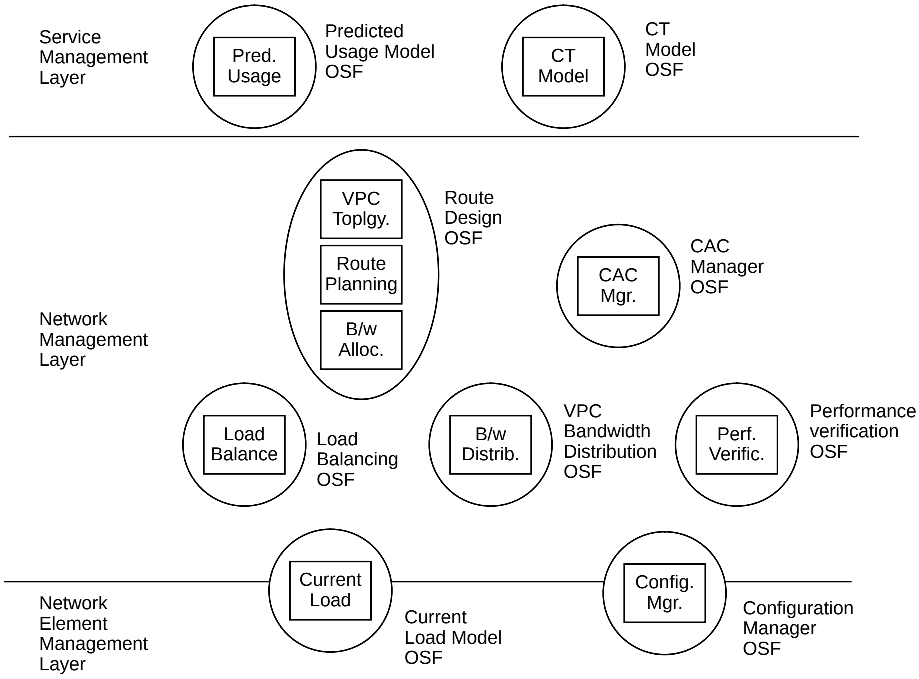

Figure 5.8 Mapping of MFCs to OSFs and OSFs to the TMN layers

5.4.3.1 Route Design OSF

Input:

- Physical network
- VPC network[1]
- SDClassRoute networks[1]
- Traffic predictions
- Current network load on the VPCs[1]

Objective:

- To redesign the VPC network and SDClassRoute networks so that the predicted usage is met within the performance constraints of the CoS.

Constraints:

- This OSF must take into account the current topology (both the physical topology of the network and the logical topologies of VPCs and routes from previous invocations of this component) and the fact that there are existing connections which must not be disturbed.

1. This is not input for the static invocation (see description)

- The link capacity sharing constraint (the sum of VPC bandwidth allocation must not exceed the link capacity) should be taken into account when reconfiguring the VPC network.

Result:

- Updates of VPC bandwidth
- Creation/Deletion of VPCs
- Creation/Deletion of routes
- Updates of CAC cell loss targets

Dynamics:

This OSF operates in an asynchronous mode, being triggered by other OSFs. Specifically, it is triggered whenever:

- The are significant changes in the predicted traffic (from the Predicted Usage Model OSF)
- There are undesirable trends in route availability (from the Load Balancing OSF)
- There are significant deviations in link loads (from the Load Balancing OSF)
- The required bandwidth of a VPC has been over estimated compared to actual load of the VPC (from the VPC Bandwidth Distribution OSF)
- There are rejection rates for a CoS which exceed the specified CoS levels (from the Performance Verification OSF)

If the Route Design OSF is unable to design a VPC network to satisfy the traffic demand because of limitations in the underlying physical network, e.g. not enough links, it will notify the network planning functions to purchase additional network resources or bring existing but unused ones into service. Service management functionality may also be triggered to invoke service migration procedures. Conversely if this OSF finds that the physical network is over resourced, it will notify the planning function that resources may be taken out of service.

5.4.3.2 VPC Bandwidth Distribution OSF

Input:

- The capacities of the transmission links
- The VPC topologies (mapping to links)
- The class or priority of the VPCs
- The bandwidth required per VPC (predictions via Route Design)
- The used bandwidth measured per VPC (from the Current Load Model)

Objectives:

- To distribute unallocated link bandwidth to all VPCs taking into account that each VPC usually spans more than one link and that each link usually contains more than one VPC.
- To allocate the required bandwidth predicted by the Route Design OSF.

Constraints:

- The VPC capacity must not be reduced below the measured used bandwidth.
- The sum of VPC bandwidth allocations over all VPCs spanning a link must not exceed the link capacity.

Result:

- VPC bandwidth updates (to network elements via the Configuration Manager)
- Monitoring requests, statistics and threshold requirements (to the Current Load Model)

Dynamics:

The Bandwidth Distribution OSF is invoked whenever there is an update in required VPC bandwidth from the Route Design OSF or whenever VPC bandwidth allocations need to be adjusted to meet current usage requirements. The latter invocations may operate in two modes: a synchronous mode where the activation interval is left as a design option; or an asynchronous mode of operation whereby it is activated by threshold crossing events. In the latter case the facilities of the Current Load Model can be used to set thresholds on the used bandwidth measured on VPCs so that event reports can be received by the Bandwidth Distribution OSF when VPCs are lightly or heavily loaded in order to cause a redistribution of bandwidth.

5.4.3.3 Load Balancing OSF

Input:

- Physical network
- VPC network
- SDClassRoute networks
- Current network load on the VPCs

Objective:

- To manage the Route Selection Algorithms of the network according to network-wide traffic conditions. The management of the Route Selection Algorithm is achieved by tuning the route selection parameters.
- To monitor the network with the purpose of warning the Route Design OSF of deterioration in route availability or load deviations at link level (indicating inefficient use of transmission facilities).

Result:

- Updates of Route Selection Parameters

Dynamics:

This OSF operates in a synchronous mode. The activation interval is left as a design option. However, an asynchronous mode of operation could be envisaged, whereby it is activated by threshold crossing events.

5.4.3.4 Performance Verification OSF

The objectives of the Performance Verification OSF are:

- To ensure that the network meets the performance targets of the different CoSs supported by the network.
- To warn performance management related components when network performance has dropped below the acceptance levels per CoS, so that corrective actions can be taken.

- To analyse customer complaints with respect to the quality of the network services they use.

The measured connection rejection rates per CoS and per source destination pair are retrieved from the Current Load Model and compared to the rejection rate targets as specified per CoS. If CoSs are found to be experiencing connection rejection rates in excess of their target, an indication is sent to the Route Design OSF to cause the number of routes, or the bandwidth allocated to the routes, to be increased.

Customer complaints are analysed and if they are justified the Route Design OSF will be triggered.

5.4.3.5 Predicted Usage Model OSF

This models the predicted usage of the network in terms of the generation pattern of connection requests of each CoS between s-d pairs. The model details how the connection generation pattern changes: hour by hour over the day; day by day over the week; and week by week over the year

Initially this is configured by the service level of the TMN but it is modified by the actual usage of the network via the Current Load Model. This is so that the model becomes more accurate as experience of the usage of the network is gained.

Whenever the Predicted Usage Model indicates that the traffic will change significantly the Route Design OSF will be provided with a prediction of traffic for the next time interval. The exact definition of a significant change is a design variable to be experimented with according to the performance of the system as a whole.

The predictions are in the form of: mean and peak number of active connections; mean and variance of inter-connection arrival times; and mean and variance of connection holding times.

5.4.3.6 Configuration Manager OSF

The Configuration Manager is responsible for maintaining a consistent model of the physical and logical configuration of the network. It will receive configuration actions from the other OSFs and be responsible for implementing the configuration requests in the network. This task may involve coordination of configuration actions over a number of network elements, for example when a VPC is created.

The Configuration Manager can provide event reports to the other OSFs whenever a configuration action has succeeded. This will be achieved by use of the object management systems management function and events will be generated according to event forwarding discriminators registering each components interest in the information.

5.4.3.7 Current Load Model OSF

The Current Load Model monitors the network usage and calculates usage statistics according to the requirements of the other OSFs. Two types of parameters are required by the VPC and Routing management system: bandwidth usage statistics on a VPC and link level; and connection statistics in terms of the number of active connections and connections rejected per CoS and s-d pair on a VPC or node basis.

The Current Load Model is capable of calculating peak, mean, EWMA, etc. statistics according to the specifications of the other components. It will identify the minimum number of network probes and measurements to meet the varied demands of its users.

5.4.3.8 CAC Manager OSF

The CAC Manager reproduces the CAC algorithm deployed in the control plane of the network. When supplied with a traffic mix in the form of a list of the number of connections of each CoS the CAC Manager returns the aggregate bandwidth of that traffic mix. The calculation has exactly the same result as the equivalent CAC algorithm in the network. Additionally the CAC Manager can deal with requests such as “how many connections of CoS x can be accommodated in a VPC of capacity y?”

5.4.3.9 Connection Type Model OSF

This models the bandwidth and performance targets for each CoS, and acts as a central database for the other OSFs requiring this information.

5.4.4 Interactions between the architectural components

Figure 5.9 shows the manager-agent relationships between the derived components.

As shown in the figure Load Balancing and VPC Bandwidth Distribution are both agents of Route Design. Load Balancing and VPC Bandwidth Distribution interact in manager and agent roles. Performance Verification is a manager of Route Design sending it alarms in the form of actions indicating undesirable trends in network availability. All these components act in a manager role with respect to the supporting components.

VPC Bandwidth Distribution and Load Balancing are both agents of Route Design but they are assigned different tasks. In particular, VPC Bandwidth Distribution takes a limited view (that of a link) and tries to allocate the link capacity to the defined VPCs according to their load level. As a result, the bandwidth of a VPC may be increased or decreased. The Load Balancing manages the RSAs in the control plane aiming at making their routing decisions network-state dependent. However, their operation is not totally independent, since the effect (in the network) of one of them is taken into account by the other. VPC Bandwidth Distribution focuses on the current load of the VPCs, which, to a great extent, is determined by the routing decisions, and Load Balancing looks at the availability of the VPCs (for accepting new connections) which is determined by VPC capacity and hence on the decisions of Bandwidth Distribution. This indicates that some coordination is required between them, to avoid possible contradictions.

To better assess the difference in their functionality and the operational dependencies, the notion of the VCC network is introduced. The VCC network is defined in a similar way to the VPC network but the links correspond to VCLs instead of VPCs. The physical network shows the long-term provisioning of network resources, the VPC network shows a medium-term provisioning of network resources and the VCC network shows the actual usage of the network resources.

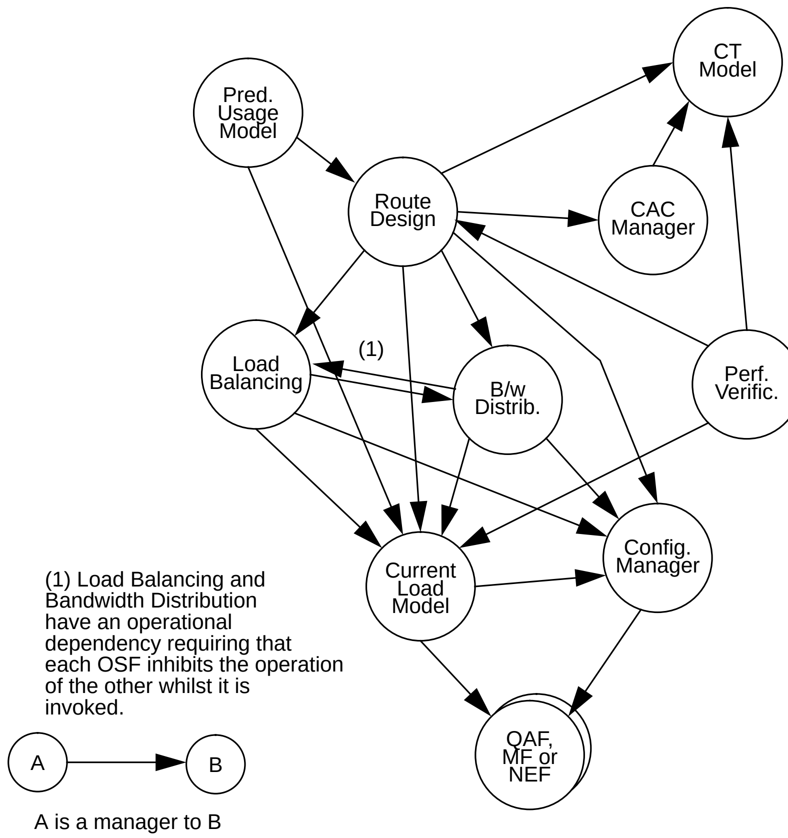

Figure 5.9 Manager-agent relationships between the OSFs

The Route Design and the VPC Bandwidth Distribution OSFs manage the VPC network. The Route Design OSF is the top level manager and has the authority to reconstruct the VPC network. VPC Bandwidth Distribution is an agent of Route Design, acting on a smaller time-scale, being assigned with the task to tune the bandwidth of the links of the VPC network, as necessary.

Load Balancing is not concerned with the construction of VPC networks. Rather, given a VPC network, its goal is to optimise its use by influencing the routing strategies in the network. In other words Load Balancing (together with the RSAs in the control plane) is concerned with the construction of the VCC network.

From the above analysis it is clear that the scope of Load Balancing and VPC Bandwidth Distribution is different, since they manage different resources. However, there are the following operational dependencies.

When Load Balancing is activated and during its operation the VPC network (topology and resources) should be stable. This means that the VPC Bandwidth Distribution OSF should be prohibited from taking actions when the Load Balancing algorithms are invoked. Conversely, when the VPC Bandwidth Distribution OSF changes VPC capacity, the Load Balancing OSF should not be active.

Another operational dependency, with a potential conflict is the following. The VPC Bandwidth Distribution OSF may decrease the bandwidth of a lightly loaded VPC, but this lightly loaded VPC may have been recommended recently (i.e. the route selection priority has been updated in the network, but a significant quantity of new connections has not yet been set up) for routing with a high priority by Load Balancing. First note that the fact that Load Balancing has recommended a VPC for routing, does not mean that traffic will be routed immediately through it; it will be routed when it arrives. Therefore, only if this VPC remains lightly loaded for some time should its bandwidth be reduced. And conversely, only when a VPC seems to have a constant trend of significant spare capacity should be recommended for routing. This means that both OSFs should not be over-sensitive to traffic or spare bandwidth fluctuations. This indicates that the network measurements that each OSF takes should be in terms of moving averages as opposed to instantaneous values, the moving window may being a function of the activation frequency of the other OSF.

To conclude, the functions of the Load Balancing, VPC Bandwidth Distribution and Route Design OSFs are quite different and do not overlap, but there are some operational dependencies requiring careful design and activation coordination.

5.4.5 Physical architecture - TMN building blocks

Figure 5.10. maps the TMN functional blocks from the functional architecture to TMN building blocks, i.e. OSs, WS-OSs, MDs and QAs.

