26

A general framework for routing management in multi-service ATM networks

P. Georgatsos
Algosystems S.A., 4 Sardeon St., N. Smyrni, 17121 Athens, Greece.
tel: +30 1 93 10 281, fax: +30 1 93 52 873, email: pgeorgat@algo.com.gr

D. P. Griffin
Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK. tel.: +44 171 419 3687,
fax: +44 171 387 1397, email: D.Griffin@cs.ucl.ac.uk

Abstract

This paper presents a framework for routing management in ATM networks supporting guaranteed quality connections. It discusses the rationale behind the decomposition of a routing management service into a hierarchical system comprising both the management and control planes of the network. The concepts and ideas behind a set of algorithms for implementing the desired functionality are developed and discussed.

Keywords

ATM, TMN, load balancing, performance management, route design, routing, routing management, VPC, multi-class environment.

1 INTRODUCTION

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 conditions.

Routing in Asynchronous Transfer Mode (ATM) networks 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

A. A. Lazar et al. (eds.), *Integrated Network Management V*

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.

To date, the majority of research work in the area of routing has been concerned with routing algorithms in isolation, rather than considering both the requirements and capabilities of the networks and the interplay with other resource management functions. One of the first attempts to design an integrated approach to the problem of routing management in multi-service ATM networks was presented by Griffin (1995).

This paper describes a general framework for tackling routing management in ATM networks. The framework encompasses both the control and management planes in a hierarchical manner assuming that route selection at call set-up time is handled by control plane functions, but according to parameters set by routing management. The routing management system is part of the management plane, and it is itself a hierarchical system operating in parallel to the network and its embedded control functionality.

The paper is organised as follows: Section 2 introduces the management service and describes our general functional model for routing management. Section 3 describes and analyses the issues behind the algorithms necessary for realising the components of the functional model. Finally, Section 4 draws conclusions on the work presented in this paper and identifies the scope for future work in this area.

2 DESCRIPTION OF THE MANAGEMENT SYSTEM AND GENERAL FUNCTIONAL MODEL

It is assumed that the network being managed is a public ATM network offering switched, on-demand services which are composed of a number of unidirectional connections. Each connection falls into a particular connection type or class, denoted throughout this paper by the term Class of Service (CoS). The CoS definition characterises the connection type in terms of bandwidth and performance requirements. We assume that it is in the responsibility of the network to guarantee the performance of the supported CoSs. Such performance guarantees are provided in terms of upper bounds on parameters characterising CoS performance in the network. A more detailed treatment of the assumptions made on the network can be found in (ICM, 1997) (Griffin, 1995).

The objective of the Routing management service is *to manage the network routing functions so as to maximise network availability whilst guaranteeing the performance requirements of current and future connections as specified in their CoS definitions.*

In the following paragraphs, we present an implementation independent functional architecture for the Routing Management service, by decomposing it into a number of distinct functional components. This architecture was first introduced by Griffin (1995) and it is further elaborated here. The basic dimensions of reasoning for the decomposition, are as follows:

- The definition of routing information is made on the basis of traffic predictions of anticipated network usage.
- Predictions may vary over time, hence routing information needs to be revised.
- Traffic estimates may prove to be inaccurate, furthermore actual traffic load may vary within a statistical range around the predicted values.

From the above it is apparent that there are basically two levels at which adaptivity to traffic variations should be provided; one at a level of traffic prediction changes and one

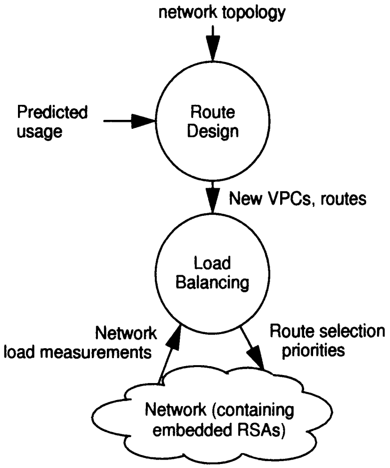

Figure 1 Routing management - a hierarchical approach

at the level of actual traffic fluctuations around the predictions. Therefore, we propose that routing management is achieved through a two level hierarchy (see figure 1).

