REAL-TIME NETWORK ROUTING IN THE AT&T NETWORK - IMPROVED SERVICE QUALITY AT LOWER COST

By

Gerald R. Ash, Jin-Shi Chen, Alan E. Frey
BaoSheng D. Huang, Chin-Kyooh Lee, Gail L. McDonald
AT&T Bell Laboratories
Holmdel, New Jersey 07733, USA

ABSTRACT

The introduction of Real-Time Network Routing (RTNR) into the AT&T Switched Network was completed in July 1991, and completely replaced dynamic nonhierarchical routing (DNHR). It has resulted in a marked improvement in AT&T network connection availability while simultaneously reducing network costs. RTNR has laid the foundation for an important AT&T service and quality advantage by introducing new class-of-service capabilities and dramatically improving network service and robustness, particularly in responding to abnormal traffic or failure conditions. RTNR provides the platform for the dynamic class-of-service routing features for emerging new services, and also provides a multiple ingress/egress routing arrangement to ensure reliability and flexibility for international and access networks. To date eight virtual networks have been established to serve major classes-of-service, which share the facilities of the AT&T network but provide individual control and monitoring of traffic as well as new priority service capabilities. RTNR allows a marked improvement in network reliability by providing for the selection of over 100 routes between every pair of cities for every call. In its first year of operation RTNR has shown outstanding performance improvements over DNHR. These improvements are illustrated across a spectrum of network conditions, which include a) high day loads on the Monday after Thanksgiving 1991, which set a record for the number of calls on the AT&T network, 157.5 million, and provided a completion rate on AT&T facilities of 99.999% on the first try, b) peak day loads on Thanksgiving and Christmas 1991, in which call blocking was reduced by more than 50% over the year before, c) a cable cut near Austin, Indiana on September 25, 1991, in which the combination of RTNR and automatic facility restoration provided excellent recovery from the failure, d) a focused overload caused by Hurricane Bob on August 19, 1991, in which priority services experienced virtually zero network blocking, and e) the international network performance to Taiwan during the Chinese New Year three-day holiday calling period, which showed considerable improvement in the quality of service following the introduction of RTNR multiple ingress/egress routing to that country. We conclude that RTNR improves network performance under all network conditions, that it simplifies the operations environment and lowers operations cost, that it enables the sharing of voice and data network capacity in the deployment of ISDN, and that it provides a possible flexible routing direction for the globalization of dynamic routing.

1. INTRODUCTION

Dynamic routing in telecommunications networks has been the subject of worldwide study and interest [1-24]. During the 1980s, dynamic nonhierarchical routing (DNHR) was fully deployed in the AT&T long distance network. DNHR has provided considerable benefits in improved performance quality and reduced costs [20], and has motivated the extension of dynamic routing to all classes-of-service. Real-time network routing (RTNR) is a new adaptive routing method which replaced DNHR in the AT&T network starting in 1991. RTNR provides the platform for a class-of-service dynamic routing strategy which supports dynamic routing on an integrated transport network for all new and existing voice, data, and wideband services. RTNR also provides a multiple ingress/egress routing arrangement to ensure reliability and flexibility for international and access networks.

In this paper we analyze the performance of RTNR in its first year of operation in AT&T's new dynamic class-of-service network. This network provides connections for voice, data, and wideband services on a shared transport network. These connections are distinguished by resource requirements, traffic characteristics, and design performance objectives. Class-of-service routing allows independent control, by class-of-service, of service specific performance objectives, routing rules/constraints, and traffic data collection. Class-of-service routing allows the definition of virtual networks by class-of-service. Each virtual network is allotted a predetermined amount of the total network bandwidth. This bandwidth can be shared with other classes-of-service, but is protected under conditions of network overload or stress. With RTNR, the originating switch first tries the direct route, and if available, sets up the call on a direct trunk to the terminating switch. If a direct trunk is not available, the originating switch tries to find an available two-link path by first querying the terminating switch through the common channel signaling (CCS) network for the busy-idle status of all trunk groups connected to the terminating switch. The originating switch then compares its own trunk group busy-idle status information to that received from the terminating switch, in order to find the least loaded two-link path to route the call over.

Three motivations for developing the new RTNR network are a) to support multiple classes-of-service on a dynamic class-of-service network, b) to move towards a robust routing strategy for self-healing networks, and c) to simplify operations support functions. These motivations are now briefly discussed. (a) It is desirable to introduce new services using dynamic routing, which is the routing strategy of choice for new services in the future. These services include 64 kbps, 384 kbps, and 1536 kbps switched digital services, international switched transit services, priority routing services,

(ORLANDO) GLOBECOM '92 CONF RECORD VOL 2

ISUP preferred service, international virtual private network, and others. Such needs have led to the concept of dynamic class-of-service routing for telecommunications networks, in which individual voice and data services utilize dynamic routing and share network bandwidth. (b) There clearly is benefit in developing a robust routing strategy for self-healing networks [19]. A self-healing network is one which responds in near real-time to a network failure, and continues to provide connections to customers with essentially no perceived interruption of service. Here adaptive routing shifts bandwidth rapidly among node pairs and services, and provides facility-diverse routing as well as multiple ingress/egress routing to help respond to link and node failures. Facility-diverse routing allows selection of diverse traffic paths through the transport network, and multiple ingress/egress routing allows selection of diverse traffic paths into and out of the switching network. (c) There is need to simplify the operational environment since operations costs are significant and sometimes dominate capital costs. There is also motivation to develop an efficient decentralized routing strategy which minimizes the need for centralized operations support, as a basic flexible routing strategy for the globalization of dynamic routing [19]. As demonstrated in this paper, these needs are met by the RTNR strategy for dynamic class-of-service networks.