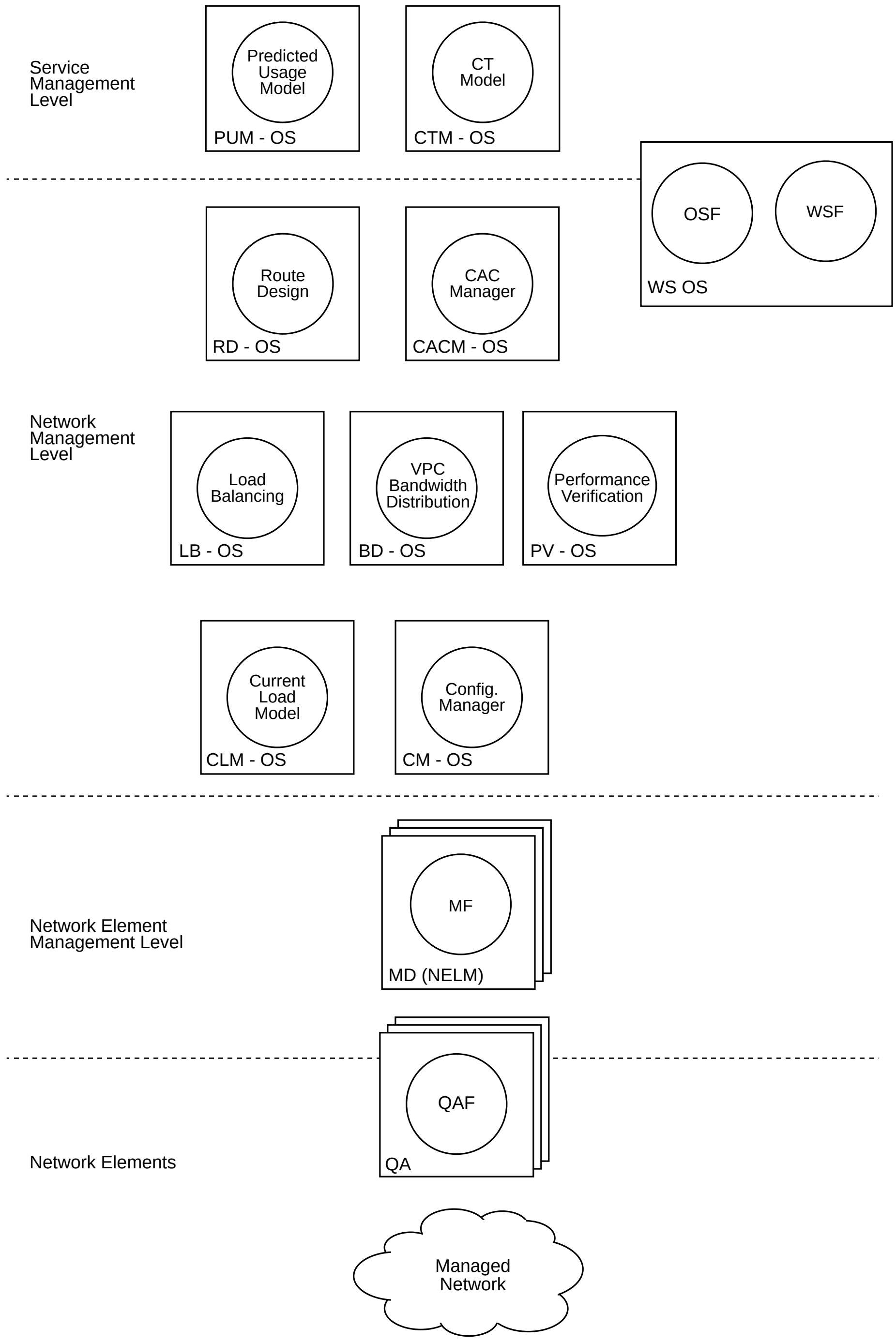

Figure 5.10 Physical architecture for VPC and Routing management

5.5 Route Design

The objectives and functionality of the Route Design OSF have been presented in section Section 5.4.3.1. This section presents an overview of specific algorithms for realising its functionality.

5.5.1 Overview

As already mentioned in previous sections, Route Design has both static and dynamic aspects. The static aspects are related to the network planning activity and is used to initially configure the network in terms of VPCs and routes. This part is performed at initialisation time. The dynamic aspects of Route Design operation caters for changes in network traffic predictions and for prediction inaccuracies that could not be resolved by the lower level management components. The result of the dynamic part is reconfiguration of the VPC and ClassRoute networks.

The design philosophy is to ensure that the dynamic algorithms employed in the Route Design OSF are much simpler than the static algorithms so that the actions to be taken at run-time are as economical as possible in terms of management overhead and required processing power.

Another basic dimension of the design philosophy is to base the algorithms on simple and comprehensive invariant principles. The static and dynamic algorithms should be built around these concepts to guarantee homogeneity in decisions. After all, management should strive to be simple.

The following notions should be taken into account:

- The function operates on an integrated services environment. The diversity in bandwidth requirements and performance targets for each CoS should explicitly be taken into account. Hence the difference of this function with the “usual” routing functions for data (packet-switched) or telephone networks.
- The design should be open, in the sense that it should be possible to experiment with different algorithms in the components while the overall architecture and the external interfaces of the components remain identical. This design objective is in line with the overall objectives of the ICM project and allows for future research and development work.

5.5.2 Terminology

Path: A sequence of links connecting two network nodes.

Route: A sequence of VPCs connecting two network nodes.

VPC network: The network formed by the VC switches interconnected by VPCs.

ClassRoute network: For a given CoS, the ClassRoute network is a sub-network of the VPC network consisting only of the VPCs that belong to routes of that CoS.

SDClassRoute network: For a given CoS and a given pair of source-destination (s-d) nodes, the SDClassRoute network is the sub-network of the ClassRoute network consisting only of the VPCs that belong to the routes interconnecting the given (s-d) pair.

SDClassPath network: For a given CoS and a given (s-d) pair, the SDClassPath network is the sub-network of the physical network consisting of the nodes and links that appear in the paths of the given (s-d) pair and CoS.

5.5.3 The static part of the algorithm

5.5.3.1 Problem description

Given:

- The physical network,
- the CoSs and their performance characteristics, and,
- the traffic predictions per (s-d) and CoS.

Objective:

- Design VPC and SDClassRoute networks subject to the constraint of meeting the predicted usage within the performance constraints of the CoSs.

5.5.3.2 The algorithm

The proposed algorithm is evolved in four steps.

1. Map traffic predictions into max flow requirements for all (s-d) pairs and each CoS. The max flow requirements are in terms of VCCs that the network must be able to establish at any instant, given the max connection rejection tolerance of each CoS.
2. Given the max flow requirements (established in previous step), determine suitable paths per (s-d) and CoS so that to satisfy the delay, jitter and cell loss constraints.
3. Given the set of paths per (s-d) and CoS (established in step 2), determine a suitable VPC network.
4. Map the set of paths (established in step 2) to the derived VPC network (established in step 3) to obtain the set of routes per (s-d) and CoS

The following sections provide some detail on the above steps.

5.5.3.2.1 Determining max flows

It is assumed that the predictions are in the form of number of connection requests for a specific interval. By modelling the network with an appropriate queueing system, it is possible to obtain the (minimum) number of VCCs which the network should provide at any instant so that the blocking probability is less than or equal to the maximum tolerable rejection ratio per CoS. The derived number also denotes the maximum flow per (s-d) and CoS that must flow within the physical network.

5.5.3.2.2 Determining suitable paths

The next step is to establish suitable paths per (s-d) and CoS to satisfy the performance requirements of each CoS.

The algorithm guarantees upper bounds on delay, jitter and cell loss ratio per CoS by defining appropriate set of paths per (s-d) and CoS. Each path is assigned a performance quality corresponding to an upper bound on the delay, jitter and cell loss that a cell of a connection set up on this path would experience. Based on its performance quality, a path then falls into one of the performance categories defined for the CoSs.

The algorithm proposes various policies for allocating (s-d) pairs of specific CoS to paths that offer a specific performance quality. The following three *performance sharing policies* are proposed:

- *Complete sharing*: where a path of a specific performance category is shared exclusively by the CoSs of the same performance category.
- *Compatible sharing*: where a path of a specific performance category is shared only by CoSs of the same or superior performance category.
- *Non-compatible sharing*: where a path of a specific category can be shared by any CoS.

In order to meet the cell loss requirements, the cell loss performance targets of the CAC algorithms deployed in the network switches need to be determined and then configured. It is assumed that the CAC algorithms corresponding to VPCs defined on the same link will have the same cell loss performance target. The allocation of cell loss performance targets for each link can be achieved according to the following three *link performance assignment policies*.

- *Complete sharing*: Links are assigned with performance targets in such a way so that paths of any cell loss category can be established on them.
- *Complete partitioning*: Links are assigned with performance targets in such a way so that paths of specific cell loss categories can be established on them.
- *Complete sharing with initial reservations*: Links are assigned with performance targets in such a way so that paths of specific cell loss categories can be established on some links while paths of any cell loss category can be established on the remainder of the links.

For a given performance sharing and link performance assignment policy, there may be many paths for a given (s-d) and CoS that satisfy the performance constraints of each CoS. The number of paths is however finite, constrained by the number of paths that can be found for each (s-d) pair in the physical network. So the following questions arise:

- are there any paths per (s-d) and CoS that satisfy the performance and max flow constraints?
- if yes, which paths should be selected per (s-d) and CoS?

To answer the latter question, the algorithm proposes a means to quantify the benefits and the costs of alternative SDClassPath networks. The selection of the optimum SDClassRoute networks design is made to satisfy one of the following overall goals:

- maximum benefits (irrespective of cost)
- minimum cost (irrespective of the benefits)
- cost-effectiveness (the maximum possible benefits before the cost becomes prohibitive)

The above can be formulated as optimisation problems, the solutions to which yield the optimum SDClassPath networks.

5.5.3.2.3 Determining the VPC network

The next step is to define a suitable VPC network. The VPC network is introduced in order to:

- simplify the route networks that need to be managed (i.e. routes are defined in terms of VPCs rather than links), therefore increasing management flexibility;
- avoid situations where the source node information is required for route selection; these situations occur in transit nodes where not all starting paths to a particular destination for a given CoS are admissible (due to delay constraints, for example).

The definition of the VPC network encompasses:

- the definition of the topology of the VPCs,
- the definition of the (required) bandwidth of each VPC.

To determine the VPC topologies, the nodes that will be the VPC termination points are selected first. These nodes are referred to as *VPC cut nodes*. The VPC cut nodes will act as VC switches, the remainder will act as VP cross-connects. The VPC cut nodes are connected with (at least) as many VPCs as the number of physically different paths that connect them.

The access nodes will always be VPC cut nodes. There are many combinations of VPC cut nodes resulting in different VPC network topologies. At one extreme only the access nodes will be VPC cut nodes. At the other extreme all network nodes will be VPC cuts.

5.5.3.2.4 Determining the SDClassRoute networks

Having determined the VPC network, the SDClassRoute networks can be produced easily by mapping the paths produced in step 1 to the VPCs established in the previous step.

5.5.4 The dynamic part of the algorithm

5.5.4.1 Problem description

Given:

- The physical network,
- the CoSs and their performance characteristics,
- the traffic predictions per (s-d) and CoS, and,
- the load on the VPCs.

Objective:

- Design VPC and SDClassRoute networks subject to the constraint of meeting the predicted usage within the performance constraints of the CoSs without disturbing the existing connections.

5.5.4.2 The algorithm

1. Map the traffic predictions to max flow requirements.
2. Determine whether the new flow requirements can be accommodated in the existing VPC and SDClassRouteNetworks, keeping the same VPC capacity, under the constraints of optimality (established by the static part algorithm).
3. If not possible, try to increase the VPC bandwidth (preserving the link capacity constraints), keeping the same SDClassRoute networks to accommodate the new flow requirements, under constraints of optimality (established by the static part of the algorithm).
4. If not possible, modify SDClassRoute networks either by adding/removing existing VPCs and by appropriately modifying their bandwidth, and/or creating/ deleting VPCs and by appropriately setting their bandwidth so that to accommodate the new flow requirements, under constraints of optimality (established by the static part of the algorithm)
5. If not possible, trigger either:
   - the network planning Management Services to establish new physical links or nodes, or,
   - the service management functions to migrate resources (change the (s-d) pairs).

5.5.5 Information modelling

The information model of the Route Design OS consists of:

- The *icmRDVPCRouteManager* models the main entry point of the Route Design functions. It enables other OSs to view the status and operational parameters of the Route Design activities. It receives appropriate actions and notifications from its related managers and agents. It is associated with a number of application objects making up the Route Design functionality, triggering the appropriate one according to the received actions or notifications.
- A collection of object instances representing the defined routes.

  The information model is updated whenever VPCs or routes are created or deleted by the activities of the Route Design OSF. A network approach was followed, whereby the information model represents the SDClassRoute networks; rather than explicitly representing the routes for all source destination pairs and CoS as a sequence of VPCs. Adopting this network view, the routes from any node to a particular destination node for a given CoS can be retrieved by means of well known algorithms for path finding in directed graphs. With this approach, the overhead for maintaining the model is minimal as no information is duplicated. The M.3100 classes for network connectivity were extended for representing the connectivity of the VPC and SDClassRoute networks.

The inheritance hierarchy is shown in Figure 5.11 and the containment schema is shown in Figure 5.12.

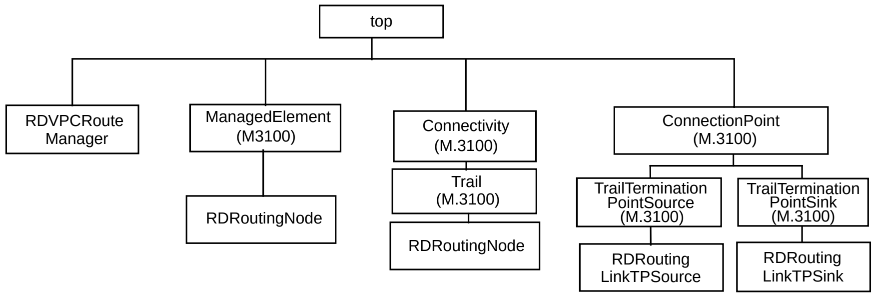

Figure 5.11 Route Design inheritance hierarchy

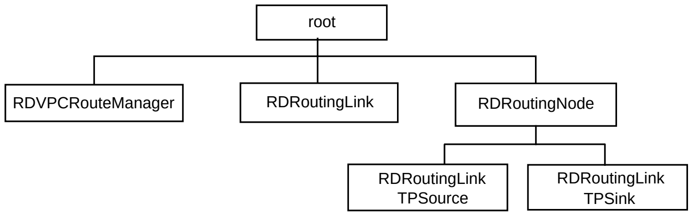

Figure 5.12 Route Design containment schema

5.6 VPC Bandwidth Distribution

The objectives and scope of VPC Bandwidth Distribution was presented in Section 5.4.3.2, this section proposes an algorithm for this component and discusses some of the design issues.

5.6.1 VPC classes

VPC classes are introduced to prioritise the handling of traffic:

- Class 1 VPCs are used for leased lines (see Chapter 6),
- Class 2 VPCs are used for highest priority traffic,
- Class 3 VPCs are used for standard priority traffic,
- Class 4 VPCs are used for low priority traffic.

Higher class VPCs get more of the unused and common pool bandwidth than lower class ones. This facility allows for some VPCs (e.g. those used for leased lines) to always be given exactly the bandwidth they require and for the other VPCs to be assigned a proportion of the unused bandwidth according to their class. Route Design will decide which VPCs should be of each class according to the performance targets of the CoSs they will transport.