The higher level of the routing management hierarchy, the Route Design component, operates at epochs where traffic predictions change, producing new sets of routes per CoS. Suitable sets of routes are constructed using, as input, predictions of network traffic at a connection level, per CoS and per s-d pair. Whenever traffic predictions change, the sets of routes may need to be reconstructed. The level of reconstruction depends on the significance of the changes (Griffin, 1995).

The lower level, the Load Balancing component, operates within the time-frame of traffic predictions and within the set of VPCs and routes defined by the higher level. Its objective is to manage the Route Selection Algorithms (RSA) embedded within the network switches according to actual traffic conditions. This 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 warns the higher level component of undesirable trends in network availability based on actual usage measurements.

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.

The proposed routing management system implies a semi-dynamic type of routing policy, combining the merits of centralised and decentralised (based on local information) routing policies. Semi-dynamic routing policies were introduced in traditional data networks and improvements in network performance has been confirmed under such schemes (Rudin, 1976) (Yum, 1981).

An essential point underlying the proposed system is that it explicitly views route definition and route selection management functionality as management plane functionality. As a result, the routing functions in the control plane embedded in the network switches need only incorporate the actual route selection functionality; the management of which is left to the overlying management components.

One of the critical issues associated with routing has always been the assessment of the impact of routing decisions, at a particular instant, on future network performance. The proposed routing management model tackles this issue through its hierarchy. The higher level reserves the resources for routing based on estimates per source-destination (s-d) pair and CoS which refer to a long-term time period. According to the information supplied by the higher level, the lower level may be able to make predictions regarding shorter-term network usage and influence accordingly the routing decisions. With the proposed hierarchical structure the impact of routing decisions is time bounded and routing is always made on the basis of anticipated traffic.

Finally, it is worth-seeing how management of guaranteed performance connections is achieved in the proposed model. It is in the responsibility of the higher level routing component - Route Design - to apply suitable route design algorithms that preserve the performance targets of the supported traffic by defining appropriate sets of routes. The lower level component manages the performance of the different classes in an indirect manner by sharing the network resources between the competing classes, through appropriately influencing the routing decisions. Various prioritisation policies can be realised through this component in cases where network availability is limited, favouring certain CoSs at the expense of others.

3 FUNCTIONAL ASPECTS

3.1 Route Selection Algorithm (RSA)

Route selection is achieved in the control plane by means of a RSA which, without loss of generality, is assumed to operate on the basis of *route selection parameters* associated with the available routes. Following the ideas on the taxonomy of routing algorithms (Rudin, 1976), (Schwartz, 1980), 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, RSAs may be:

- *Deterministic:* route selection is made according to a predefined order. In this case, a priority is assigned to each alternate route and higher priority routes are selected first.
- *Random:* route selection is made based on probabilistic criteria. Routes are assigned probabilities or frequencies and the selections are made to guarantee the frequencies.
- *Locally Adaptive:* 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).

Apart from the routing information available at the network switches, RSAs may base their decisions upon purely local information (e.g. congestion at the VPCs originating from that switch), network-wide information (e.g. congestion on the routes to specific destinations), or upon no information at all. According to their degree of adaptivity, RSAs may be: static (not adaptive at all) or dynamic. A significant parameter associated with adaptivity is how adaptivity is provided; it can be provided: locally at connection acceptance/release times; through inter-node exchange, periodically or at exception; or from the management system periodically or at exception.

From the many algorithms that may be derived according to the above taxonomy, locally adaptive RSAs incorporating adaptivity provided by a network-wide routing management component (semi-dynamic routing schemes) seem to result in better network performance compared to other adaptive schemes, as indicated in the literature for data networks (Rudin, 1976) (Yum, 1981). However, the optimality of particular RSAs should not be judged in isolation but in conjunction with the algorithms incorporated in the higher routing management components.