In Section 2 we describe the RTNR platform for dynamic class-of-service networks and for flexible ingress/egress routing, in Section 3 we illustrate RTNR network performance with a spectrum of examples from actual network conditions, in Section 4 we discuss RTNR network operations and simplifications introduced, and in Section 5 we present a summary and conclusion.

2. DESCRIPTION OF RTNR FOR DYNAMIC CLASS-OF-SERVICE NETWORKS

First we briefly review DNHR operation, since DNHR is the routing system replaced by RTNR in 1991. DNHR [3-7,20] uses a hybrid time varying and real-time routing system to respond to network load variations, and incorporates one and two-link path routing between originating and terminating switches. Time varying routing allows pre-specified routing patterns to change as frequently as every hour to respond to expected traffic patterns, and supplementary real-time routing, which searches for idle capacity on a call-by-call basis, if needed. The real-time routing method appends to each sequence of two-link paths, engineered by the DNHR design algorithm [3] to meet the expected load, several additional two-link (real-time) paths to be used only when idle capacity is available. Dynamic trunk reservation is used to recognize idle network capacity on real-time paths. Access to trunks on a particular trunk group is allowed only when a specified number of trunks - the reservation level - is available, which guarantees minimal interference with engineered traffic. A Network Management Operations System (NEMOS), under conditions of network stress, is capable of generating up to seven additional paths to append to the 14 pre-planned DNHR engineered and real-time paths generated by the design. Network blocking performance is improved by both real-time routing and NEMOS reroutes under a variety of overload and failure conditions as demonstrated by network simulation studies [4] and network performance measurements [20].

In contrast to DNHR, switches using RTNR first select the direct trunk group between the originating switch, denoted here as OS_j, and the terminating switch, denoted here as TS_k. When no direct trunks are available, the originating switch checks the availability and load conditions of all of the two-link paths to TS_k on a per call basis. If any of these two-link paths are available, the call is set up over the least loaded two-link path. Traffic loads are dynamically balanced across trunks throughout the network to maximize the call throughput of the network. As illustrated in Figure 1, an available two-link path from OS_j to TS_k goes through a via switch

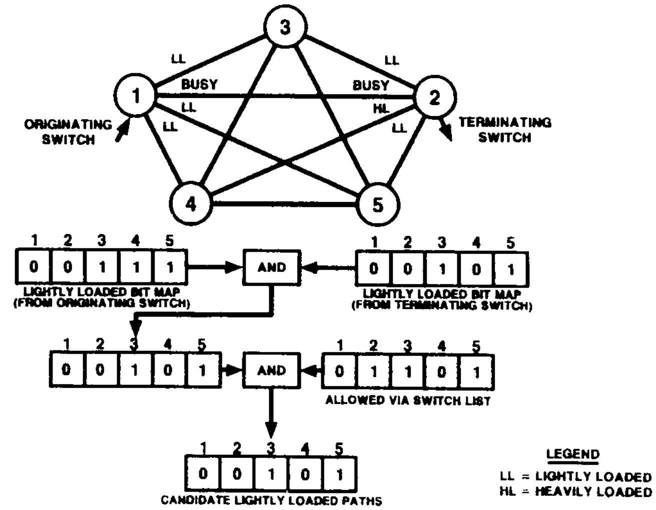

FIGURE 1. DETERMINING LEAST LOADED PATHS

to which both OS_j and TS_k have idle trunks, that is, neither link is busy. An available two-link path is considered to be lightly loaded if the number of idle trunks on both links exceeds a threshold level. In order to determine all of the switches in the network which satisfy this criterion, the originating switch sends a message to TS_k over the CCS network, requesting TS_k to send a list of the switches to which it has lightly loaded trunk groups. Upon receiving this list of switches from TS_k, the originating switch compares this list with its own list of switches to which it has lightly loaded trunk groups. Any switch that appears in both lists currently has lightly loaded trunk groups to both OS_j and TS_k, and therefore can be used as the via switch for a two-link connection for this call. In Figure 1 there are two lightly loaded paths found between OS_j and TS_k.

The switch identifiers used in the switch list sent by TS_k must be recognized by the originating switch. Each switch in the network is assigned a unique network switch number (NSN); these NSNs are used as switch identifiers. In the example depicted in Figure 1, there are five switches in a network which have been arbitrarily assigned NSNs. With these NSN assignments, a list of switches can be represented by a bit map that has a 1-bit entry for each NSN in the network. In Figure 1, a "1" entry is made in the bit map for each NSN having a lightly loaded trunk group to TS_k. OS_j also maintains its own bit map listing each NSN having a lightly loaded trunk group to OS_j. Using bit maps makes it very easy and efficient for the originating switch to find all lightly loaded two-link paths. The originating switch simply ANDs the bit map it receives from TS_k, which lists all of the lightly loaded trunk groups out of TS_k, with its own bit map to produce a new bit map which identifies all the via switches with lightly loaded trunk groups to both OS_j and TS_k. Bit maps also are a very compact way to store a list of switches. This is an important consideration since the list is sent in a CCS message. In the simplest case, only 16 bytes of data are needed for a network with 128 switches.

Since some of the available two-link paths may not provide good voice transmission quality, network administrators can restrict path selection to the two-link paths which provide good transmission quality through use of another bit map, the allowed via switch list, which specifies the acceptable via switches from OS_j to TS_k. As illustrated in Figure 1, ANDing this bit map with the bit map containing the via switches of all of the available two-link paths removes the via switches of paths with unacceptable transmission quality. When two or more available paths have the same load status, as in Figure 1 where two lightly loaded paths are found, the originating switch randomly picks one of these paths to use for the call. To pick a path for this particular call to TS_k, the originating switch starts a circular search through the bit map list of paths with the same load status, beginning with the entry immediately following the via switch it last used for a call to TS_k.