5.6.2 Design issues

The top-level execution flow diagram of Bandwidth Distribution is given in Figure 5.13. There are several approaches which could be taken in the interaction between Bandwidth Distribution and the Current Load Model. The two main ones are:

- For Bandwidth Distribution to set up monitoring objects in the Current Load Model at initialisation. Thereafter, on a periodic basis, Bandwidth Distribution would perform a GET on these objects to obtain the current bandwidth on the VPCs. Bandwidth Distribution itself would then determine if threshold have been crossed. In this way no thresholds are actually set up in the Current Load Model. The main advantage of this approach is the control which Bandwidth Distribution has over the amount of information which comes from the Current Load Model. Bandwidth Distribution will only poll the Current Load Model when it is ready to process the information. In this way no (possibly out-of-date) messages from the Current Load Model are queued at Bandwidth Distribution. Also, Bandwidth Distribution can change thresholds with little effort since they are stored internally to Bandwidth Distribution.
- The second approach would be to create monitoring objects in the Current Load Model in the usual way but to also setup thresholds on those objects which emit notifications when thresholds are crossed. Bandwidth Distribution would then create an Event Forwarding Discriminator to convert these into the threshold events it needs. This has the advantage of Bandwidth Distribution only being invoked when it is needed.

5.6.3 The Bandwidth Distribution algorithm

5.6.3.1 Overview

The following steps describe the Bandwidth Distribution algorithm:

- Order the VPCs by Class with Class 1 VPCs first.
- Process any requests from Route Design for reduced bandwidth regardless of Class.
- Allocate the bandwidth increases requested by Route Design to all Class 1 VPCs. (Class 1 VPCs do not have thresholds set on them.).
- Allocate the required bandwidth increases to all Class 2 VPCs. If a given bandwidth increase can not be satisfied then allocate as much as possible.
- Allocate the required bandwidth increases to all Class 3 VPCs. If a given bandwidth increase can not be satisfied then allocate as much as possible.
- Allocate the required bandwidth increases to all Class 4 VPCs. If a given bandwidth increase can not be satisfied then allocate as much as possible.
- For all Class 2 VPCs for which the upper threshold has been crossed, increase the allocated bandwidth to that given by the upper threshold crossing algorithm (see later).
- For all Class 3 VPCs for which the upper threshold has been crossed, increase the allocated bandwidth to that given by the upper threshold crossing algorithm (see later).

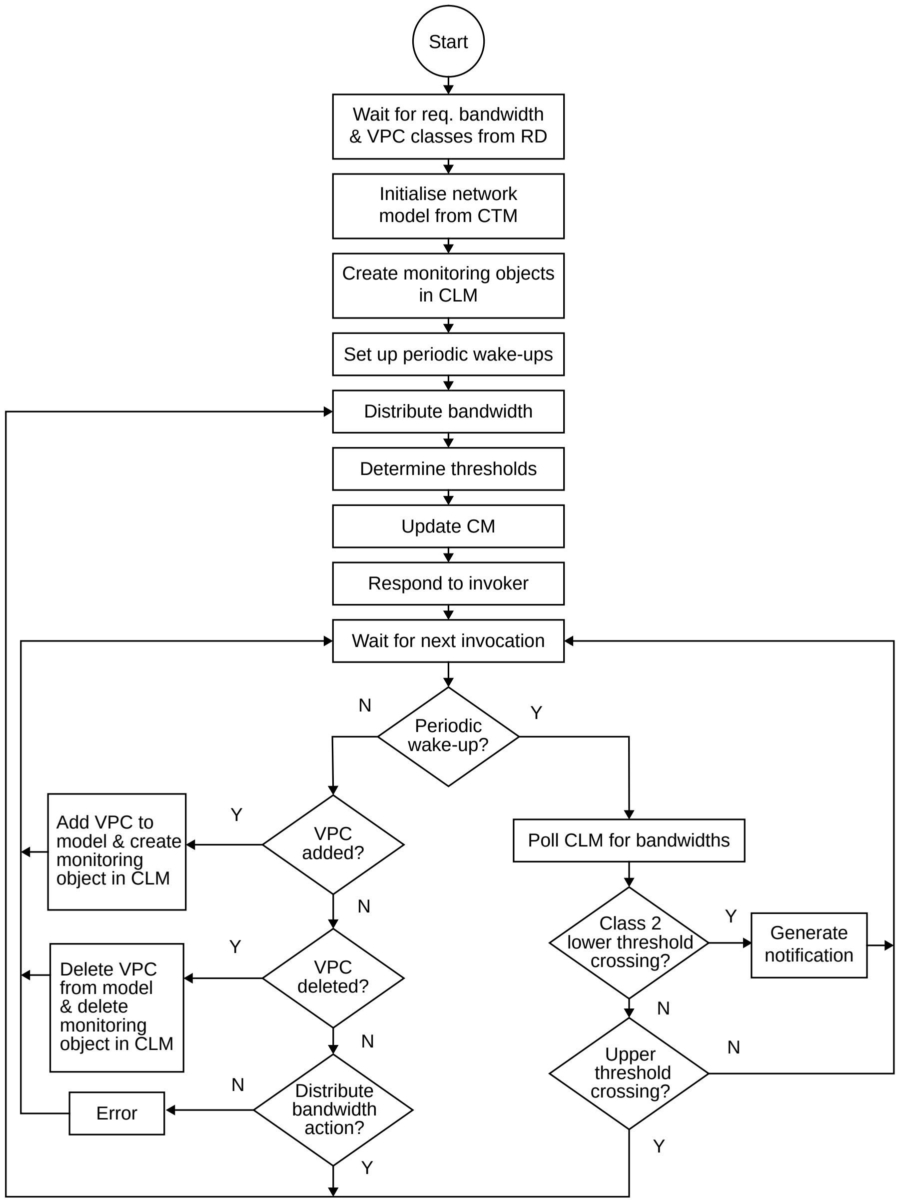

Figure 5.13 Top level execution flow diagram for Bandwidth Distribution

- For all Class 4 VPCs for which the upper threshold has been crossed, increase the allocated bandwidth to that given by the upper threshold crossing algorithm (see later).
- Any spare bandwidth is distributed between Class 2, 3 and 4 VPCs.

An error message is output for each Class 1 VPC which could not be allocated the required bandwidth. Warning messages are output for each Class 2, 3 or 4 VPC which could not be allocated the required bandwidth.

It is assumed that the bandwidth consumed on a VPC, by the connections routed over it, will never exceed the bandwidth allocated to that VPC (assuming that the CAC algorithms in the network are working correctly).

5.6.3.2 Details of the algorithm

The Route Design OSF supplies the initial bandwidth allocation and a class per VPC.

For each VPC the following data are kept:

- $ABW_{VPC\ j}$, bandwidth currently allocated. This is initially set to $RBW_{VPC\ j}$ but is later updated according to the behaviour of the Bandwidth Distribution function.
- $RBW_{VPC\ j}$, bandwidth required. This is determined by the by the Route Design OSF.
- $C_{VPC\ j}$, VPC class.
- $UBW_{VPC\ j}$, effective used bandwidth. This is calculated and updated by the Current Load Model.

  For class 1 VPCs which are used for end-to-end leased lines, the Current Load Model is unable to monitor the connections on the VPC as they are created and terminated by VC switches outside of the public network so UBW_{VPCj} cannot be calculated. In this case $UBW_{VPC\ j}$ is set to the same value as $RBW_{VPC\ j}$.

The following data are kept for each link:

- $BW_{link\ i}$, the capacity of the link obtained from the Configuration Manager.
- $VPCs_{link\ i}$, a list of VPC ids for the VPCs traversing the link, obtained from the Configuration Manager.
- $UBW_{link\ i}$, effective used bandwidth on the link calculated and updated by the Current Load Model.
- $SBW_{link\ i}$, the spare bandwidth on the link calculated by: $SBW_{link\ i} = BW_{link\ i} - UBW_{link\ i}$
- $WBW_{link\ i}$, the wasted bandwidth on the link, i.e. that bandwidth not allocated to any VPC.

The task now is to distribute $SBW_{link\ i}$ for each link among the set $VPCs_{link\ i}$ with the objective of minimising $WBW_{link\ i}$, according to the current usage of the VPCs, i.e. inversely proportionally to the spare capacity of each VPC.

This is not a straightforward task as VPCs span more than one link and bandwidth allocated to a VPC on one link must be allocated to the same VPC on all other links it traverses. A possible algorithm is described below[1].

1. Note that the algorithm assumes that a linear combination of effective bandwidths is suitable for determining aggregate bandwidth requirements.

1. Order links according to $SBW_{link\ i}$, lowest first.
2. Take the 1st link in the list[1].
3. Split $VPCs_{link\ 1}$ into 4 sets $VPCs_{link\ 1,\ class\ c}$ according to $C_{VPC\ j}$.
4. For each VPC in $VPCs_{link\ 1,\ class\ 2}$, $ABW_{VPC\ j}$ is set to $RBW_{VPC\ j}$, $SBW_{link\ 1}$ is reduced by ($RBW_{VPC\ j}$ - $UBW_{VPC\ j}$), and VPC j is tagged as being modified[2].
5. DBW[3] (distributed bandwidth) is calculated by dividing $SBW_{link\ 1}$ by the number of VPCs in $VPCs_{link\ 1,\ class\ 3}$.
6. For each VPC in $VPCs_{link\ 1,\ class\ 3}$,

   if ($UBW_{VPC\ j}$ + DBW) > $RBW_{VPC\ j}$ then $ABW_{VPC\ j}$ is set to $RBW_{VPC\ j}$, $SBW_{link\ 1}$ is reduced by ($RBW_{VPC\ j}$ - $UBW_{VPC\ j}$), and VPC j is tagged as being modified[4].

   else $ABW_{VPC\ j}$ is set to ($UBW_{VPC\ j}$ + DBW), $SBW_{link\ 1}$ is reduced by DBW, and VPC j is tagged as being modified.
7. DBW is recalculated by assigning it the value of $SBW_{link\ 1}$ by the number of VPCs in $VPCs_{link\ 1,\ class\ 4}$.
8. For each VPC in $VPCs_{link\ 1,\ class\ 4}$, $ABW_{VPC\ j}$ is set to ($UBW_{VPC\ j}$ + DBW), $SBW_{link\ 1}$ is reduced by DBW, and VPC j is tagged as being modified[5].
9. $WBW_{link\ 1}$ is set to $SBW_{link\ 1}$ [6]
10. Take the next link in the list.
11. Split $VPCs_{link\ i}$ into 4 sets $VPCs_{link\ i,\ class\ c}$ according to $C_{VPC\ j}$. Remove all VPCs from the lists that have been tagged as modified[7] and reduce $SBW_{link\ i}$ by the increase in $ABW_{VPC\ j}$ for each of these VPCs.

---

1. The reason the link with the smallest spare bandwidth is examined first is that this ensures that it is possible to make corresponding bandwidth increases the other links that the VPC traverses. There is a potential weakness here: the bandwidth increases on lower class VPCs made on this link will have to be made on the other links, so this may consume bandwidth (on the later links) which might have otherwise been given to higher class VPCs.
2. Class 1 VPCs are not considered because there is no question about the amount of bandwidth they should be allocated - it is dictated in a service level contract.

   The VPCs are tagged as modified so that they will not be allocated different bandwidth on the other links they traverse in steps 10 and onwards.
3. Distributed bandwidth. This step calculates how the remaining bandwidth should be distributed among the class 3 VPCs.
4. In other words, “don’t give class 3 VPCs more bandwidth than they require.” This bandwidth is distributed among class 4 VPCs in the next step.
5. The lowest class VPCs are given any bandwidth that is left over from step 6.
6. By this stage $SBW_{link\ 1}$ will hold the unallocated bandwidth on the link.
7. The modified VPCs are removed because they have already been allocated bandwidth and this part of the algorithm should not give them a different allocation. $SBW_{link\ i}$ is reduced because this bandwidth has already been allocated to the already-modified VPCs, and should not be reallocated here.

From this point the previous steps are repeated.

12. to 17. (Steps 12 to 17 are a copy of steps 4 to 9).

18. Repeat from step 10 until all links are exhausted.

After $ABW_{VPC\ j}$ has been calculated, the Bandwidth Distribution function, via the Configuration Manager, modifies the VPC bandwidth allocation in the network for each VPC tagged as being modified.

5.6.3.3 Threshold setting algorithm

A configurable (this is a design variable) fixed percentage of the VPC allocated bandwidth is used for the upper threshold and another for the lower threshold (on Class 2 VPCs). Figure 5.14 shows the possible upper threshold scenarios on a VPC. As mentioned previously, it is assumed that the effective used bandwidth is not allowed to exceed the allocated bandwidth.

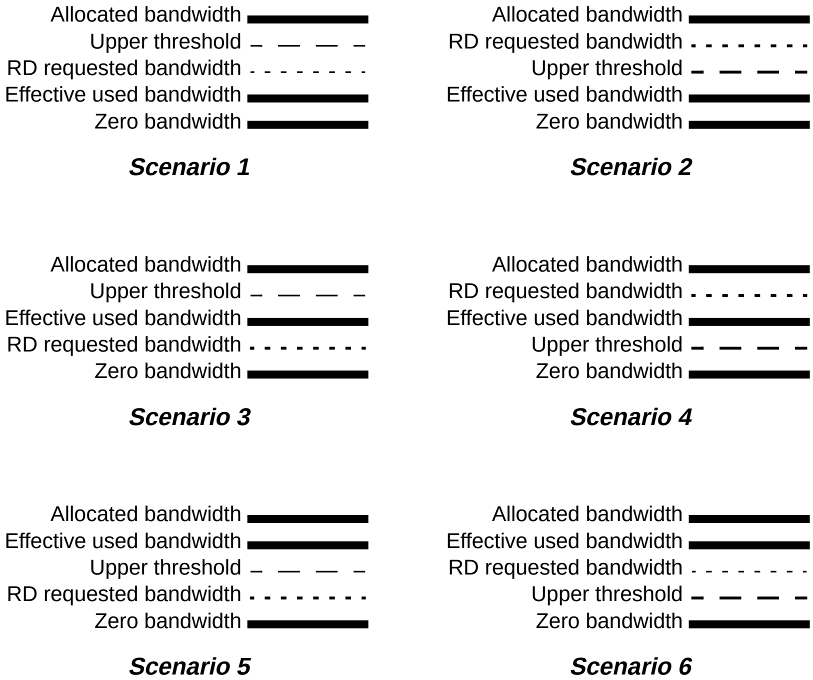

Figure 5.14 Possible VPC upper threshold scenarios

- *Scenario 1*. This is a normal scenario in which the effective used bandwidth is less than both the bandwidth requested by Route Design, and is less than the Upper threshold. Bandwidth Distribution does not need to be activated in this case.

- *Scenario 2*. This is similar to Scenario 1 except that the upper threshold has been calculated to be less than the Route Design requested bandwidth. Bandwidth Distribution does not need to be activated in this case.
- *Scenario 3*. This is an abnormal situation in which the effective used bandwidth has exceeded the Route Design requested bandwidth but has not exceeded the upper threshold. In this case Route Design needs to be notified, via the Predicted Usage Model, so that it can invoke Bandwidth Distribution with new bandwidths. Bandwidth Distribution is not activated by the Current Load Model this case.
- *Scenario 4*. This is an abnormal situation in which the effective used bandwidth has exceeded the upper threshold but not the Route Design requested bandwidth. In this case, Bandwidth Distribution is invoked by the Current Load Model via an event. Bandwidth Distribution subsequently re-distributes the VPC bandwidth and determines new upper thresholds.
- *Scenario 5*. This is an abnormal situation in which the effective used bandwidth has exceeded both the Route Design requested bandwidth and the upper threshold. In this case Bandwidth Distribution is invoked by the Current Load Model via an event. Bandwidth Distribution subsequently re-distributes the VPC bandwidth and determines new upper thresholds. Route Design should also be invoked via the Predicted Usage Model so that it can invoke Bandwidth Distribution with new bandwidths.
- *Scenario 6*. The action taken in this case is the same as Scenario 5.