3.2 The Load Balancing Component

The scope of the Load Balancing component is to:

- manage the RSAs in the network nodes according to network-wide traffic conditions;
- monitor the network with the purpose of providing warnings on deterioration in route availability and load deviations which would result in an unbalanced network.

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 proved by the huge quantity of literature on the subject (Girard, 1983), (Eshragh, 1987), (Rudin, 1976), (Schwartz, 1980). 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 (Gelenbe, 1994), (Kershenbaum, 1991). Apart from its active role in routing management, the Load Balancing component also contributes to preventive management. By taking a future perspective, it notifies the higher-level routing management component (Route Design) of undesirable trends in network availability. Thus, appropriate actions to increase network availability may be taken before the network availability deteriorates below acceptable levels.

The essence of any route selection management 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 (Schwartz, 1980). In connection-oriented networks, such as ATM, the figure of merit should refer to the *potentiality of the route* to accommodate new connections.

Route potentiality is calculated for all possible routes from a given switch to a particular destination node and for each CoS. It is calculated in terms of a metric associated with each VPC which reflects the potentiality of the VPC to accept connections of a particular CoS given its current load; this metric is referred to as *VPC acceptance potential.* Taking into account that on a particular VPC originating at a switch more than one route to a particular destination for a given CoS may originate, the potentialities of all these routes can be accumulated, giving rise to a figure of merit for selecting this VPC, referred to as *VPC selection potential.* Route selection management algorithms then grade VPCs at each switch according to their selection potential and configure the RSAs accordingly. This is achieved by setting the route selection parameters so that VPCs with a higher figure of merit have advantage over those with lower figures of merit.

To achieve network load surveillance, network availability for new connections can be estimated by extending the notion of potentiality to the node level. *Node potentiality* to a given destination for a particular CoS reflects the potential of the network to establish new connections from this node to the given destination node for a particular CoS, taking into account actual network traffic conditions.

Specific route selection management algorithms can then be distinguished according to the way they calculate route potentiality, the location in which they execute their algorithms for determining VPC selection potentials, and the way they perceive changes in VPC selection potentials and then trigger appropriate management actions.

Load balancing algorithms along the previously introduced concepts have been presented by Georgatsos (1996) and ICM (1997), and improvement in network performance has been verified assuming a deterministic type of RSA.

3.3 The Route Design Component

The scope of the Route Design component is to design and redesign - whenever necessary - a network of VPCs and a set of admissible routes per (s-d) and CoS satisfying predicted traffic requirements, given the constraints imposed by the network model and the performance targets of the CoSs. The following terminology and notions are introduced:

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 (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.

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.

A *design problem*, defined by its literal meaning, is characterised by its objectives, its constraints and by its uncontrolled variables for which solutions are sought. The objectives of a design problem are usually defined in qualitative terms (e.g. optimum solutions) rather than in quantitative terms. Associated with a design problem, a *design alternative* is defined as an arrangement of the problem's uncontrolled variables, satisfying the problem's constraints.

A design problem is said to be a *well defined design problem* if there exist two functions of the problem's uncontrolled variables one corresponding to the cost (*cost function*) and the other corresponding to the benefits (*benefit function*) of alternative designs. Therefore, each design alternative of a well defined problem can be explicitly evaluated in terms of the cost that it implies and the benefits that yields. Then, the following are defined, where c(.), b(.) are the problem's cost and benefit functions respectively and D is the set of all design alternatives:

A *least-cost design* of a well defined problem is that design alternative which has the minimum cost among all other possible designs, irrespective of benefits. Design $d' = (\underline{x}')$ is a least-cost design if and only if $c(\underline{x}') \le c(\underline{x}), \forall d = (\underline{x}) \in D$.

A *maximum-benefit design* of a well defined problem is that design alternative which achieves the maximum possible benefits among all other possible designs, irrespective of cost. Design $d' = (\underline{x}')$ is a maximum-benefit design if and only if $b(\underline{x}') \ge b(\underline{x}), \forall d = (\underline{x}) \in D$.