Requesting trunk status on a per call basis ensures that the current network conditions are known for selecting a two-link path for the call. However, sending a request message and waiting for the response adds to the set up time for the call. This additional call set up delay can be avoided. When the originating switch does not find any available direct trunks to TS_k, it requests the current list of switches to which TS_k has idle trunks. Rather than waiting for a response from TS_k to set up the call, the originating switch can select a two-link path using the most recently received status response from TS_k. When the new status response is received, the originating switch stores it away for use the next time it needs to select a two-link path to TS_k. Since the status of the idle trunks is not as current using this method, the two-link path selected has a greater probability of being blocked. When this happens, the via switch uses a CCS crankback message to return the call to the originating switch. The originating switch must then wait for the status response from TS_k, and pick a new via switch for the call using this up-to-date information. With these methods, RTNR uniformly and substantially reduces call set up delay in comparison to other dynamic routing strategies. Further details of RTNR operation are given in Reference [22].

2.1 RTNR Class-of-Service Routing

RTNR provides a platform to implement a dynamic class-of-service network, and therefore is used for voice services, 64 kbps data services, 384 kbps data services, and 1536 kbps data services. The simplicity of RTNR operation and the lack of routing data administration make RTNR the method of choice for dynamic class-of-service networks. As illustrated in Figure 2, the various classes-of-service share bandwidth on the network links. For this purpose the dynamic class-of-service network is engineered to handle the combined forecasted call loads for many classes-of-service. The network engineering process allots a certain number of direct virtual trunks between OS_j and TS_k to class-of-service i, which is referred to as $VTeng_k^i$. $VTeng_k^i$ is the minimum guaranteed bandwidth for class-of-service i, but if class-of-service i is meeting its GOS^i blocking objective, other classes-of-service are free to share the $VTeng_k^i$ bandwidth allotted to class-of-service i. The quantities $VTeng_k^i$ are chosen in the network engineering process such that their sum over all classes-of-service sharing the bandwidth of the direct link is equal to the total bandwidth on the link, as do the three $VTeng_k^i$ segments illustrated in Figure 2.

For each major class-of-service i, RTNR maintains node-to-node blocking rates NN_k^i, node-to-node traffic load estimates $VTtraf_k^i$, and current node-to-node calls-in-progress counts CIP_k^i. RTNR controls the use of direct and two-link paths based on parameters defining each "virtual network" for each class-of-service. A virtual network is comprised of the above traffic measurements, bandwidth allocation parameters ($VTeng_k^i$), routing priority, performance objective parameters, and voice/data transport capability. Through use of the virtual network parameters, RTNR implements a routing pattern which is able to meet different blocking objectives and load levels for different classes-of-service. For example, classes-of-service with different transport capabilities, e.g. voice and 384 kbps switched digital services, can have different blocking objectives, as can classes-of-service that require the same transport capabilities, e.g. domestic voice and international voice services. In this way, RTNR maximizes the performance of a dynamic class-of-service network in meeting blocking and call throughput objectives for all classes-of-service. For purposes of call establishment, RTNR class-of-service routing executes the following steps:

1. at the originating switch, the class-of-service, virtual network, and terminating switch TS_k are identified;
2. the class-of-service, virtual network, and TS_k information are used to select the corresponding routing pattern data, which includes voice/data transport, data rate, performance objectives, dynamic reservation thresholds, routing priority, bandwidth allocation, and traffic data registers; and
3. an appropriate path is selected through execution of dynamic path selection logic and possible exchange of network status bit maps, and the call is established on the selected path.

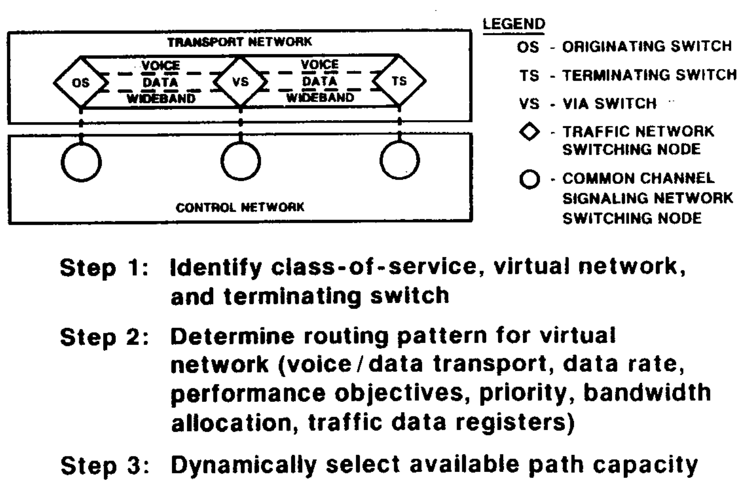

FIGURE 2. RTNR CLASS-OF-SERVICE ROUTING

Switch OS_j uses the quantities $VTtraf_k^i$, CIP_k^i, NN_k^i, and $VTeng_k^i$ to dynamically allocate link bandwidth to classes-of-service. Under normal non-blocking network conditions, all classes-of-service fully share all available capacity. Because of this, the network has the flexibility to carry a call overload between two switches for one class-of-service if the traffic load for other classes-of-service are sufficiently below their engineered levels. An extreme call overload between two switches for one class-of-service may cause calls for other classes-of-service to be blocked, in which case trunks are reserved to ensure that each class-of-service gets the amount of bandwidth allotted by the network engineering process. This reservation during times of overload results in network performance that is analogous to having a number of trunks dedicated between the two switches for each class-of-service. When blocking occurs for class-of-service i, RTNR trunk reservation is enabled to prohibit alternate routed traffic and traffic from other classes-of-service from seizing direct link capacity designed for class-of-service i. When the originating switch detects that the current blocking level NN_k^i for calls for class-of-service i exceeds the GOS^i objective, trunk reservation is triggered for direct traffic for class-of-service i, and the reserved bandwidth on the direct link for any class-of-service is at most $VTeng_k^i$ - CIP_k^i.