5.6.4 Information modelling

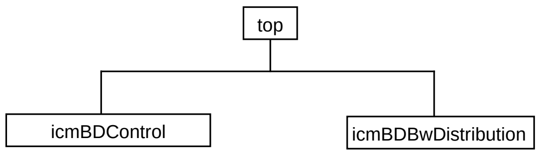

Figure 5.15 Bandwidth Distribution inheritance hierarchy

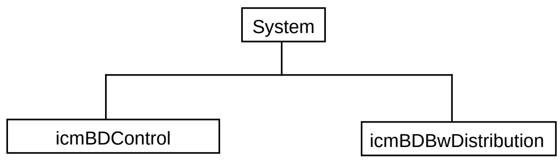

Figure 5.16 Bandwidth Distribution containment schema

5.7 Load Balancing

The scope and functionality of the Load Balancing OSF has been presented in Section 5.4.3.3. This section proposes specific algorithms for realising its functionality.

5.7.1 Overview

5.7.1.1 Load Balancing and Routing Management

As mentioned in Chapter 2, where the basic principles of ATM were presented, all possible routes for a given (s-d) pair and particular CoS are downloaded by the TMN to the network switches, where the actual routing decisions are taken at call set-up time. Route selection is done by means of a Route Selection Algorithm (RSA). Without loss of generality, it is assumed that RSAs operate on the basis of parameters (route selection parameters) associated with the available routes. Following the ideas on the taxonomy of routing algorithms [5.43], [5.68], several types of RSAs can be distinguished according to the selection method they employ, the information they utilise and the degree of adaptivity they offer.

According to the selection method employed, a Route Selection Algorithm can be:

- *Deterministic*, whereby route selection is made according to a predefined order. In this case, a priority is assigned to each alternate route and the routes of higher priority are selected first.
- *Random*, whereby route selection is made based on probabilistic criteria. Each route is assigned a probability or a frequency and the selections are made so that the frequency is guaranteed.
- *Locally Adaptive*, whereby route selection is made based on a policy taking into account the current load on the VPCs, as seen locally (e.g. select the least loaded VPC).

A Route Selection Algorithm may base its decisions upon purely local information, network-wide information, or upon no information at all. According to its degree of adaptivity (rate at which used information is renewed), a Route Selection Algorithm can be: static (not adaptive at all) or dynamic. Another parameter associated with adaptivity is how adaptivity is provided. It can be provided through: inter-node exchange, periodically or at exception; locally at connection acceptance/release times; or from the TMN periodically or at exception.

Examples of random, dynamic Route Selection Algorithms which do not require information about the network load are the Dynamic Alternate Routing (DAR), Linear Reward Penalty and Linear Reward Inaction algorithms, proposed for telephone traffic routing [5.65], [5.66]. An example of dynamic deterministic algorithm is the DNHR algorithm used by AT&T long distance telephone network [5.2].

The adaptivity of RSAs should not be confused with the quasi-adaptive nature of the Routing Plan. The Routing Plan has been constructed on the basis of predicted network usage; it is redefined whenever significant changes in network predictions are verified. The adaptivity of RSAs, on the other hand, is *within* the timeframe of network usage predictions, when the Routing Plan is stable. Such adaptivity is desirable since it compensates for inaccuracies in traffic predictions and/or network usage fluctuations around the predicted values. It can be argued that adaptive RSAs have the potential of responding efficiently to varied load patterns and network resource failures; thus improving network performance.

The above analysis indicates that there is scope for RSA management and proposes that the issue of routing management encompasses two levels: a higher level for the management of the Routing Plan and a lower level for the management of RSAs. Within this framework the Load Balancing OSF undertakes the task of managing RSAs, assuming a given VPC network and a specified Routing Plan.

It should be stressed that the introduction of the Load Balancing OSF, does not make the RSAs obsolete nor does it imply that the management plane is involved in routing decisions at call set-up times. On the contrary, it enhances the RSAs by conveying to them network-wide information. By placing the Load Balancing functionality in the management plane, the network elements are relieved from the burden of implementing intelligent RSAs which not only impose complexity on the network elements but also require inter-node exchange mechanisms to make information about network-wide conditions available at individual nodes. Therefore, the required routing intelligence in the network switches is reduced, resulting in faster routing decisions - an essential target of future broadband networks.

The Load Balancing OSF introduces a semi-dynamic routing policy, combining the merits of centralised and decentralised routing policies. Semi-dynamic routing policies have been utilised in traditional data networks and it has been shown that they improve network performance [5.43], [5.44].

There is a significant research in the area of network routing and the problem of RSA management has been tackled in the overall context of routing algorithms (e.g. [5.4], [5.64], [5.22], [5.65], [5.1], [5.43], [5.68], [5.3] and [5.44]). However, the majority of these studies do not take into account the different bandwidth and performance requirements of the multi-class network environment. Moreover, these studies do not address the issue of RSA management in the overall context of network management and they do not offer a clear distinction between management and control plane functionality.

5.7.1.2 Benefits

The Load Balancing OSF contributes to the efficient operation of networks from several aspects which in turn further justify its existence. Through its actions Load Balancing makes routing decisions network-state adaptive. Network-state adaptive routing has been recognised as a useful merit of routing algorithms as it is proved by the huge quantity of literature in the subject; indeed, network performance improvement has been verified under adaptive routing [5.64], [5.66], [5.43], [5.68]. Moreover, through RSA management, distribution of network load may be regulated; therefore enabling network load balancing. Balanced networks have been widely accepted as a valid objective of network design and routing policies [5.22], [5.67]; they allow better utilisation of free bandwidth across the network, they are better able to withstand variations in offered traffic and they ensure that network availability for new connections is as even as possible over the whole network. Apart from its active role in routing management, Load Balancing also contributes to preventive management. By taking a future perspective, it notifies the Route Design OSF of undesirable trends in network availability. Thus, appropriate actions to increase network availability may be taken before the network availability deteriorates below acceptable levels.

5.7.2 The Load Balancing algorithm

5.7.2.1 The route selection management algorithm

The essence of the proposed algorithm is to assign a figure of merit to each route, and to influence the RSAs so that traffic is routed over those routes with higher figures of merit. This view is in accordance with the traditional view where routing schemes are variants of shortest path algorithms [5.68]. In connectionless data networks, route merit is usually a function of the delay imposed by a route. However, in connection-oriented networks, such as ATM, the figure of merit should refer to the potentiality of the route to accommodate new connections. The figure of merit should be a function of spare capacity and it should take into account the fact that different CoSs may share part of, or all of, the routes.

Adopting the above approach, *route potentiality* is calculated for all possible routes between a given source and destination node, for each CoS.

The routes available at a node are defined in terms of route selection entries associating a particular network destination and CoS with a VPC starting from the node. Therefore, route selection in fact refers to the selection of a particular VPC. A VPC may belong to more than one route, and at a given node all these routes will use the VPC as an exit from this node. The potentialities of all these routes can therefore be accumulated, giving rise to a figure of merit of selecting this VPC as the next step in the route. The figure of merit of VPC selection reflects the potentiality of the network to accommodate new connections in the route(s) originating at this VPC.

The VPCs at each node are therefore graded with a figure of merit, *VPC selection potential*. The algorithm then recommends VPCs for routing according to their selection potential. This is achieved by setting appropriate route selection parameters so that VPCs with a higher figure of merit have advantage over those with lower figures of merit. In the case of deterministic RSAs, VPCs are prioritised in the order of their merit; and RSAs make selections according to this order. In the case of random RSAs, VPCs with higher figures of merit are assigned higher frequencies. In the case of locally adaptive RSAs, VPCs are classified into equivalent groups according to the significance of the differences in their figures of merit; VPC selection is done in the order of the equivalent sets and by applying local criteria for the VPCs within a set. The latter routing policy enhances the concept of δ-routing [5.43] proposed for data networks.

It should be noted that the proposed algorithm is not simply a widest path routing algorithm trying to route traffic over routes with the highest potentiality. It is a highest potentiality (HP) path routing algorithm, trying to achieve routing over the network part(s) that have the highest potentiality to accommodate new connections. Figure 5.17 outlines this point. Under widest path routing, VPC A corresponding to route R1 should be selected for traffic between nodes 1 and 6. Under highest potentiality path routing, VPC B, corresponding to route R2 or R3, is selected. Therefore, under highest potentiality path routing, full advantage of route alternatibility, not only locally (at the vicinity of a node) but also remotely, is taken. In this sense highest potentiality path routing outperforms widest path routing.

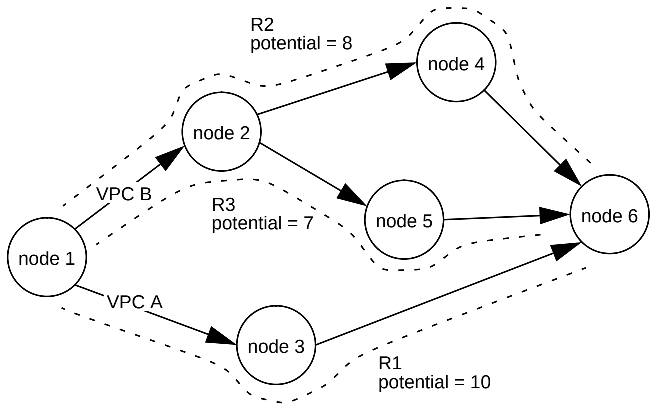

Figure 5.17 Widest path vs. HP path routing.

In the following, we formulate the notions of route potentiality and VPC selection potential introduced previously and propose specific formulae for their calculation. Route potentiality is defined in terms of *VPC acceptance potential* which is introduced next.

For a given VPC, say v, VPC acceptance potential, denoted by $VPt_a(v; c)$ for CoS c, is defined as the number of VCCs of this CoS that can be potentially accommodated in the VPC, taking into account the current load on the VPC. The fact that a number of CoSs may share the same VPC in their routes needs also be taken into account. The following heuristic is proposed for their calculation.

Considering a VPC, the $VPt_a(.)$s are calculated as solutions of the following linear system with respect to k(.)s:

$$\sum_{i=1}^{C} \delta(i) B(i) k(i) = S \tag{1}$$

$$\frac{B(i) k(i)}{B(j) k(j)} = \frac{g(i)}{g(j)}, \; i \;\; j, \; i, j = 1 \ldots C, \; \delta(i) = \delta(j) = 1 \tag{2}$$

where:

$\delta(i)$ is a Boolean taking the value 1 if CoS i uses the VPC in its routes, and the value 0 otherwise.

B(i) is an estimate of the VPC bandwidth that CoS i will consume when it is accepted on the VPC. It can be the mean or the peak bandwidth requirement of CoS i, or it can be its effective bandwidth as calculated by the CAC network algorithm.

S is an estimate of the VPC spare bandwidth. It is recommended to be in the form of a moving average and not in the form of instantaneous value for reducing sensitivity to traffic fluctuations and increasing estimate accuracy.

C is the number of the different CoSs supported by the network.

g(i) are weights differentiating the access of each CoS on the VPC. They reflect the frequency with which CoS i is using the VPC.

The weights g(.)s are closely related to the Routing Plan. They can be calculated either dynamically, from network measures, or directly from the Routing Plan. In the latter case, the following is proposed:

$$g(i) = \sum_{j=1}^{O(i)} \frac{1}{a(j)} \qquad (3)$$

where:

O(i) is the number of selection occasions on which CoS i may select the VPC. Note that for a given CoS, a particular VPC may belong to routes to one or more destinations. Therefore, O(i) equals the number of possible destinations that CoS i may reach through the given VPC.

a(j) is the alternatibility factor i.e. the number of routing alternatives for routing occasion j.

Equation (1) is intuitively evident, taking into account that a VPC can accommodate connections of different classes. Equation (2) says that the ratio of the bandwidth to be consumed by two different CoSs is taken to be proportional to the visit ratio of these CoSs to the VPC; note that the product B(i)k(i) is the amount of VPC's bandwidth to be given away to CoS i connections.

The system of the equations (1), (2) yields the following solution:

$$VPt_a(v;i) = \frac{g(i)}{B(i)} \cdot \frac{S}{\sum_{j=1}^{C} g(j)}, \; i = 1 \ldots C \qquad (4)$$

As it can be seen from (4), the VPC acceptance potential for a CoS depends on:

- the bandwidth characteristics of the CoSs,
- the spare bandwidth of the VPC,
- the alternatibility with which the VPC is used for routing.

Note that because the alternatibility with which a CoS uses a VPC has been taken into account, a CoS with alternative routes is discouraged from occupying a VPC at the expense of the CoSs that use that VPC as a unique option.

Having defined the notion of VPC acceptance potential, the notion of route potentiality is defined next.

For a given route, say r, route potentiality, denoted by RPt(r; c) for CoS c, is defined as the number of VCCs of this CoS that can be potentially established in the route, taking into account the current load of the VPCs along the route. The fact that a number of

CoSs may share parts of the route needs also be taken into account. Route potentiality is defined in terms of VPC acceptance potential as follows:

$$RPt(r;c) = \min\{VPt_a(v;c)\ \forall (v \in V(r))\} \qquad (5)$$

where V(r) denotes the set of VPCs that constitute route r.

Finally, the notion of VPC selection potential is defined.

Considering a VPC, say v, starting at a specific network node, say n, VPC selection potential, denoted by $VPt_s(v; c,d)$ for CoS c and destination node d, is defined as the number of VCCs that can be potentially established in all possible routes to the particular destination starting with this VPC. It can be defined as follows:

$$VPt_s(v;c,d) = \min\left\{VPt_a(v;c), \sum_{r \in R(n,d;c)} RPt(r;c)\right\} \qquad (6)$$

where R(n,d;c) denotes the set of all routes from node n to destination node d defined for CoS c.

5.7.2.2 Network load surveillance

Taking a future perspective, Load Balancing monitors network load with the purpose to notify the management functions responsible for the definition of the Routing Plan of deterioration in network availability for new connections and significant deviations in link load - which might indicate inefficient use of the physical resources. Based on the analysis presented in the previous section, network availability for new connections can be estimated by extending the notion of potentiality at the node level.

For a given network node, say n, Node Potentiality, denoted by NPt(n; c,d) for CoS c and destination node d, is defined as the number of VCCs that can be potentially established from node n to node d over all possible routes starting at node n, taking into account the load in the network i.e.

$$NPt(n;c,d) = \sum_{v \in V(n)} VPt_s(v;c,d) \qquad (7)$$

where V(n) denotes the set of VPCs starting from node n.

Considering access nodes, the above formula provides a measure for network availability for specific source-destination pairs and CoSs.

Load deviations at the link level can be measured by calculating the difference of link utilisation from the network-wide average value.

Warnings are emitted as threshold crossings events. The threshold values as well as the parameters of the measurements (e.g. moving average method, observation period) are regarded as operation parameters and are defined by the Route Design OSF.

5.7.2.3 The Load Balancing algorithm

Figure 5.18 summarises the previous algorithms, offering a complete view of the Load Balancing algorithm. The algorithm may be activated periodically or at exception by threshold crossing events related to network usage. Moreover it is triggered whenever the Route Design OSF updates the VPC infrastructure and the sets of routes.