A *cost-effective design* of a well defined problem is that design alternative which achieves the maximum possible benefits at the least cost among all other possible designs. Design $d' = (\underline{x}')$ is a cost-effective design if and only if the following hold:

$$\frac{c(\underline{x}')-c(\underline{x})}{c(\underline{x}')} \le \frac{b(\underline{x}')-b(\underline{x})}{b(\underline{x}')},\ \forall d=(\underline{x}) \in D \text{ such that } c(\underline{x}) \le c(\underline{x}') \text{ and } b(\underline{x}) \le b(\underline{x}')$$

$$\frac{c(\underline{x})-c(\underline{x}')}{c(\underline{x}')} \ge \frac{b(\underline{x})-b(\underline{x}')}{b(\underline{x}')},\ \forall d=(\underline{x}) \in D \text{ such that } c(\underline{x}) \ge c(\underline{x}') \text{ and } b(\underline{x}) \ge b(\underline{x}')$$

there are no $d = (\underline{x}) \in D$ such that $c(\underline{x}) \le c(\underline{x}')$ and $b(\underline{x}) \ge b(\underline{x}')$

It can be proved that a cost-effective design corresponds to the solution of an optimisation problem with the objective function being the ratio of the problem's cost by the benefit functions, i.e. c(.)/b(.), subject to the problem's constraints.

The existence of *optimum designs* (least-cost or maximum-benefit or cost-effective designs) to a well defined design problem depends on the existence of solutions to the corresponding optimisation problems (i.e. on the topological properties of the problem's state space as constrained by the set of constraints and on the mathematical properties of the problem's cost and benefit functions).

The essence of the proposed approach is to tackle the Route Design problem within the previously introduced conceptual framework of problem alternative designs. This entails decomposition of the overall Route Design problem into a number of distinct problems and tackling each of the problems by means of optimum designs i.e. defining appropriate cost and benefit functions and seek for optimum designs. Specifically, the approach is to decompose the overall problem of Route Design as follows:

- Map traffic predictions into max (maximum) flow requirements for all (s-d) pairs and each CoS. The max flow requirements are in terms of Virtual Channel Connections (VCCs) that the network must be able to establish at any instant, given the max connection rejection tolerance of each CoS.
- Determine SDClassPath networks, i.e. paths over which connections will be routed so that CoS performance is guaranteed. That is, given the max flow requirements determine suitable paths per (s-d) and CoS so that to satisfy the delay, jitter and cell loss constraints.
- Determine a suitable VPC network, based on the identified SDClassPath networks. That is, given the set of paths per (s-d) and CoS, determine a suitable VPC network.
- Map the SDClassPath networks to the VPC network to derive the SDClassRoute networks. That is, map the set of paths to the derived VPC network to obtain the set of routes per (s-d) and CoS.

Determining max flows

It is assumed that traffic predictions characterise the number of connection requests for a specific interval, in which statistical convergence is achieved. By modelling the network with an appropriate queuing 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 within the physical network.

Determining SDClassPath networks

The Route Design component 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 per (s-d) is assigned a *performance quality* corresponding to an upper bound on the delay, jitter and cell loss that a connection traversing this path would experience.

Various policies for allocating (s-d) pairs of a specific CoS to paths offering specific performance qualities can be identified. The following *performance sharing policies* are proposed:

Complete sharing: a path of a specific performance category is shared exclusively by the CoSs of the same performance category.

Compatible sharing: a path of a specific performance category is shared only by CoSs of the same or superior performance category.

Non-compatible sharing: paths of a specific category can be shared by any CoS.

In order to meet the cell loss requirements, the cell loss performance targets - determining the level of multiplexing at the cell level - of the Connection Admission Control (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 performance targets in such a way so that paths of any cell loss category can be established on them.

Complete partitioning: Links are assigned 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 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?

The above questions state the design problem of the *SDClassPath configuration.* The uncontrolled variable of this design problem is the topology of the SDClassPath networks. The constraints of the problem are: the max flow constraints ensuring that the max flow requirements per CoS and (s-d) should be accommodated within the defined SDClassPath networks and path sharing constraints corresponding to the adopted performance sharing policy. The problem is a well defined design problem as explicit cost and benefit functions can be defined. For example for a given design, i.e. SDClassPath networks' configuration, the cost and benefit functions could be the respective cost and benefit functions of the related optimum flow assignment problem.