Sharing of bandwidth on the direct link is implemented by allowing calls for class-of-service i to always seize a virtual trunk on the direct link if the calls-in-progress CIP_k^i are below the level $VTeng_k^i$. But if CIP_k^i is equal to or greater than $VTeng_k^i$, then calls for class-of-service i can seize a virtual trunk on the direct link only when the idle bandwidth on the trunk group is greater than the bandwidth reserved by other classes-of-service that are not meeting their blocking objectives. That is, if

$$CIP_k^i \geq VTeng_k^i$$

and the idle link bandwidth $ILBW_k$ is greater than a reserved bandwidth threshold, then we select a virtual trunk on the direct link. Hence traffic for other classes-of-service is restricted from seizing direct link capacity being reserved for class-of-service i to meet its $VTeng_k^i$ engineered call load level. In this manner each class-of-service is assured a minimal level of network throughput determined by the network engineering process. When the number of calls-in-progress CIP_k^i exceeds $VTeng_k^i$, the engineered capacity allotted for class-of-service i is used up, and calls can then be routed to unreserved direct link capacity or to available via paths. Finally, when CIP_k^i exceeds the $VTtraf_k^i$, then reservation is no longer needed to meet the GOS^i objective for class-of-service i, and virtual trunks are shared by all traffic, including alternate routed traffic.

With the introduction of dynamic routing based on real-time network status, the network is more adaptive or robust in coping with traffic fluctuations and network failure situations. RTNR implements the concepts of facility-diverse routing, illustrated in Figure 3, and multiple ingress/egress routing, illustrated in Figure 4 and described in the following Section. Facility-diverse routing provides immediate access to all surviving capacity following a network facility failure. RTNR achieves a maximally efficient facility-diverse routing method. This is because with fixed routing node pairs such as A to D in Figure 3 can be isolated by a single facility cable cut, in the given example, while with RTNR all surviving capacity is immediately accessed, such as through nodes E and F.

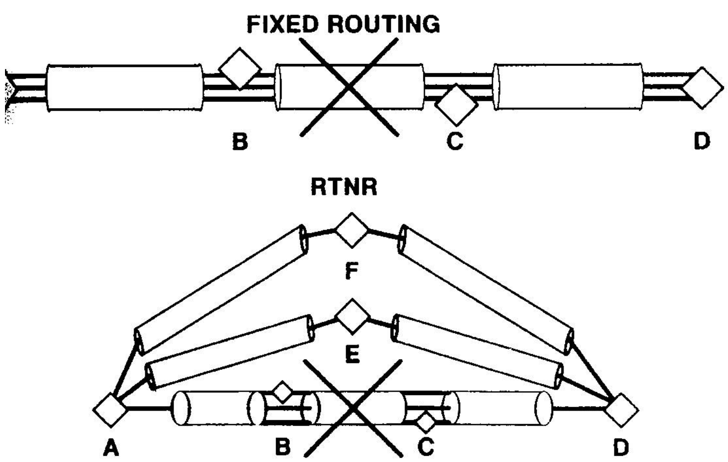

FIGURE 3. FACILITY DIVERSE ROUTING

2.2 RTNR Multiple Ingress/Egress Routing

Multiple ingress/egress routing also is incorporated in RTNR and as illustrated in Figure 4 allows international networks, toll switches, end-offices, or PBXs to be multiply-connected to more than one switch in the network in order to provide flexible access to all available connecting capacity, as well as robust routing diversity and protection from node failures, facility failures, or failures with the ingress/egress network.

Traditionally, the traffic to and from a particular geographical area goes to a particular switching node nearest to it in order to carry the access and egress traffic. For example, in Figure 4, if all the traffic for location P enters and exits the network through switch A, then if switch A fails location P gets isolated. To protect against this location P is connected to more than one switch, A and B in Figure 4. RTNR path selection might first be tried to switch A to reach location P, but if that is unsuccessful for whatever reason, the call is returned to the originating switch with CCS crankback and RTNR path selection is tried to switch B to reach location P. In a similar manner, international flexible routing to other national networks is achieved with this multiple ingress/egress routing concept. If a call from the originating switch in Figure 4 cannot reach the UK through switch A, perhaps because all international trunks from switch A to the UK are busy, then the call is returned to the originating switch via CCS crankback for RTNR path

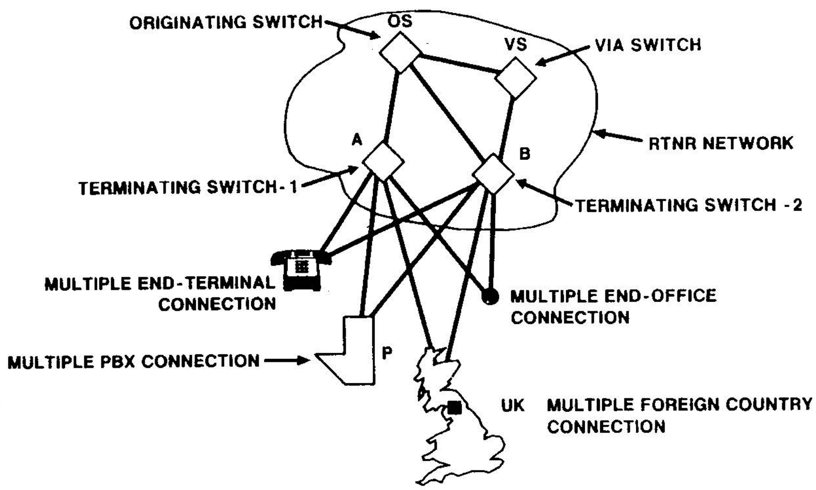

FIGURE 4. RTNR MULTIPLE INGRESS/EGRESS ROUTING

selection to switch B and then to the UK. Full access to international capacity provides additional throughput flexibility and robustness to failure.