- For each VPC get its current load and calculate its acceptance potential for each CoS (cf. (4)).
- For each CoS consider its network of routes and view it as a (directed) graph.
  - For each node from which there are more than one VPC, say there are k alternatives, k>1
    - Find the k-HP paths restricted to have different first VPCs i.e. calculate the VPC selection potential, for each destination (cf. (6)). An algorithm for establishing HP routes is presented in the next section.
    - Grade the VPCs and determine the values of the associated route selection parameters; if they differ from the ones determined in the previous invocation time, send appropriate management actions.
    - Calculate node (access or transit) potentiality measure (cf. (7)) and emit appropriate alarms if necessary.
- Measure link utilisation, determine deviations around the network-wide average and if necessary emit alarm.

Figure 5.18 The Load Balancing algorithm

5.7.3 Establishing HP routes

A centralised algorithm for establishing HP routes between all source-destination pairs of a directed graph is presented in this section. This algorithm is used by Load Balancing for grading the VPCs out of each node (see Figure 5.18). The algorithm is based on the well-known shortest path algorithms of Dijkstra and Floyd [5.63]. The notation introduced earlier is preserved. For simplicity, the CoS dependency is dropped. Note, from Figure 5.18, that the algorithm runs for each CoS separately; CoS multiplexing is taken into account while calculating the measures related to VPC and route potentialities.

The essence of the algorithm is as follows: starting from a given destination node, remote nodes are added according to their distance (in number of VPCs) from the destination node. A node is said to be l-level predecessor if its distance from the destination node is l-VPCs long. The VPC selection potentials are continuously updated as the algorithm progresses level (distance). Indeed, the VPCs starting from the newly added nodes are updated according to the updates of node potentiality made in the previous level. The algorithm terminates if no more nodes can be added and it is repeated for all destination nodes. Figure 5.19 describes the HP algorithm in pseudo-code.

5.7.4 Design issues

5.7.4.1 Discussion

The Load Balancing OSF is concerned with the management of network entities, relating information from a number of network elements, therefore it has been placed at the network management layer following the directives implied by the decomposition of the logical TMN architecture. Alternatively, it could be placed at the network element management layer, requiring the existence of a distributed algorithm for finding HP

```
For every destination node, d {
        Initialise: NPt(n)=0 for every node n, VPt_s(v)=0 for every VPC v, l = 1 //the level
                SetOfNodes(1) = {}
                //it keeps the nodes to be considered in the next level;
                // they are kept in the form:
                // <n, prevn>: n is the node to be considered,
                // prevn: is the node that put node n into the current level
        //Deal with 1st level predecessors
        For every node n: predecessor of d
                For every route from n to d, say corresponding to VPC v
                        VPt_s(v) = VPt_a(v),
                        NPt(n) += VPt_s(v).
        //Find the 2nd level predecessors
        l = 2
        For every node n: predecessor of d: n≠d. Add <n,d> to SetOfNodes(l)
        //Deal with the general level: l
        while SetOfNodes(l) ≠ {} {// l is the current level
                For every <n,pn> in SetOfNodes(l)
                        For every route from n to d, say corresponding to VPC v
                                VPt_s(v) = min [VPt_a(v), NPt(pn)]
                                NPt(n) += VPt_s(v)
                        // Determine the (l+1)-level nodes
                        For every node prevn: predecessor of n
                                if <prevn, n> ∉ SetOfNodes(l)
                                        Add <prevn, n> to SetOfNodes(l+1)
                l += 1
        } //while
} //for d
```

Figure 5.19 The HP algorithm

routes. In this case, the interactions between the network and network element management layers for conveying network measures would be substituted by the interactions between the network element management layers for the exchange of the information required by the distributed algorithm. Additionally in this case, the interactions with the other components of the network management layer (e.g. Route Design) as well as the cost of meta-management (e.g. software maintenance) would be increased. The magnitude of these trade-offs depends on the TMN transmission infrastructure and the physical location of the TMN hosts. The decision as to which architectural option to choose is therefore left open to the TMN system designers.

It should be stressed that the choice of the distributed solution should not be taken as an implication that the Load Balancing functionality belongs to the control plane nor that it should be embedded in the network elements. By placing the Load Balancing functionality in the management plane, using a hierarchical TMN architecture the intelligence required for the efficient operation of the network is distributed over nodes other than the actual network elements.

5.7.4.2 Information modelling

The icmLBLoadBalancer object class comprises the interface of the Load Balancing OS. It enables other OSs to view and if necessary to modify (e.g. lock) its state and it is used to receive the required actions. Moreover, it schedules the periods of the Load Balancing activation. The inheritance hierarchy is shown in Figure 5.20 and the containment schema is shown in Figure 5.21.

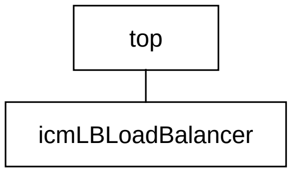

Figure 5.20 Load Balancing inheritance hierarchy

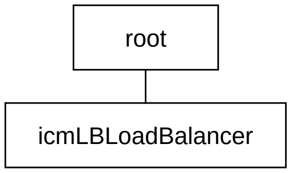

Figure 5.21 Load Balancing containment schema

5.8 Performance Verification

5.8.1 Overview

The Performance Verification OSF is needed because the network supports multiple CoSs with a guaranteed performance. For networks offering connections without a guaranteed performance Performance Verification, and performance management in general, is limited to performance monitoring.

The Performance Verification OSF does not directly manage the network. It monitors the network to quantify its performance, and indirectly the performance of the management system, being an indisputable measure its efficiency.

This section outlines a specific approach and suitable algorithms for performance monitoring and network performance verification to support the goals of the Performance Verification OSF. The approach and algorithms are built around the rich and powerful features of OSI systems management, fully exploiting the hierarchical nature of the TMN, to achieve sophisticated performance monitoring and performance verification without imposing a large communications overhead in the TMN and hence in the managed network itself. Specifically, by pushing the monitoring functionality down to the NEs, polling from the network management layer is avoided; the Performance Verification OSF residing in the network management layer receives only the emitted threshold crossing notifications indicating unacceptable network performance.

Requirements on supporting monitoring objects are identified as well as enhancements on the 'classical' OSI metric and summarisation monitoring objects [5.61] [5.62].

5.8.2 Functional aspects and approach

The Performance Verification OSF:

- evaluates and verifies network performance with the purpose of notifying the Route Design OSF of unacceptable network performance, and,
- analyses customer complaints with regard to network performance.

Evaluation and verification of network performance is done on the basis of measurements concerning the performance of the CoSs that the network supports. It is assumed that the following parameters define the performance of the CoSs:

- rejection ratio (blocking probability),
- cell delay,
- cell loss ratio,
- jitter.

The above parameters are those which are directly influenced by the activities of the VPC and Routing Management Service. Other performance parameters may be associated with network CoSs, such as set-up delay; but because the role of Performance Verification is to feed back performance analysis results to the performance management system, to influence its future behaviour, these parameters are outside the scope of Performance Verification in this context.

The above parameters have been widely accepted as meaningful connection performance parameters and they are in accordance with the performance parameters defined by ATM Forum [5.23]. It is assumed that for each CoS there is an upper bound (performance target) defining the range of acceptable values. Network performance is within acceptable levels if these bounds are preserved within certain confidence levels.

Measurements of rejection ratio statistics can be made directly from appropriate raw data (e.g. counters of connection requests and rejections) available at the management interfaces of the network elements. The measurement of cell related statistics (delay, loss ratio, jitter) requires the existence of measurement instruments at the source and destination end-points or the existence of OAM (operation and maintenance) capabilities at access switches including appropriate management interfaces.

Based on measurements on the above parameters, Performance Verification evaluates network performance and by comparing it with the maximum allowable values (the performance targets) verifies whether or not network performance is within acceptable levels. Evaluation and verification should be done in the scope of the total population of CoSs and source-destination pairs. Therefore the measurements should cover all (or a representative portion of) the source-destination pairs and supported CoSs, to increase the validity of the produced results. On the other hand, the verification process should introduce minimum overhead to the management system for collecting the required network information and it should not be sensitive to transitory situations.

For a given CoS, network performance evaluation and verification could be done in two modes:

- Per source-destination (s-d) pair, or,
- Network-wide (statistically averaged over all possible s-d pairs by sampling).

Network-wide performance measurements could be taken either exhaustively (over all possible s-d pairs) or by adopting sampling techniques in cases where the number of all s-d pairs is huge.

The proposed approach to network performance verification adopts the first mode, since it is considered more useful to the VPC and Routing management functions. Indeed, the routing management systems construct and manage routes of guaranteed quality for all source-destination pairs for a given CoS; therefore, as a means for quantifying routing management efficiency, network performance estimates per s-d pair are required. Network-wide performance estimates may hide unacceptable performance situations for some s-d pairs. Moreover, network-wide performance estimates inevitably require the use of polling for retrieving the necessary sample values which subsequently creates a communications overhead in the management system. As it will be shown in the following sections, network performance verification per CoS and per s-d pair, paradoxically enough, does not require as great a management overhead as in the network-wide case; in fact it creates minimal management overhead.

For a specific s-d pair and a given CoS the essence of the proposed approach is to calculate the following probabilities:

$$\mathrm{Prob}[{}^{(i)}R_{sd} <= {}^{(i)}T_r] >= A \qquad (8)$$

$$\mathrm{Prob}[{}^{(i)}R_{sd} > {}^{(i)}T_r] < 1 - A \qquad (9)$$

where:

${}^{(i)}R_{sd}$ denotes a measurement of a CoS related performance parameter R (rejection ratio, cell delay, cell loss ratio, jitter) for CoS i and source destination pair s-d. For rejection ratio, ${}^{(i)}R_{sd}$ is calculated as an instantaneous value or as a moving average estimate. For the cell related performance parameters (cell delay, cell loss ratio, jitter), ${}^{(i)}R_{sd}$ is calculated as the arithmetic average over a specific number of connections.

${}^{(i)}T_r$ is the maximum allowable value of performance parameter R for CoS i.

A is the confidence level.

Formula (8) says that it is almost certain (with a confidence level A) that network performance, with regard to the performance parameter R, for connections of CoS i between the source-destination pair s-d, will be within the acceptable levels specified for that CoS. Violation of condition (8) or (9) means that network performance has fallen below acceptable levels. In such cases performance alarms should be raised to trigger the necessary corrective actions.

The calculation of the above probabilities can be done continuously, or during specific verification periods, with dynamic duration and inter-period times, depending on network state.

Note that by taking the probability of the event ER = $[{}^{(i)}R_{sd} \geq {}^{(i)}T_r]$, transitory fluctuations in network performance are not taken into account. The measurement of the probability is taken at a given window, within the verification interval; then the probability is approximated by the frequency (ratio) of the times of occurrences of the event

ER over all the observations that were made within this window. The choice of the length of the probability window should be made by taking into account the nature of the measurements (instantaneous values or moving average type of values or number of connections over which the cell related performance parameters are calculated) for the performance parameter R. These are left as design options.

Figure 5.22 summarises the proposed approach outlining the main steps of an algorithm fulfilling the objectives of the Performance Verification OSF.

Retrieve performance targets ${}^{(i)}T_r$ for all CoS related performance parameters R.

- For all access source nodes
  - For all CoSs, determine destination nodes over which network performance will be verified
    - For each CoS related performance parameter, initiate appropriate monitoring activities for the selected s-d pairs for monitoring the probability of unacceptable performance (see (9)).
    - Collect threshold crossing notifications.
      - If one received, wait for a specific time period and collect any other notification that may come within this time period. At the end of the time period send the collected notifications (network unacceptable performance alarms) to the interested Management Services.

Schedule the next verification interval.

Figure 5.22 Performance Verification functionality

It is worth noting that the Performance Verification OSF operating as above creates minimal overhead into the management system. As it will be shown in the following section, all the required monitoring activities may be delegated down to the Q-Adaptor level providing the management interface of the network elements. Only the notifications resulting from threshold crossings are forwarded to the management system. The number of notifications depends on the performance of the network and on the sensitivity of the measures. By appropriately regulating the measurement characteristics of the connection related performance parameters and the probability window, the trade-off between the validity of the measurements and their sensitivity can be managed.

5.8.3 Design aspects

Performance Verification is concerned with collecting, collating and comparing performance data obtained from more than one network element. For these reasons, it requires a global view of the network, and therefore must be positioned at the Network Management Layer. However, the performance monitoring functionality is distributed in a hierarchical fashion over the network management and element management layers. By virtue of the proposed hierarchical structure, the management overhead for acquiring the required network statistics may be minimised, as monitoring activities can be delegated down the hierarchy as close as possible to the network elements themselves where the raw performance data is generated.

By following the TMN approach which uses OSI systems management concepts, performance monitoring is achieved by virtue of the OSI systems management functions (SMFs). In particular, event reporting [5.57], alarm reporting [5.58], log control [5.59], test management [5.60], workload monitoring [5.61] and measurement summarisation [5.62] SMFs are used. These are standard facilities of OSI systems management, pertinent to any Q_3 interface, and demonstrate the advantage of adopting the TMN approach for implementation. It is believed that these facilities provide a rich and powerful set of generic management tools, which - when properly used - may considerably reduce the cost of developing and of operating the management system.

The hierarchical TMN architecture and the OSI SMFs imply and furthermore facilitate distribution of the required monitoring activities over the network management, network element management and network element (management interface) layers. This distribution enables frequent data collection for a single network element to be carried out close to the source of the data. The results of monitoring activities in the network elements are then forwarded to the higher management layers either on request or at exception, at threshold crossing instances. This design consideration ensures that the management communications overhead is as small as possible as the bulk of the data in the management plane will be transferred locally.

Furthermore, by adopting the per s-d verification approach, rather than the network-wide one, decomposition at the functional level is achieved, in the sense that individual network element performance measures do not need to be further summarised in higher management layers. By extending the measurement summarisation functions to include probability calculations, as described previously, the calculation of the required performance measures (see (8) and (9)) can indeed take place at the network element (management interface) level, incurring no polling cost to the management system whatsoever.

Whenever the Performance Verification OSF requests a particular performance measure, a monitoring activity is created in the Current Load Model OSF. In turn, the Current Load Model delegates element level monitoring activities to the network element layer. The monitoring activities and data retrieval between the Performance Verification OSF and the Current Load Model OSFs is achieved by the creation of monitoring activities in the form of metric and summarisation objects. In turn, the interaction between the Current Load Model OSF and the underlying OSFs/MFs/QAFs/NEFs is achieved by the same mechanisms. Figure 5.23 illustrates the proposed design approach.

Only if the network elements themselves do not support the required SMFs is synchronous polling required between the lowest level management functions and the network elements. This means that high load communications inherent in polling mechanisms is limited to local area communications, reducing the load in the rest of the TMN and the underlying managed network.

5.8.3.1 Information model

The icmPVPerformanceVerifier object class comprises the interface of the Performance Verification OS. It sends appropriate actions to the Route Design OS and enables other OSs to view and if necessary to modify (lock) its state and operational parameters. The inheritance hierarchy is shown in Figure 5.24 and the containment schema is shown in Figure 5.25.