Tightly coupled with the SDClassPath definition problem is the *optimum flow assignment* problem concerned with the optimum distribution of the max flow commodities (a commodity is the tuple: (s-d) and CoS) along the defined paths, given the topologies of the SDClassPath networks. The uncontrolled variable is the flow per commodity along each path or equivalently their distribution frequency at the source nodes. The constraints of the problem are the flow balance constraints in and out each node, the capacity constraints for each link and the max flow constraints. The cost function may be related to the cost of used network resources (the utilisation of the links) and the benefit function may be related to network load balancing (the discrepancy of a link's utilisation from the average link utilisation). By considering appropriate cost or benefit functions, optimum flow configurations can be established as solutions to appropriately defined optimisation problems.

Determining the VPC network

The definition of the VPC network encompasses the following tasks:

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

which are viewed as design problems for which optimal designs can be sought.

As far as the *VPC topology design* problem is concerned, the uncontrolled variable is the set of *VPC cut nodes*. A VPC cut node is defined to be a node where a VPC is terminated. VPC cut nodes are connected with (at least) as many VPCs as the number of physically different paths in the SDClassPath networks that connect them. The paths connecting two different VPC cut nodes are termed as *VPC platforms* upon which VPCs may be defined depending on the adopted *bandwidth segregation policy* (see below). Therefore, the definition of VPC cut nodes in effect determines the minimum number of VPCs which may exist in the network. Note that each VPC platform is characterised by its required bandwidth which can be determined by the expected number of commodities flowing between its edges. The mapping between the anticipated flow to the required bandwidth should be made taking into account the multiplexing scheme as applied at the cell level by the network CAC algorithms as well as taking into account multiplexing at the connection level. The anticipated flow along the VPC platforms was determined during the definition of the SDClassPath networks. At one extreme a VPC topology design treats all VC/VP switches as VPC cut nodes, whilst at the other extreme only the access nodes are VPC cut nodes. Optimum VPC topology designs could be sought for, considering explicit cost and benefit functions in terms of the VPC cut nodes. For example, for a given design - selection of VPC cut nodes - a cost function could reflect the cost of managing the VPCs (the number of VPC platforms) and a benefit function could reflect the benefits of reduced set-up times (inverse of the average number of hops on VPC cut nodes from source to destination over all SDClassPath networks). Appropriate optimisation problems can be formulated to determine the optimum VPC topology designs.

Tightly coupled with the VPC topology design problem is the *VPC bandwidth assignment problem*. The VPC assignment problem can be viewed as a design problem with uncontrolled variable the bandwidth assigned to the VPCs. The constraints of the problem are the link capacity constraints. The following policies can be identified for VPC instantiation, corresponding to different *bandwidth segregation policies*. Considering a simple model consisting of a single resource (bandwidth available on a VPC platform) for which a number of different classes (corresponding to the (s-d)-CoS commodities) compete, the following resource sharing policies can be identified:

Complete sharing: the resource is shared among all competing entities on the basis of their arrival.

Complete partitioning: the resource is appropriately segmented and access by the competing classes is restricted to a segment. Access between the classes competing for the same segment is done on the basis of their arrival.

Compete sharing with initial reservations: similar to the complete partitioning policy, however a portion of the resource is also defined for being shared by all classes on the basis of their arrival. The optimality of complete sharing with an initial reservation scheme has been experimentally verified (Sykas, 1991).

Optimum designs for VPC bandwidth assignment can then be sought for, provided that explicit cost and benefit functions can be instituted. For example, for a given design i.e. a given set of VPCs and their allocated bandwidth, a cost function may reflect the increased cost in using network resources (increase in link utilisation) caused by bandwidth segmentation on the links compared to the case where no VPCs have been defined. The utilisation of links is measured assuming the optimum flow distribution in the network of VPCs. An example of a benefit function could be related to the gains in call blocking probability attained by the defined VPC set of the particular design as opposed to the case where no VPC has been defined.