Figure 5 compares multiple ingress/egress routing with RTNR to the previous routing method with DNHR. Note that in DNHR routing to the UK had a choke point. That is, traffic destined for the UK from Anaheim had to go through the Pittsburgh international switching center. Once such a call gets to Pittsburgh, it can alternate route through the New York NY55 switch, but that is the point of vulnerability. If the Pittsburgh switch fails, or if there is a facility failure into or out of Pittsburgh, a large part of the incoming and outgoing international traffic would be blocked. This vulnerability is eliminated by the multiple ingress/egress routing introduced by RTNR: a switch failure at Pittsburgh or a facility failure on that route can be circumvented by the additional route through the NY55 international switching center, leading to a significant improvement in network robustness for international traffic.

The methods used to achieve RTNR also are attractive for a worldwide dynamic network, in which a dynamic routing network is implemented among the countries of the world [19]. This is because, as pointed out in Reference [19], a global dynamic routing strategy should be simple, distributed in nature and not centralized, provide service flexibility for emerging services, have maximum flexibility in selecting transit paths through all possible via countries, and provide a robust response to failures and overloads. The methods used to achieve RTNR have promise in meeting all of these objectives in an efficient manner, and efforts to standardize the required CCS messages is being undertaken to achieve this direction.

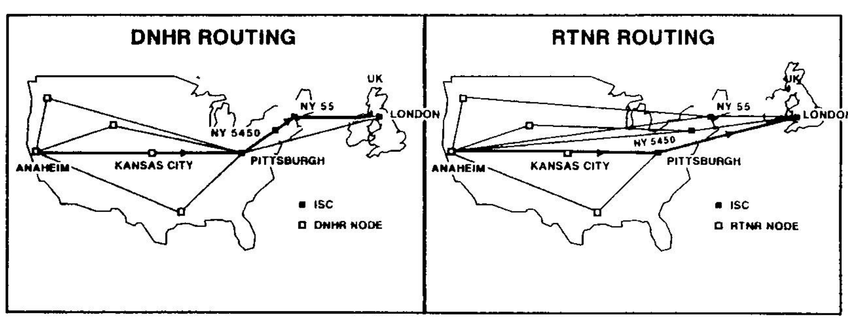

FIGURE 5. EXAMPLE OF RTNR MULTIPLE INGRESS/EGRESS ROUTING TO THE UNITED KINGDOM

determined by the network engineering process. When the number of calls-in-progress CIP^i_k exceeds $VTeng^i_k$, the engineered capacity allotted for class-of-service i is used up, and calls can then be routed to unreserved direct link capacity or to available via paths. Finally, when CIP^i_k exceeds the $VTtraf^i_k$, then reservation is no longer needed to meet the GOS^i objective for class-of-service i, and virtual trunks are shared by all traffic, including alternate routed traffic.

With the introduction of dynamic routing based on real-time network status, the network is more adaptive or robust in coping with traffic fluctuations and network failure situations. RTNR implements the concepts of facility-diverse routing, illustrated in Figure 3, and multiple ingress/egress routing, illustrated in Figure 4 and described in the following Section. Facility-diverse routing provides immediate access to all surviving capacity following a network facility failure. RTNR achieves a maximally efficient facility-diverse routing method. This is because with fixed routing node pairs such as A to D in Figure 3 can be isolated by a single facility cable cut, in the given example, while with RTNR all surviving capacity is immediately accessed, such as through nodes E and F.

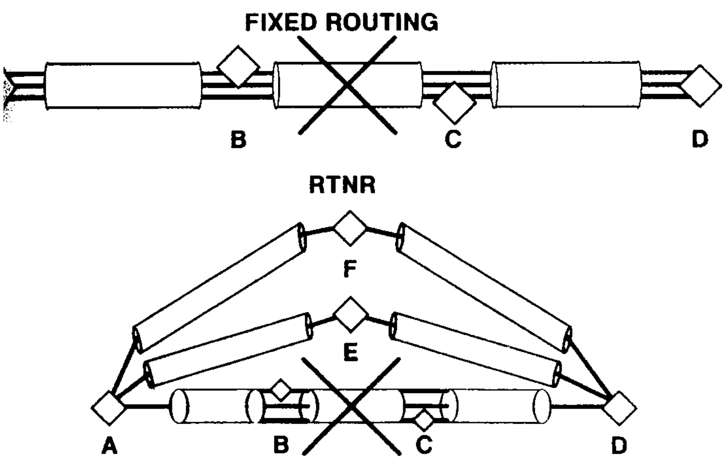

FIGURE 3. FACILITY DIVERSE ROUTING

2.2 RTNR Multiple Ingress/Egress Routing

Multiple ingress/egress routing also is incorporated in RTNR and as illustrated in Figure 4 allows international networks, toll switches, end-offices, or PBXs to be multiply-connected to more than one switch in the network in order to provide flexible access to all available connecting capacity, as well as robust routing diversity and protection from node failures, facility failures, or failures with the ingress/egress network.