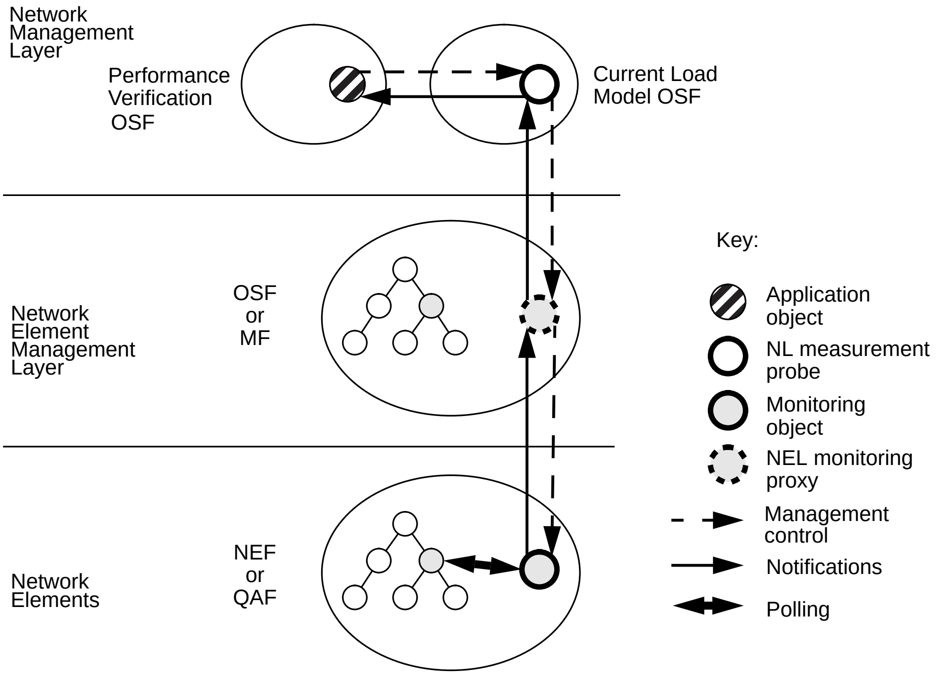

Figure 5.23 Performance Verification management system design

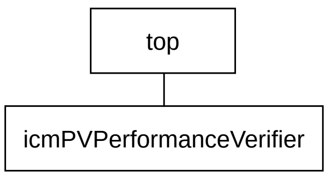

Figure 5.24 Performance Verification inheritance hierarchy

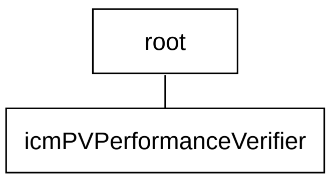

Figure 5.25 Performance Verification containment schema

5.9 Predicted Usage Model

5.9.1 Overview

According to the objectives detailed in Section 5.4.3.5, the Predicted Usage Model OSF (PUM) provides:

- A data capture framework for modelling user demand on network resources in terms of call traffic variables.
- A trend detection framework for identifying a model of traffic intensity over past data.

The Predicted Usage Model collects measured statistics on call patterns within a network and then establishes past traffic trends by carrying out regression analysis on measured traffic data. Based on past trends, the Predicted Usage Model forecasts likely future traffic trends. Information on expected future trend is provided to the Route Design OSF. Measured traffic statistics is provided by the Current Load Model OSF.

The generic traffic modelling functions of the Predicted Usage Model are well recorded within the narrow-band ISDN standards ([5.47], [5.48], and [5.49]) under the following topics:

- measurement and recording of traffic,
- estimation of traffic offered in the network, and,
- forecasting of traffic.

The concepts described in these standards apply equally well to broadband ISDNs. In this case, there is a need to take into consideration the specialised nature of broadband ISDN's User Demand Modelling. An overview of these functions are described in the following sub-sections. Standardisation of user demand modelling in broadband ISDN is currently in progress [5.50].

5.9.2 User demand modelling

This section outlines the basic definitions and concepts of ATM traffic engineering.

5.9.2.1 Traffic parameters

The term user demand means the demand of the users for telecommunication services through the network to satisfy the users' information transfer needs. Call demand is the basic manifestation of the user demand at the UNI. A call demand consists of sequences of call attempts made by a user or by its customer premises equipment and the subsequent call, if it is finally established.

Call attributes are those attributes of a call demand that identify resources needed by the call demand from the network. Such resources include both user-plane resources (e.g. transmission paths) and control plane resources (e.g. signalling capabilities).

Call pattern is defined in terms of events at the UNI, and the times between these events.

The combination of call attributes and call patterns define measures for quantifying user demand.

Call attributes include:

- establishment of communication,
  - control plane, management plane,
  - semi-permanent, demand, reservation,
- communication configuration,
  - point-to-point, multi-point, broadcast,
  - number of points and their locations,
- number of ATM connection,
  - number in each direction between each pair of points,
- use of VCCs or end-to-end VPCs (when end-to-end VPCs are provided, calls within such VPCs are not defined for traffic engineering purposes),
- traffic contract for ATM connections,
  - source traffic descriptor,
  - QoS descriptors.

In defining call traffic variables, an event is the transmission across the UNI corresponding to call set-up, renegotiation or release signals. The fixed attribute traffic variables describe the call attempt arrival process in terms of:

- mean number of re-attempts in case of non-completion (this measure needs to be estimated when actual measurements are not available),
- mean time between call attempts,
- mean total holding time of the call.

5.9.2.2 Measurement and recording of traffic

5.9.2.2.1 Overview of traffic measurement

Two main kinds of measurement are required in forecasting traffic:

- measurement of the amount of traffic carried. This is an average value of source destination traffic over a certain period of time.
- measurement of the number of bids: This is a count of entities denoting events, e.g. calls accepted, during a certain period of time.

Measurements are taken continuously during the day, the set of recorded days during a given period of time is referred to as measurement days. The large recurrent frame for taking measurements in the analysis of traffic trends is yearly.

A traffic profile is defined to be stable when the individual daily traffic profiles do not differ significantly.

There are different policies for obtaining traffic measurements:

- *Yearly continuous measurements*: the measurement days are post-selected from a base period with a length of the whole year. The post selected days include the peak intensity values measured during the base period.
- *Yearly non-continuous measurement*: the measurement days are scheduled (pre selected) from a base period of a few months. The pre-selected days include expected high load days established from expectation analysis (regression) or from earlier observations.

- *Busy hour traffic*: an average traffic intensity (or bids) for the same hour period averaged over several days. Within the busy hour, traffic is considered to be stationary and thus the recorded intensity is the mean value during the busy hour.
- *Time Consistent Busy Hour* (TCBH): The recommended standard method of calculating the daily average requires continuously measuring all quarter hours for all days concerned and selecting the busiest hour in the average profile for all days. This method is valuable in situations of stable traffic profiles. The daily continuous measurements provide the data necessary for confirming profile stability.
- *Average Daily Peak Hour* (ADPH): In this case, the daily average involves continuously measuring all quarter hours, but only the busy hour of each day is retained for averaging. The advantages of ADPH are that it requires less data storage and manipulation than TCBH and that it gives a more representative value in the situation of unstable traffic profiles.
- *Fixed Daily Measurement Period* (FDMP) or Fixed Daily Measurement Hour (FDMH): In this case, the administration concerned do not measure the traffic continuously over the day, but only for the hour or few hours expected to be busiest. The advantage of FDMP is that it requires less measurement resources than TCBH or ADPH. The disadvantage is that in individual situations, the difference between FDMP and TCBH results may vary widely.

5.9.2.2.2 Non-continuous measurements

Within the scope of ICM, it is not reasonable to expect the Predicted Usage Model to capture data as yearly continuous measurements as the quantity data generated is too large for practical analysis. For this reason, the Predicted Usage Model operates by adopting a non-continuous measurement approach.

- Yearly non-continuous measurements

  A limited sample of days in each year is selected. Such sample measurement days would normally be working days. The mean (M) and the standard deviation (S) of the daily busy hour traffic loads are calculated. Load estimates are given by the relation:

  $L = M + k*S$

  where various values of k are used for normal and high load levels.

  The standard deviation would normally be calculated from measurements collected over a sample measurement days, using the mean of each day's time consistent busy hour (TCBH) traffic. For less than 30 measurement days, the standard formula for S would not yield reliable values hence special measurement values need to be provided for load estimation. This requirement for special estimate becomes mandatory if Fixed Daily Measurement Period (FDMP) is used is used instead of TCBH.
- Approximate estimation of TCBH from FDMP

  This method is effected by taking measurements over a fixed period of each day (e.g. 3 hours). This period should correspond to the highest part of the traffic profile, thus it is expected to include the TCBH. Measurement values are accumulated separately for each quarter hour, and the busiest hour is determined at the end of the measurement period. This method has been reported to give results which are good approximations to those of TCBH traffic level when the time of fixed daily measurement is defined for every single circuit group (or VPCs), (see Rec E.500 [5.47]).

Stability of traffic profiles need to be confirmed several times over days of the yearly measurement days.

- Summary on adjustments to measured traffic

  The fixed daily measurement period (FDMP) is adopted for the Predicted Usage Model. Two statistical analysis procedures need to be carried out on the measured data:
  - stability of traffic profiles,
  - variation factor for mean loads.

  Stability analysis of traffic profiles is carried out to ensure that FDMP is suitable instead of the more rigorous TCBH. This analysis should be sensitive to the condition that the actual load level varies over measurement days. Stability analysis is aimed at justification of the selected time window for FDMP.

  Variation factor for mean loads is aimed at ensuring that the estimated load is within reasonable value of the mean irrespective of the variation of changes in magnitude of offered traffic over measurement days.

5.9.2.3 Forecasting of traffic

The Predicted Usage Model OSF maintains a recurrent pattern of call demands. The information consists of a structured set of data which represents call patterns over recurrent time periods. The Predicted Usage Model OSF receives updates of patterns of bids issued by calls in terms of call traffic over source destination nodes. These updates are provided by the Current Load Model OSF.

Figure 5.26 illustrates an example three day pattern of daily traffic intensity. Sampling windows are selected by trial and error on actual network data in order to capture faithfully a traffic profile such as that shown in the example, e.g. three distinct seg-

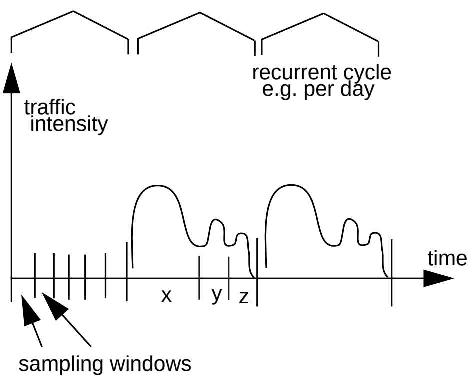

Figure 5.26 Recurrent pattern of traffic intensity

ments x, y, and z per day. The granularity of sampling windows is left as a design option.

Given an expected pattern, it is the function of the Predicted Usage Model to perform statistical analysis on a sequence of pattern data to establish stability. Thus, the pattern recurrence (e.g. daily) is analysed over a suitable cycle (e.g. a sequence of days over a year's measurement days) to confirm pattern stability. This analysis is based on computation of standard deviation, as outlined in Rec E.500 [5.47].

Assuming that there is a stable pattern within a stability analysis cycle, the goal of the Predicted Usage Model is to identify a model for trends of a pattern's change over a time horizon. This time horizon is normally larger than a stability cycle, say six stability cycles. A trend model could be linear, exponential, etc. It is established by looking at past data through curve fitting and correlation as outlined in Rec E.507 [5.49].

After establishing a trend model, a prediction of magnitude of future call patterns can be made by the Predicted Usage Model. Such predictions need to be validated against actual data to confirm the magnitude of predicted changes (Rec E.507 [5.49]).

Figure 5.27 illustrates the procedure for forecasting source-destination traffic (Rec E.507 [5.49]).

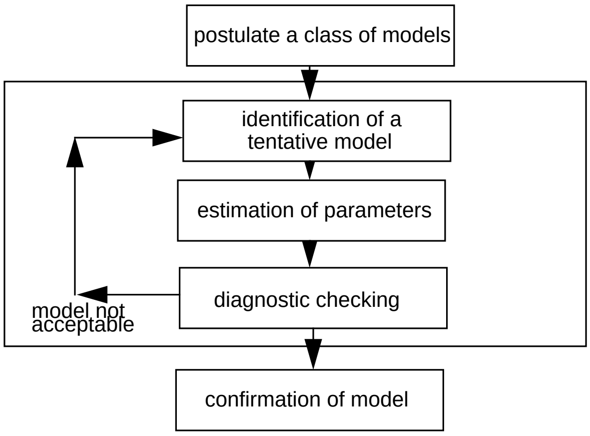

Figure 5.27 Trend identification and traffic intensity prediction

A number of forecasting models is selected as a class of value generators for stable pattern traffic. Such generators could be linear, parabolic, exponential, etc. [5.49]. The Predicted Usage Model OSF adopts linear and exponential models.

Given a number of models, one model is selected for a forecasting period, depending on its performance when applied to measured data. Model selection is performed by looping the three operations as shown in Figure 5.27: identification of tentative model, estimation of model parameters and diagnostic checking of model.

Correlation statistics and confidence intervals related to parameter uncertainty provide the model selection criteria [5.49]. Once a model has been selected, there is a need to confirm it by testing it on future data as opposed to identification through regression. This task involves computation of forecast error such as: mean error, mean percentage error, root mean square error, etc. [5.49].

5.9.3 Design approach

The Predicted Usage Model is part of a closed loop relating to OSFs in the higher levels of the VPC and Routing management hierarchy presented in previous sections. Based on monitored data from the Current Load Model the Predicted Usage Model makes predictions about the call traffic for Route Design.

The Predicted Usage Model works on two cycles:

- The polling cycle: the network data is monitored and analysed against the usage profile model. If the model does not fit then an alternative model is selected for the source/destination/cosId tuple and updated in the current database.
- The predictions cycle: the Predicted Usage Model interrogates the current database and if the data is different from the previous predicted data it sends the data to Route Design.

The Predicted Usage Model acts as a manager to the Route Design OSF in providing information on predicted changes to expected traffic pattern.

The Predicted Usage Model acts as a manager to the Current Load Model in periodically collecting measurement data. The Current Load Model provides the call traffic variables described previously. This information is provided per source-destination pair and CoS. The measurement period is based on Fixed Daily Measurement Period (FDMP) described earlier.

5.10 Configuration Manager

5.10.1 Function

The Configuration Manager (CM) maintains a model of the logical and physical network, together with any related configuration information (e.g. link capacities). It is assumed that the Configuration Manager is the *only* such database. Thus any changes to physical or logical topology or other configuration *must* be reflected in the Configuration Manager. As a consequence, other OSs in the TMN can rely on the Configuration Manager to provide an accurate picture of the current state of the network.

This implies that any manager-role activities to be carried out on the network are performed only by the Configuration Manager. Other OSFs must act through the Configuration Manager to perform such actions. For example, to change the bandwidth allocated to a particular VPC the value in the Configuration Manager must be changed. The Configuration Manager will then carry out any action required to pass this information down to the network.

Routing information is not stored in the Configuration Manager. Routing information in the form of routing table entries is kept in the NE management layer. Network-wide routing information is kept in the Route Design OSF.

5.10.2 Modelling approach

The information model makes extensive use of the generic objects defined in [5.51]. In particular, trail and connection classes and their corresponding termination point classes have been adapted to represent physical links and VPCs. The use of these classes is illuminated by [5.52], with an SDH implementation in [5.53]. In addition, a draft ETSI document [5.54] adapts trail and connection termination points to ATM at the NE layer.

At the network management layer it is only necessary to model physical links and VPCs, not VCCs. Using a layered network model, the links are servers to the VPCs. The VPC trails are made up of VPLs which are represented as objects of class "connection," defined in M.3100 [5.51].

New classes have been defined only where no suitable standard generic objects could be found. Although much work has been done in the standards bodies at the NE level for ATM, little seems to be available as yet for higher levels. Thus it is found that new classes for such functions as the CAC configuration need to be defined.

The containment and inheritance hierarchies are shown in Figure 5.28 and Figure 5.29.