The dynamic part

The dynamic part of the Route Design could be achieved in a quasi-static way by rerunning the static part algorithms. However, we envisage a more efficient way of dealing with asynchronous triggers for the purpose of ensuring management operations are as economical as possible in terms of management overhead and reaction response. In particular, we envisage a hierarchical approach (Griffin, 1995) towards handling asynchronous triggers, as indicated below:

- Map the traffic predictions to max flow requirements and determine the optimum designs of the involved design problems corresponding to the new flow requirements.
- Determine whether the new flow requirements can be accommodated in the existing VPC and SDClassRoute networks, keeping the same VPC capacity.
- If not possible (either not feasible or flow distribution optimality constraints are violated), the bandwidth of the existing VPCs is appropriately modified (within link capacity constraints), keeping the same SDClassRoute networks to accommodate the new flow requirements.
- If not possible (either not feasible or flow distribution optimality constraints are violated), SDClassRoute networks are modified either by adding/removing existing VPCs and by appropriately modifying their bandwidth.
- If not possible (either not feasible or flow distribution optimality constraints are violated), the VPC network is appropriately modified (by creating/deleting VPCs) and by appropriately setting VPC bandwidth and defining SDClassRoute networks so that to accommodate the new flow requirements.
- If not possible (either not feasible or flow distribution optimality constraints are violated), triggers to higher level management functions (network planning, service migration) are emitted.

4 CONCLUSIONS AND FUTURE WORK

In this paper we presented a general framework for routing management in multi-service ATM networks. The framework encompasses both the control and management planes of the network, specifically covering the RSAs embedded within the network nodes and the management functions required to manage the RSAs. This hierarchical separation between the control and management functions allows network elements to be as simple as possible, reducing their complexity and hence their cost. At the same time the capabilities of the routing mechanisms are not compromised because the overlying management functions are able to influence the operation of the local routing algorithms by means of route selection parameters. By building the routing management intelligence into a parallel environment (the management plane) to that of the network itself (control and user planes) the complexity of the management algorithms is not constrained by the capabilities of the network elements.

In addition to the control plane/management plane hierarchy, the paper discussed the rationale behind the decomposition of the routing management service into a hierarchical system of management components fulfilling the roles of Route Design and Load Balancing. One of the main advantages of this separation of functionality is that management decisions taken on a longer term basis (by Route Design) can be continually refined (by Load Balancing) in the light of recent developments without the disadvantage of continually invoking computationally-intensive management processes.

Following the implementation independent analysis of the functions required for routing management, the resulting components were mapped to the Telecommunications Management Network (TMN) architecture (ITU1, 1992) by Griffin (1995) and ICM (1997) as a framework for implementation. The resulting system was implemented in the RACE II ICM project and the architecture and concepts described in this paper were validated through experimentation. The Load Balancing component, in particular, was subject to extensive experimental work (ICM, 1997) and at the time of writing, Route Design algorithms are being implemented following the concepts and ideas of section 3.

An important conclusion is that in our opinion the TMN framework is mature enough for realizing complex systems such as routing management. It was necessary to extend the ITU M.3020 methodology (ITU2, 1992) to give guidance on the decomposition of TMN systems (ICM, 1997), but having done this, the development of a large scale TMN system was facilitated through a well defined approach to system design and development.

An important aspect of our future work is concerned with the development of a more resilient routing management scheme by integrating the management service presented in this paper with fault management facilities. Our future work in this direction will be according to the TINA architecture. One of the advantages of the TINA approach is that there is no longer a distinct separation between control and management plane functions - a common framework is used for all telecommunications software in a distributed processing environment. This approach will have many benefits when interworking is required between route design, load balancing and control functions, and will facilitate interworking with fault management which also spans both the management (alarm handling and correlation) and control planes (self-healing OAM techniques, etc.).