Traditionally, the traffic to and from a particular geographical area goes to a particular switching node nearest to it in order to carry the access and egress traffic. For example, in Figure 4, if all the traffic for location P enters and exits the network through switch A, then if switch A fails location P gets isolated. To protect against this location P is connected to more than one switch, A and B in Figure 4. RTNR path selection might first be tried to switch A to reach location P, but if that is unsuccessful for whatever reason, the call is returned to the originating switch with CCS crankback and RTNR path selection is tried to switch B to reach location P. In a similar manner, international flexible routing to other national networks is achieved with this multiple ingress/egress routing concept. If a call from the originating switch in Figure 4 cannot reach the UK through switch A, perhaps because all international trunks from switch A to the UK are busy, then the call is returned to the originating switch via CCS crankback for RTNR path

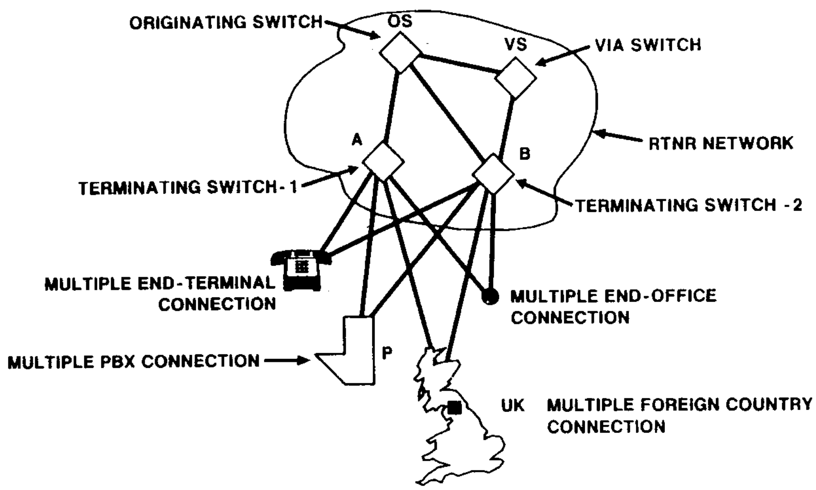

FIGURE 4. RTNR MULTIPLE INGRESS/EGRESS ROUTING

selection to switch B and then to the UK. Full access to international capacity provides additional throughput flexibility and robustness to failure.

Figure 5 compares multiple ingress/egress routing with RTNR to the previous routing method with DNHR. Note that in DNHR routing to the UK had a choke point. That is, traffic destined for the UK from Anaheim had to go through the Pittsburgh international switching center. Once such a call gets to Pittsburgh, it can alternate route through the New York NY55 switch, but that is the point of vulnerability. If the Pittsburgh switch fails, or if there is a facility failure into or out of Pittsburgh, a large part of the incoming and outgoing international traffic would be blocked. This vulnerability is eliminated by the multiple ingress/egress routing introduced by RTNR: a switch failure at Pittsburgh or a facility failure on that route can be circumvented by the additional route through the NY55 international switching center, leading to a significant improvement in network robustness for international traffic.

The methods used to achieve RTNR also are attractive for a worldwide dynamic network, in which a dynamic routing network is implemented among the countries of the world [19]. This is because, as pointed out in Reference [19], a global dynamic routing strategy should be simple, distributed in nature and not centralized, provide service flexibility for emerging services, have maximum flexibility in selecting transit paths through all possible via countries, and provide a robust response to failures and overloads. The methods used to achieve RTNR have promise in meeting all of these objectives in an efficient manner, and efforts to standardize the required CCS messages is being undertaken to achieve this direction.

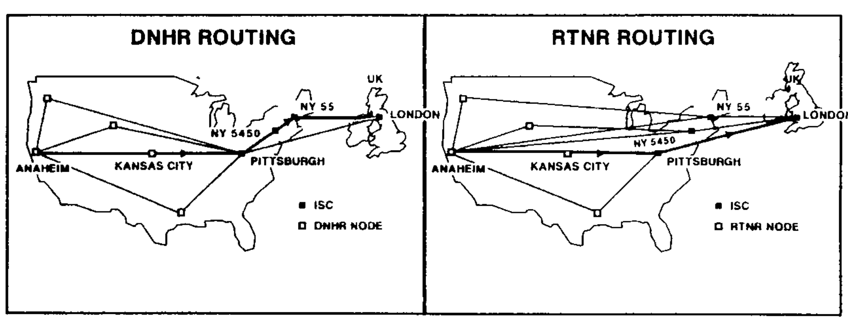

FIGURE 5. EXAMPLE OF RTNR MULTIPLE INGRESS/EGRESS ROUTING TO THE UNITED KINGDOM

3. RTNR NETWORK PERFORMANCE RESULTS

RTNR was introduced into the AT&T network over the 5 month period between March 16, 1991 and July 20, 1991, in a phased introduction, and over this time period completely replace DNHR in the AT&T network. This very large conversion process, which completely replaced DNHR technology with RTNR technology, was made on an actively working network without impacting any of the existing services. After RTNR was fully cut over in July 1991, eight virtual networks were established across the major classes-of-service listed in Table 1. As discussed in Section 2, these virtual networks share the facilities of the AT&T network, but provide individual control and monitoring of traffic by class-of-service, as well as providing new priority service capabilities. Multiple ingress/egress routing was implemented in the final stage, and in this stage international networks as well as other multiply-connected applications discussed in Section 2 were provided the added flexibility of RTNR.

TABLE 1.EIGHT VIRTUAL NETWORKS IMPLEMENTED BY RTNR CLASS-OF-SERVICE

Virtual Network No.	Class-of-Service
1	BUSINESS VOICE
2	LONG DISTANCE VOICE
3	INTERNATIONAL VOICE
4	64-KBPS DATA
5	384-KBPS DATA
6	1536-KBPS DATA
7	PRIORITY VOICE
8	PRIORITY 64-KBPS DATA

RTNR allows a marked improvement in network reliability by providing for the selection of over 100 routes between every pair of cities for every call. RTNR also enables the flexible ingress/egress routing arrangement to ensure reliability and flexibility for international and access networks. Implementation of RTNR represents the first nationally deployed adaptive routing network in the world, and has provided superb performance for our customers.

Now we illustrate performance experiences with RTNR.