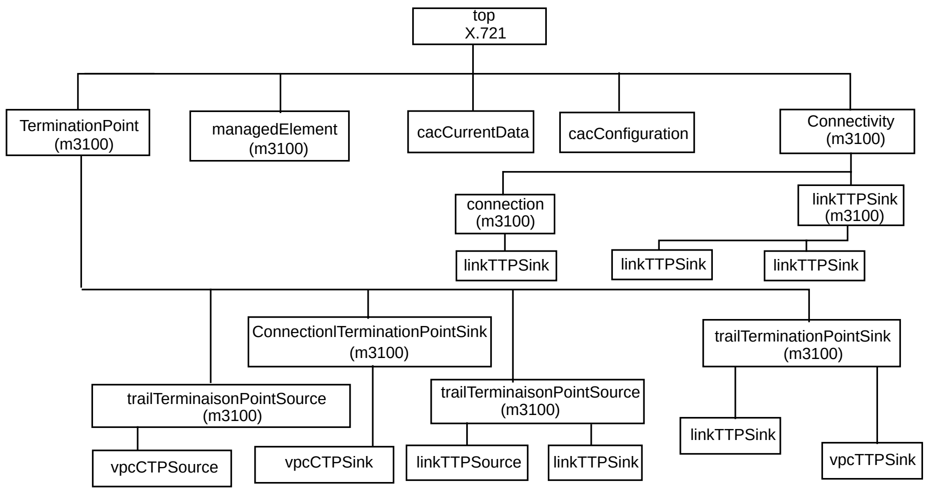

Figure 5.28 Configuration Manager inheritance hierarchy

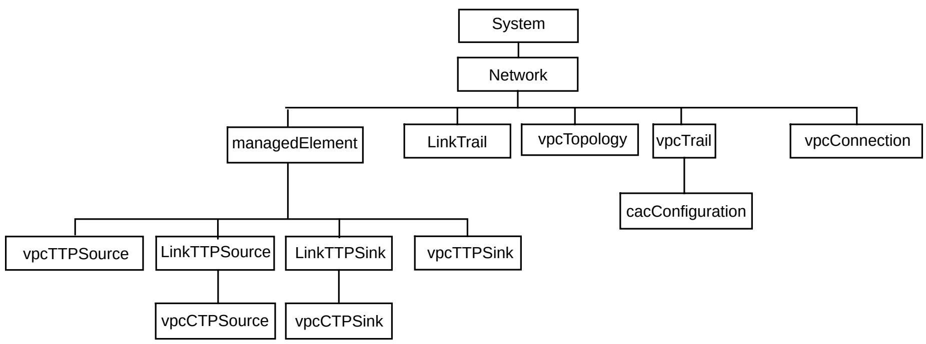

Figure 5.29 Configuration Manager containment schema

5.11 Current Load Model

5.11.1 Overview

The role of the Current Load Model OSF is to provide statistics on VPC usage, link usage and connection requests.

The approach is as follows: First, an OSF requests the creation of a monitoring activity. Then the Current Load Model interacts with the Configuration Manager and the NE layer OSFs to implement the requested monitoring activity. The requesting OSF is responsible for controlling the monitoring activity: it can retrieve the measured data, change the monitoring parameters, or delete the monitoring activity. The measured data can be retrieved either by polling or asynchronously provided that appropriate thresholds have been defined.

5.11.2 Design issues

5.11.2.1 Instantiation of monitoring activities

To exemplify the approach of instatiating the requested monitoring activities, VPC monitoring is used as a case study. The internal and external interactions are shown in Figure 5.30. Link and connection monitoring activities are performed in a similar way.

When an OSF (e.g. Load Balancing) requests the monitoring of a VPC, the following steps are performed:

- the requesting OSF sends an action to the Current Load Model specifying the global VPC id, the average method, the moving time period (the window over which the statistics are averaged), and the granularity period (the sampling frequency) as parameters.
- the Current Load Model OSF discovers[1] the identity of the element level manager corresponding to the source node of the VPC, and the particular object instance representing the VPC in the element manager MIB.
- the Current Load Model creates a metric object [5.61] in the network element level manager to monitor the attribute corresponding to the requested statistic of the VPC.
- the Current Load Model then creates an object on its own MIB acting as a proxy to the created metric object. This DN of this object is made available to the requesting OSF so it can be managed.
- the metric object in the element level manager polls the object corresponding to the VPC holding the required data.

5.11.2.2 Information modelling

The inheritance hierarchy and the containment schema of the information model are shown in Figure 5.31 and Figure 5.32.

1. ideally this is achieved through location transparency services (see Section 10.2.6), however this example assumes that the Configuration Manager provides these facilities.

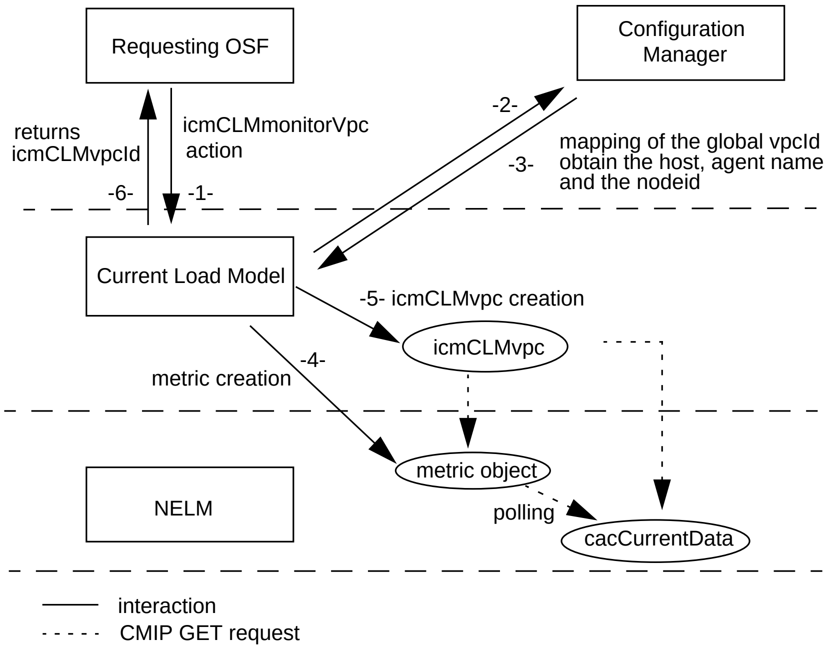

Figure 5.30 VPC monitoring in the Current Load Model

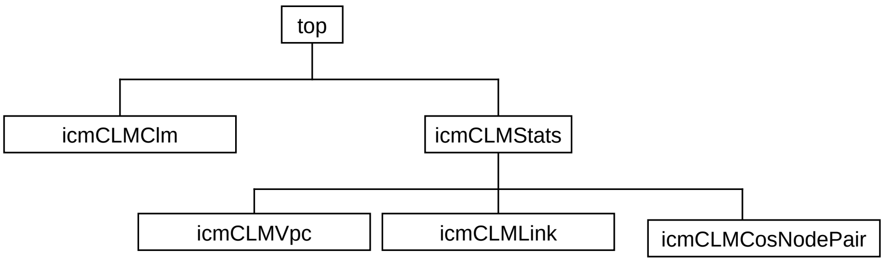

Figure 5.31 Current Load Model inheritance hierarchy

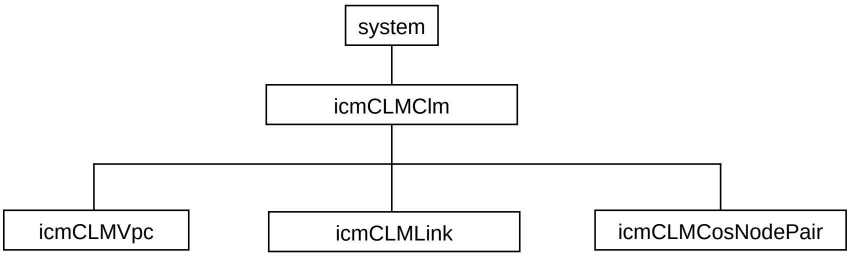

Figure 5.32 Current Load Model containment schema

5.12 Connection Type Model

5.12.1 Introduction

The Connection Type Model (CTM) is a repository of information of the types of connections (or Classes of Service - CoSs) supported by the network. The Connection Type Model OSF does not perform any management function as such. It is a repository of information to be accessed by other OSFs.

5.12.2 Concepts

Bearer services are characterised by the requirements of users and the capability of the network under consideration. The user requirements are captured in Quality of Service (QoS) parameters and the network capabilities are reflected in Network Performance Parameters (NPPs). QoS pertains to user-oriented performance concerns of an end-to-end service, while Network Performance (NP) is concerned with parameters that are of concern to network planning, provisioning and operations activities.

5.12.2.1 Quality of Service parameters

The collective effect of service performance determines the degree of satisfaction of a user of the service. The QoS is characterised by the combined aspects of service support performance, service operability performance, serviceability performance, service integrity and other factors specific to each service [5.71]. ITU recommendation I.350 defines QoS for bearer services and teleservices.

5.12.2.2 Network Performance parameters

NP is the ability of a network or network portion to provide the functions related to communications between users.

- Note 1 - NP contributes to serveability performance and service integrity.
- Note 2 - NP measures are meaningful to network providers and are quantifiable at boundaries of network portions to which they apply.

QoS measures are only quantifiable at a service access point. [5.71]

For each bearer service the Connection Type Model maintains a list of NPPs. The actual derivation of QoS of the bearer services is not within the scope of the Connection Type Model. Each NPP is associated with a maximum tolerable value identifying its acceptable values, according to the associated QoS requirements. As described in previous sections, it is the role of the management system (a combination of Performance Verification, Route Design, etc.) to guarantee that the performance of each bearer service is within acceptable levels.

When there is under provisioning of network resources, implying that more users contend for shared resources, there is a need for introducing a mechanism for selectively rejecting user requests for resource consumption. *Payoff* is a specification of the degree of desirability to allocate a quantity of resources to a contending consumer instead of allocation of the resource to any of the remaining contenders. It is therefore based on QoS, NP and other predefined social/enterprise aspects of resource consumption.

5.12.3 Connection Type parameters

Connection Type attributes have been defined in RACE CFS D120 [5.69]. The ATM Forum in [5.23] describe a traffic contract for the provision of network services.

The Connection Type Model OSF provides parameters which are collectively called CoS characteristics and describe both the traffic bandwidth characteristics, and acceptable limits on related NPPs.

In Table 5.2, the CoS characteristics are described and the mapping they have to [5.23] and [5.69].

Connection Type Parameter	Corresponding Parameter from [5.23] or [5.69]
Connection Type Id.	None
Peak Bandwidth (Mbits/sec)	Peak Bit Rate
Mean Bandwidth (Mbits/sec)	Mean Bit Rate
Maximum Delay (millisecs)	Mean Cell Transfer Delay
Maximum Delay Jitter (millisecs)	Cell Delay Variation
Tolerable Cell Loss Ratio	Cell Loss Ratio
Connection Rejection Ratio	Connection Set-up Denials Ratio

Table 5.2 Connection Type parameters

5.12.3.1 Pay-offs

The Connection Type Model stores information concerning pay-offs which could additionally be used by the VPC and Routing management OSFs. The pay-offs considered are:

- *VPC Type Payoff*. This is a rating of the VPC as a resource and is based on the PNOs' business and networking policies.
- *User Payoff*. These are parameters that determine the desirability of a user in being offered a service, as seen by the service provider. This characterisation includes the emergency service provision and the other social considerations required by legislation.
- *Connection Type Payoff*. A connection can be characterised by the amount of resource requested in the connection. Such a characterisation can be used to bias the desirability of accepting a connection - thus a connection type payoff.

5.12.4 Information modelling

Figure 5.33 and Figure 5.34. show the inheritance hierarchy and the containment schema of the Connection Type Model.

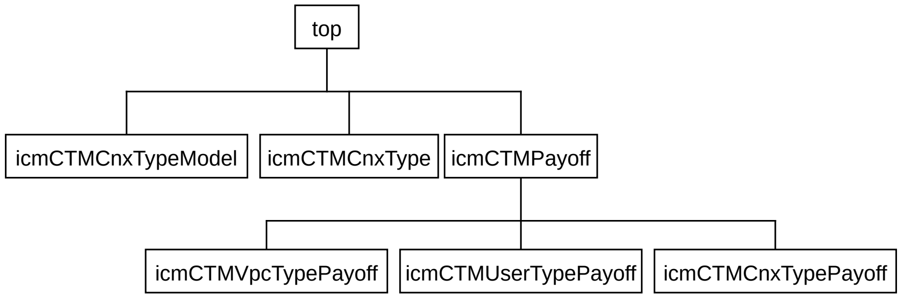

Figure 5.33 Connection Type Model inheritance hierarchy

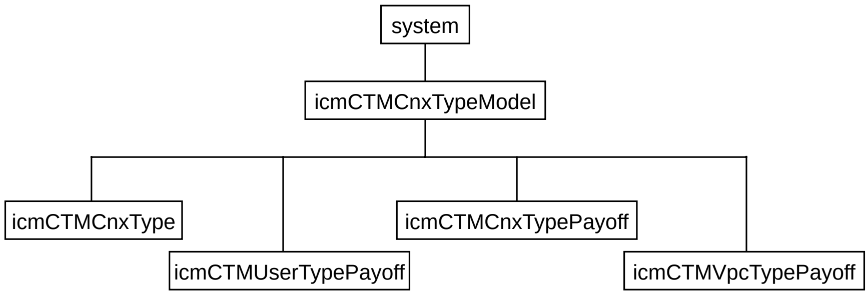

Figure 5.34 Connection Type Model containment schema

5.13 CAC Manager

The CAC Manager reproduces the CAC algorithms in the network. This way, the TMN system is consistent and independent of the control functions in the network.

The CAC Manager is supplied with a traffic mix in the form of a list of the number of connections of each CoS and returns the effective bandwidth of that traffic mix. The calculation has exactly the same result as the equivalent CAC algorithm in the network.

The task of CAC Manager OSF is not to accept or reject connection requests, but its purpose is to estimate the bandwidth consumed on a VPC by a certain mix of connections.

Figure 5.35 shows the inheritance hierarchy and the containment schema of the MIB presented to the other OSFs.

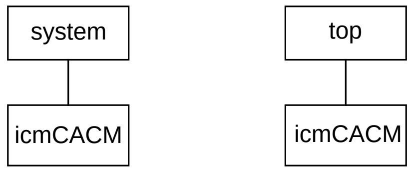

Figure 5.35 CAC Manager inheritance hierarchy and containment schema

5.14 Conclusions and future work

This chapter dealt with a VPC and Routing Management Service for multi-class ATM networks. Considering the requirements, the issues were analysed in order to decompose the problem into a number of distinct but cooperating functional components. Following the methodology described in Chapter 3 the Management Service was decomposed into MSC and MFCs which were then mapped to OSFs and then OSs in the TMN architecture. The functionality of each of the identified components was described and their inter-relationships were analysed to highlight their operational dependencies and information exchange in the context of the overall system operation. Management algorithms and models for realising the functionality and interfaces of each OSF were also presented.

The proposed system offers the generic functions of performance monitoring, load monitoring and configuration management of ATM networks. In addition, it provides specific functions for routing and bandwidth management in a hierarchical structure. The components in the hierarchy differ in terms of the level of abstraction, complexity and timescale over which they operate. The management functions to be invoked most frequently are close to the network elements and are as lightweight as possible to reduce management overhead and to increase their speed of response. The more comprehensive and computationally intensive functions are placed in the higher levels of the hierarchy and are only invoked when the lower levels are unable to resolve issues within the scope of their functionality and operational parameters. Such a hierarchy provides for continuous refinement of the management decisions and avoids the problems of a fully centralised approach.