In this direction there are a number of open issues related to the compatibility between TMN design methodologies and TINA modelling techniques. Work currently being undertaken in the context of the ACTS project AC208 REFORM (REFORM, 1997) concerns the relationship of the routing management system presented here with the resource configuration management, connection management and fault management subsystems of the TINA management architecture (TINA, 1994).

5 ACKNOWLEDGEMENTS

This paper describes work undertaken in the context of the RACE II R2059 ICM project and work currently in progress in the ACTS AC208 REFORM project. The RACE and ACTS programmes are partially funded by the Commission of the European Union.

6 REFERENCES

Eshragh, N., Mars, P. (1987) "Study of dynamic routing strategies in circuit-switches networks," 3rd UK comp. and telecomm. perf. engin. workshop, Edinburgh.

Gelenbe, E., Mang, X. (1994) "Adaptive Routing for Equitable Load Balancing," ITC 14, Elsevier Science B.V.

Georgatsos, P., Griffin, D. (1996) "A Management System for Load Balancing through Adaptive Routing in Multi-Service ATM Networks", proc. of IEEE INFOCOM'96.

Girard, A., Hurtubise, S. (1983) "Dynamic Routing and Call Repacking in Circuit-Switched Networks," IEEE Trans. on Comm., Vol. 31.

Griffin, D., Georgatsos, P. (1995) "A TMN system for VPC and routing management in ATM networks", Proc. of 4th ISINM, Chapman & Hall, UK.

ICM (1997) "Integrated Communications Management of Broadband Networks," ed., Griffin, D., Crete University Press, Heraklion, Greece, ISBN 960 524 006 8.

ITU-T Rec. M.3010 - Principles for a telecommunications management network, 1992.

ITU-T Rec. M.3020 - TMN interface specification methodology, 1992.

Kershenbaum, A., Kermani, P., Grover, G. (1991) "MENTOR: An Algorithm for Mesh Network Topological Optimization and Routing," IEEE Trans. on Comm., Vol. 19.

Rudin, H. (1976) "On routing and delta routing: A taxonomy and performance comparison of techniques for packet-switched networks", IEEE Trans. on Comm., Vol. 24.

REFORM (1997) "REFORM System Requirements and Analysis," Deliverable D1, A208/Bell/WP1/DS/P/002/b1.

Schwartz, M., Stern, T.E. (1980) "Routing techniques used in computer communication networks," IEEE Trans. on Comm., Vol. 28.

Sykas, E., Vlakos, K., Protonotarios, E., (1991) "Simulative Analysis of Optimal Resource Allocation and Routing in IBCNs", IEEE JSAC, Vol. 9.

TINA-C (1994), "Management Architecture," Version 2.0, Document label TB_GN.010_2.0_94.

Yum, T.P. (1981) "The Design and Analysis of a Semi-Dynamic Deterministic Routing Rule", IEEE Trans. on Comm., Vol. 29.

Panos Georgatsos received the B.S. degree in Mathematics from the National University of Athens, Greece, in 1985, and the Ph.D. degree in Computer Science, with specialisation in network routing and performance analysis, from Bradford University, UK, in 1989. Dr. Georgatsos is working for Algosystems SA, Athens, Greece, as a network performance consultant. His research interests are in the areas of network and service management, analytical modelling, simulation and performance evaluation. He has been participating in a number of telecommunications projects within the framework of the EU funded RACE and ACTS programmes.

David Griffin received the B.Sc. degree in Electronic, Computer and Systems Engineering from Loughborough University, UK in 1988. He joined GEC Plessey Telecommunications Ltd., UK as a Systems Design Engineer, where he was the chairperson of the project technical committee of the EU RACE I NEMESYS project while working on TMN architectures and ATM traffic experiments. In 1993, Mr. Griffin joined ICS-FORTH in Crete, Greece as a Research Associate on the EU RACE II ICM project. He was the leader of the ICM group on performance management case studies and TMN systems design. Mr. Griffin joined University College London in 1996 and is currently employed as a Research Fellow working on a number of EU ACTS projects in the area of resource management for TINA systems covering performance, fault, configuration and accounting management.