3.1 Results for High-Day Traffic Loads

Table 2 illustrates RTNR network performance for the highest day load on the AT&T network to date, December 2, 1991, the Monday after Thanksgiving, which is normally the highest business calling day of the year, and it was. We also show the blocking performance of the network.

TABLE 2.NETWORK BLOCKING PERFORMANCE FOR HIGH-DAY TRAFFIC LOADS MONDAY AFTER THANKSGIVING -- 12/2/91

Virtual Network Number	Class-of-Service	Blocking Performance (% blk.)
1	BUSINESS VOICE	0.00
2	LONG DISTANCE VOICE	0.00
3	INTERNATIONAL VOICE	0.00
4	64-KBPS DATA	0.00
5	384-KBPS DATA	0.00
6	1536-KBPS DATA	0.00
7	PRIORITY VOICE	0.00
8	PRIORITY 64-KBPS DATA	0.00

Hence we see that for the Monday after Thanksgiving in 1991, which set a record for the number of calls on the AT&T network, 157.5 million, of which only 228 calls were blocked on AT&T facilities, providing a completion rate on AT&T facilities of 99.999% on the first try. This is in contrast to the Monday after Thanksgiving in 1990, when DNHR was fully implemented, in which the number of calls on the AT&T network was 136.5 million, and 17,086 were blocked on AT&T facilities. Hence with DNHR about 1 out of every 10,000 calls were blocked on the busiest day of the year, while with RTNR only 1 out of every 1,000,000 calls were blocked. This illustrates that the dynamic routing network under normal load conditions is virtually non-blocking. This means that dynamic routing, especially with RTNR, leads to additional service revenues as formerly blocked calls are completed, and greatly improves service quality to the customer. In addition the results show that voice and data services integration leads to excellent network performance with independent traffic control for each voice or data class-of-service, and at the same time providing efficient sharing of transport network capacity.

3.2 Results for Peak Day Traffic Loads

Peak days such as Christmas, Mother's Day, and Thanksgiving are the greatest tests for dynamic routing, because the traffic loads and patterns of traffic deviate most severely from normal business day traffic for which the network is designed. If the network performs well for such severe overload patterns, it can perform well for essentially any load pattern generated by customers with equal or better quality. This is an advantage in today's competitive environment where the traffic loads and the services that generate those loads can change rapidly. RTNR gives the ability to respond to those changes quickly, and with high service quality to the customer.

Figure 6 illustrates two examples of peak day performance: Thanksgiving and Christmas. In both of these cases, we contrast the performance after the conversion to RTNR in 1991, with the performance of DNHR after its full deployment in 1987, with the performance of the hierarchical (HIER) network in 1986, when there was a far smaller DNHR implementation. The performance of these three networks is based on measured data. AT&T's network is not engineered for peak days such as Christmas, it is engineered for normal average business and weekend traffic loads. On a peak day such as Christmas we try to complete as much traffic as possible, but often it is not possible to complete calls on the first try.

We can see from these results that the performance improvements of the DNHR network over the hierarchical network, and the RTNR network over the DNHR network, are dramatic. In the case

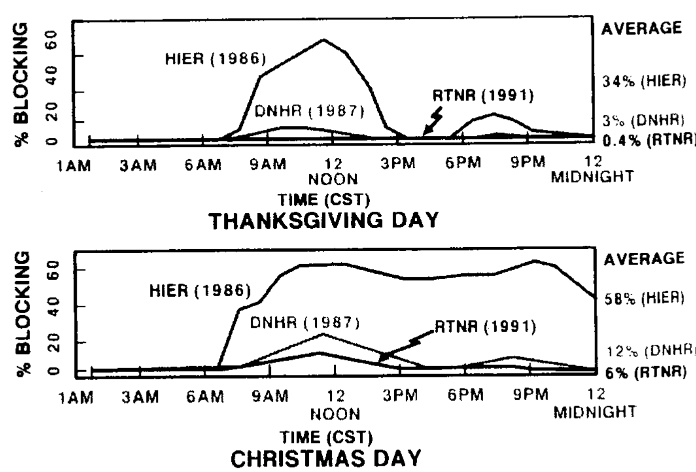

FIGURE 6. PEAK DAY NETWORK BLOCKING FOR HIERARCHICAL ROUTING (HIER, 1986), DYNAMIC NONHIERARCHICAL ROUTING (DNHR, 1987) AND REAL-TIME NETWORK ROUTING (RTNR, 1991)

of Thanksgiving, average networking blocking, which in 1986 had run 34%, was down to 3% in 1987, and then down to 0.4% in 1991, which is nearly a factor of 100 improvement. In particular, during the mid period of the day, the calling peak on Thanksgiving, blocking in 1986 had reached 60%, blocking in 1987 was down in the 10 to 15% region, and in 1991 down to the 2 to 3% region. Similarly on Christmas, which is the heaviest peak calling day of the year, blocking in 1986 had run at the 50 to 60% level, virtually throughout the entire day. In contrast, the 1987 DNHR network yielded Christmas Day performance where blocking was on the order of 20% at the mid-day peak, and the 1991 RTNR network had blocking on the order of 15% at the mid-day peak. During other portions of the day, blocking was down to just a few percent. The average blocking for Christmas day was down from 58% for the hierarchical network, to 12% for the DNHR network, to 6% for the RTNR network, which is almost a factor of 10 improvement. Also, in 1991 versus 1986, there were fewer trunks relative to demand, because of the efficiencies of dynamic routing design, and there was more traffic load.

3.3 Results for Unexpected Network Overloads

Table 3 illustrates the network blocking performance during the evening hours of August 19, 1991, when Hurricane Bob caused severe overloads in the Northeastern United States.