The VPC and routing management system provides the following benefits:

- It allows the network to operate efficiently within the constraints of the physical resources. It will indicate when the network resources are insufficient for the traffic and hence additional resources need to be deployed. Alternatively it will show when resources are under-used and may be taken out of service or redeployed to avoid congestion elsewhere.
- It implements the necessary management hooks to allow the service management layer to provide for users according to the business policy of the network operator. A range of service qualities can be defined, which the VPC and Rout-

ing management system will implement and guarantee, for which the service management layer may charge different prices.

- It allows a large range of service types to be defined with different bandwidth parameters (CBR and VBR) and different levels of availability, cell loss, delay, delay jitter.
- It designs logical overlay VPC and routing networks so that the different service types can exist on the same physical network.
- The system enhances the role and operation of relatively simple CAC and RSA mechanisms by dynamically configuring their operational parameters to the strategy of the network operator and to allow global network conditions to influence the local algorithms in the switches. The design is flexible enough to incorporate different algorithms or different levels of functionality to adapt to the specific CAC and RSAs in the network elements. Static algorithms in the elements can be transformed to quasi-static algorithms by TMN actions.
- It regulates load distribution throughout the network: to maximise network availability and minimise disruptions in the case of failures. The system, by taking into account predictions as well as network-wide measurements inherits the merits of state-dependent routing. Moreover, by aiming at influencing the routing decisions so that the least congested route is selected, the signalling overhead and hence the connection set-up time is reduced. Furthermore, by monitoring route availability and comparing link loads, preventive management is achieved to avoid network call-blocking, enabling appropriate actions to be taken by the TMN before the network actually becomes unavailable for some services.
- The system monitors parameters characterising network performance and achieves this within the constraints of: minimising the management overhead imposed by collecting the required network information; and ensuring the validity and sensitivity of the required measurements.
- It can make dynamic configurations to adapt the network configuration to fluctuating traffic. It builds a Predicted Usage Model based on service level information. The predictions are used to make changes before they actually happen. But because it is expected that the predictions will be inaccurate, it matches the actual usage to the predicted usage and dynamically changes bandwidth distribution and route selection parameters when more accurate monitoring information is obtained. At the same time the Predicted Usage Model is updated with the real usage data so that the predictions may become more accurate as time passes.
- The system not only provides functions for measuring the performance of the network, but also for quantifying the performance of the network management system itself, acting as an indisputable measure of the efficiency of the deployed management systems.

By building intelligence into the TMN (the management plane) the requirements on the NEs (in the control and user planes) are simplified. The TMN functions provide an alternative to more elaborate algorithms in the switches that must interact via signalling procedures to allow global network conditions to influence local algorithms. By placing these functions in the TMN no additional requirements are placed on the NEs apart from the most basic of management interfaces.

Specific algorithms fulfilling the functionality of the OSFs were derived. These algorithms explicitly take into account the multi-service nature of communications environments and harmoniously coexist with each other within the management hierarchy of the proposed system. At the same time a clear distinction between the control and management planes is preserved.

Policies and algorithms for constructing and dynamically managing VPC and route networks respecting the performance needs of the different CoSs were proposed.

Specific algorithms for route selection management were proposed, taking into account the wide range of traffic types coexisting in broadband multi-service networks. The proposed algorithms are built around the concept of ‘route potentiality’ (for setting-up new connections). The routes with the highest potentiality are established and subsequently are recommended for routing to the RSAs. A centralised algorithm for finding highest potentiality routes was described.

By taking advantage of the rich and powerful features of OSI network management, an efficient design approach was proposed or implementing the identified functionality. Requirements on supporting monitoring objects to cover the required performance monitoring needs were drawn and enhancements on the ‘classical’ OSI metric and summarisation monitoring objects were identified.

The proposed system is needed for implementing private and virtual private network services since it manages bandwidth reservation and routing within specified performance targets. The system can provide an abstract interface to the service management functions responsible for the private services to implement their requests.

The architectural framework can be used as a testbed for testing and validating bandwidth management, routing management and load balancing algorithms.

Future work is mainly concerned with the detailed performance assessment of the proposed system both at architectural and algorithmic levels. Other aspects of future work include further investigation of functional issues such as: determining appropriate activation periods or conditions; enhancement of bandwidth and routing management algorithms; and refinement of VPC metrics for deriving route potentiality. Furthermore there is scope for further research into enhancing the identified functional components such as the customer complaint analysis and fault management functionalities, and identifying their relationships and dependencies with the system proposed in this chapter.

5.15 References

[5.1] K. R. Krishnan, “Markov Decision Algorithms for Dynamic Routing,” IEEE Communications Magazine, Oct. 1990.

[5.2] Y. Watanabe, T. Oda, “Dynamic Routing Schemes for International Networks,” IEEE Communications Magazine, Oct.1990.

[5.3] E. Sykas, K. Vlakos, E. Protonotarios, “Simulative Analysis of Optimal Resource Allocation and Routing in IBCNs,” IEE J. Select. Areas Commun., Vol.9, No.3, April 1991.

[5.4] G. R. Ash, B. D. Huang, “An analytical Model for Adaptive Routing Networks,” IEEE Trans. Commun., Vol.41, No.11, Nov.1993.

[5.5] B. Gavish, I. Neuman, “Routing in a Network with Unreliable Components,” IEEETrans. Commun., Vol.40, No.7, July 1992.

[5.6] J. Y. Hui, M. B. Gursoy, N. Moayeri, R. D. Yates, “A Layered Broadband Switching Architecture with Physical or Virtual Path Configurations,” IEEE J. Select. Areas Commun., Vol9, No.9, Dec.1991.

[5.7] J. Y. Hui, “Resource Allocation for Broadband Networks,” IEEE J. Select. Areas Commun., Vol.6, No.9, Dec.1988.

[5.8] J. Y. Hui, “Layered Required Bandwidth for Heterogeneous Traffic,” Infocom 1992.

[5.9] S. Chang, “Fair Integration of Routing and Flow Control in Communication Networks,” IEEE Trans. Commun., Vol.40, No.4, April 1992.

[5.10] S. Ohta, K. Sato, “Dynamic Bandwidth Control of the Virtual Path in an Asynchronous Transfer Mode Network,” IEEE Trans. Commun., Vol.40, No.7, July 1992.

[5.11] G. Woodruff, R. Kositpaiboon, “Multimedia Traffic Management Principles for Guaranteed ATM Network Performance,” IEEE J. Select. Areas Commun., Vol.8, No.3, April 1990.

[5.12] Y. Sato, K. Sato, “Virtual Path and Link Capacity Design for ATM Networks,” IEEE J. Select. Areas Commun., Vol.9, No.1, Jan.1991.

[5.13] R. Guerin, L. Gun, “A Unified Approach to Bandwidth Allocation and Access Control in Fast Packet Switched Networks,” Infocom 1992.

[5.14] D. Hong, T. Suda, “Congestion Control and Prevention in ATM Networks,” IEEE Network Magazine, July 1991.

[5.15] M. Wernic, O. Aboul-Magd, H. Gilbert, “Traffic Management for B-ISDN Services,” IEEE Network, Sept.1992.

[5.16] J. Labetoule, G. W. Hart, “Blocking Probabilities in Multitraffic Loss Systems: Insensitivity, Asymptotic Behaviour, and Approximations,” IEEE Trans. Commun., Aug.1992.

[5.17] C. Lea, A. Alyatama, “Bandwidth Quantization in Broadband ISDN,” Infocm 1992.

[5.18] H. Saito, K. Shiomoto, “Dynamic Call Admission Control in ATM Networks,” IEEE J. Select. Areas Commun., Vol.9, No.7, Sept.1991.

[5.19] M. Lebourges, “Routing with economic regulation for telecommunication networks,” ITC 14 / J.Labetoule and J.W.Roberts (Eds), 1994 Elsevier Science B.V.

[5.20] T. Murace, H. Suzuki, S. Sato, T. Takeuchi, “A Call Admission Control Scheme for ATM Networks Using a Simple Quality Estimate,” IEEE J. Select. Areas Commun., Vol.9, No.9, Dec.1991.

[5.21] A. Arvidsson, “Management of Reconfigurable Virtual Path Networks,” ITC 14 / J.Labetoule and J.W.Roberts (Eds), 1994 Elsevier Science B.V.

[5.22] E. Gelenbe, X. Mang, “Adaptive Routing for Equitable Load Balancing,” ITC 14 / J. Labetoule and J.W.Roberts (Eds), 1994 Elsevier Science B.V.

[5.23] ATM Forum, “ATM User-Network Interface Specification,” Version 3.0, Sept. 1993.

[5.24] K. Geihs, P. Francois, D. Griffin, C. Kaas-Petersen, A. Mann, “Service and traffic management for IBCN,” IBM Systems Journal, Vol. 31, No. 4, 1992.

[5.25] ITU-T Recommendation I.320, “ISDN protocol reference model.”

[5.26] ITU-T Recommendation I.321, “B-ISDN protocol reference model and its application.”

[5.27] ITU-T Recommendation I.150, “B-ISDN asynchronous transfer mode functional characteristics.”

[5.28] ITU-T Recommendation I.362, “B-ISDN ATM Adaptation Layer (AAL) functional description.”

[5.29] ITU-T Recommendation M.3010, “Principles for a telecommunications management network.”

[5.30] ITU-T Recommendation M.3020, “TMN interface specification methodology.”

[5.31] RACE R1005 NEMESYS Deliverable 5, “Specification of the NIKESH Experimental System,” 05/GPT/BNM/DS/X/009/b1, May 1990.

[5.32] RACE R1005 NEMESYS Deliverable, “NIKESH - experiment 2 in NEMESYS,” 05/KT/AS/DS/B/017/b1, June 1991.

[5.33] RACE R1005 NEMESYS Deliverable 9, “Specification of the EMMANUEL Experimental System,” 05/KT/AS/DS/B/021/b1, December 1991.

[5.34] RACE R1005 NEMESYS Deliverable 10, “Experiment 3 design,” 05/DOW/SAR/DS/B/024/a1, May 1992.

[5.35] D. Griffin, P. Patel, “Traffic Management for IBC Networks,” in Proc. Fifth RACE TMN Conference., Church House, London 1991.

[5.36] J. Hui, “Resource Allocation for Broadband Networks,” IEEE Journal on Selected Areas on Communications, Vol. 6, No. 9, Dec. 1988, pp. 1598-1608.

[5.37] K. 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.

[5.38] RACE R1003 GUIDELINE deliverable ME8, “TMN Implementation Architecture,” 03/DOW/SAR/DS/B/012/b3, March 1992.

[5.39] RACE R2059 ICM Deliverable 5, “Revised TMN Architecture, Functions and Case Studies,” R2059/ICS/DPG/DS/P/007/b1, David Griffin, editor, September 1993.

[5.40] RACE R2059 ICM Deliverable 14, “ICM Case Studies,” R2059/QMW/BM1/DS/P/014/b1, Babul Miah, editor, September 1994.

[5.41] RACE R2059 ICM Deliverable 17, “Specification of Interfaces (Q, X and M),” R2059/DNAC/NAZ/DS/P/017/a1, Nazim Agoulmine, editor, May 1995.

[5.42] RACE Common Functional Specification H200, “An Architecture for the Telecommunications Management Network,” Document 6, Issue C, December 1992.

[5.43] H. Rudin “On routing and delta-routing: A taxonomy and performance comparison of techniques for packet-switched networks,” IEEE trans on comm, vol COM-24, No.1, pp.43-59, 1976.

[5.44] T. P. Yum “The design and analysis of a semidynamic deterministic routing rule,” IEEE trans on comm, vol. COM-29, No.4, pp.498-504, 1981.

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

[5.46] S. Gupta, K. Ross, M. El Zarki “Routing in Virtual Path Based ATM networks.”

[5.47] ITU-T Recommendation E.500, “Measurement and Recording of Traffic,” 1989, pp71-82.

[5.48] ITU-T Recommendation E.501 (rev.1), “Estimation of Traffic Offered in the Network,” 1992, pp1- 16.

[5.49] ITU-T Recommendations E.506, E.507, E.508, “Forecasting of Traffic,” 1989, pp111-156.

[5.50] ITU-T Recommendation E.71X, “User Demand Modelling in Broadband ISDN,” COM 2-R 29-E, March 1995.

[5.51] ITU-T Recommendation M.3100, “Generic network information model,” October 1992.

[5.52] ITU-T Recommendation G.803, March 1993.

[5.53] ITU-T Recommendation G.774, September 1992.

[5.54] ETSI ETSI/NA5 WP BMA November 1993 (Draft ETS).

[5.55] D. Griffin, P. Georgatsos, “A TMN System for VPC and Routing Management in ATM Networks,” in proceedings of the 4th international symposium on Integrated Network Management, Sethi, Raynaud and Faure-Vincent, editors, Chapman & Hall, London, UK, 1995.

[5.56] P. Georgatsos, D. Griffin, “A Management System for Load Balancing through Adaptive Routing in Multi-Service ATM Networks,” to be published in the proceedings of IEEE INFOCOM'96 to be held March 26-28, 1996.

[5.57] ITU-T Recommendation X.734, “Event Report Management Function”(ISO/IEC 10164-5), 1993.

[5.58] ITU-T Recommendation X.733, “Alarm Reporting Function” (ISO/IEC 10164-4), 1992.

[5.59] ITU-T Recommendation X.735, “Systems Management: Log Control Function,” (ISO/IEC 10164-6), September 1992.

[5.60] ITU-T Recommendation X.745, “Test Management Function” (ISO/IEC 10164-12).

[5.61] ITU-T Recommendation X.739, “Metric Objects and Attributes” (ISO/IEC 10164-11), March 1993.

[5.62] ITU-T Recommendation X.738, “Measurement Summarization Function” (ISO/IEC 10164-13), October 1992.
[5.63] A. Aho, J. Hopcroft, J. Ullman “Data structures and algorithms,” Addison-Wesley pub.co., 1983.
[5.64] R. R. Boorstyn, A. Livne “A technique for adaptive routing in networks,” IEEE Trans. on Comm., Vol.29, 1981.
[5.65] A. Girard, S. Hurtubise “Dynamic Routing and Call Repacking in Circuit-Switched Networks,” IEEE Trans. on Comm., Vol.31, No.12, Dec.1983.
[5.66] N. Eshragh, P. Mars “Study of dynamic routing strategies in circuit-switches networks,” 3rd UK comp. and telecomm. perf. engin. workshop, Edinburgh, Sept.1987.
[5.67] A. Kershenbaum, P. Kermani, G. Grover “MENTOR: An Algorithm for Mesh Network Topological Optimization and Routing,” IEEE Trans. on Comm., Vol.19, No.4, April 1991.
[5.68] M. Schwartz, T. E. Stern “Routing techniques used in computer communication networks,” IEEE Trans. on Comm., Vol.28, 1980.
[5.69] RACE CFS D120, “Specification of Overall B-ISDN Connection Types,” December 1993.
[5.70] RACE CFS D510, “General Aspects of Quality of Service and System Performance in IBC,” December 1993.
[5.71] ITU-T Recommendation E.800, “Quality of Service and Dependability Vocabulary, Terms and Definitions Related to the Quality of Telecommunications Services,” 1993.
[5.72] ITU-T Recommendation F.500, “International Public Directory Services, Directory Service Operations and Definitions of Service,” August 1992.