TABLE 3. NETWORK BLOCKING PERFORMANCE FOR UNEXPECTED TRAFFIC OVERLOADS HURRICANE BOB -- 8/19/91

Hour of Day	Normal Voice Services (% blk.)	Priority Voice Services (% blk.)
6 to 7 PM	0.04	0.00
7 to 8 PM	0.99	0.00
8 to 9 PM	0.47	0.00

As is clear from the results in Table 3, the priority service capability of the RTNR class-of-service network provides essentially zero blocking performance even under such severe network conditions.

3.4 Results for Network Failure

Figures 7 and 8 give simulation results for an actual cable cut in Sayreville, New Jersey, a cut of approximately 33,000 trunks. The results clearly demonstrate that RTNR not only significantly improves network throughput (Figure 7) but also is much better in

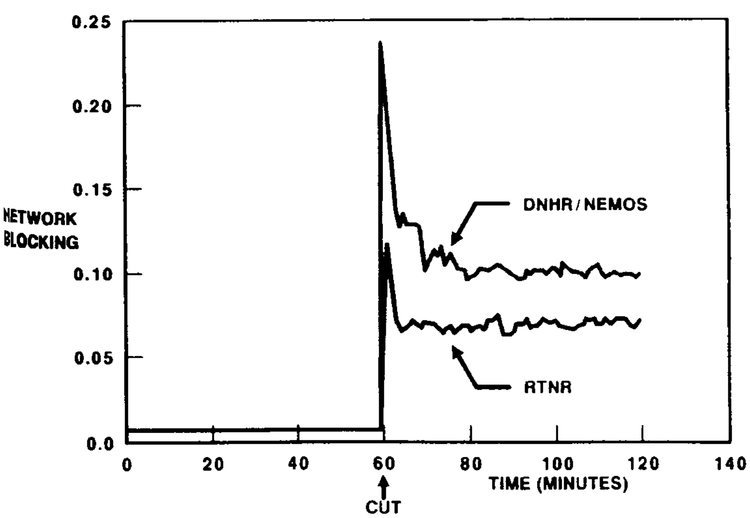

FIGURE 7. NETWORK BLOCKING PERFORMANCE COMPARISON FOR THE 11/18/88 CABLE CUT IN SAYREVILLE, NEW JERSEY

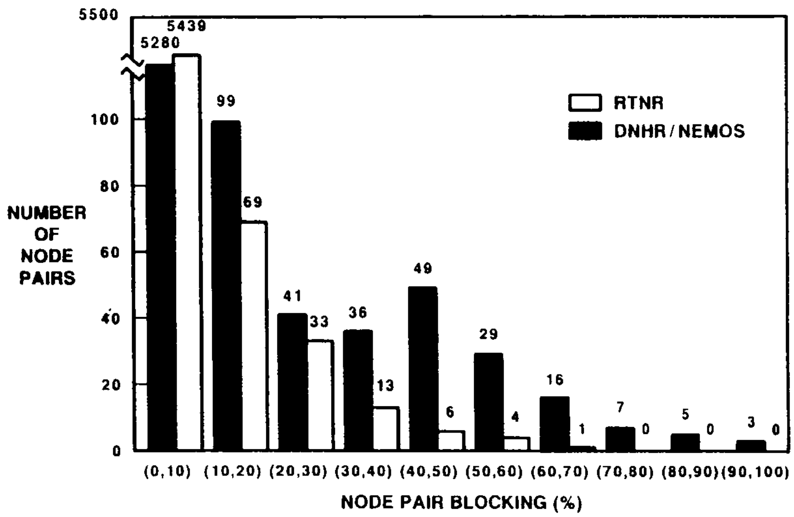

FIGURE 8. NETWORK BLOCKING DISTRIBUTION COMPARISON BETWEEN RTNR AND DNHR/NEMOS

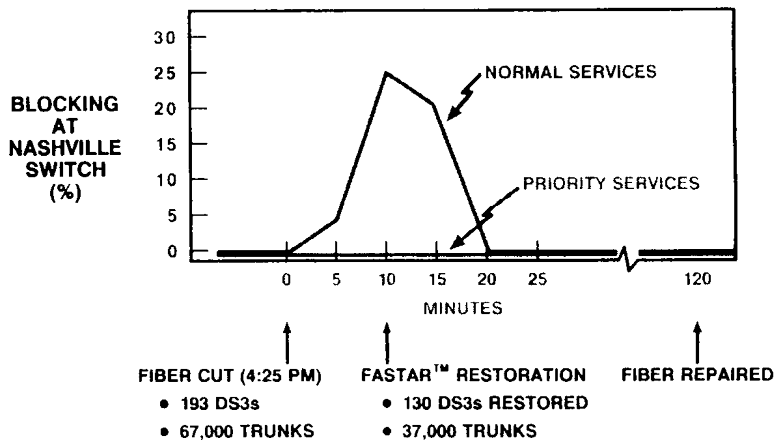

FIGURE 9. RTNR NETWORK PERFORMANCE AFTER FIBER CUT IN AUSTIN, IN (9/25/91)

preventing node pair traffic isolation (Figure 8). Call attempts also are greatly reduced which means that customers require many fewer attempts to get through. This self-healing property is very valuable for network reliability.

Figure 9 shows the performance of the network during a fiber cut which occurred at 4:26 p.m. EST on September 25, 1991. The fiber cut, which was caused by a bullet, was near Austin, IN, on a cable which runs between Columbus, IN and Louisville, KY. In all, more than 129 thousand circuits were lost due to the cut, and within these circuits, 67,000 trunks were lost in the AT&T switched network. An AT&T automatic facility restoration system, FASTAR, restored the first 672 circuits (one DS3) in 10 minutes, and the next 87 thousand circuits (130 DS3's) were restored within the next minute, which included 37,000 trunks in the switched network, leaving a total of 30,000 trunks still out of service in the switched network after the FASTAR restoration. Over the duration of this event, more than 12 thousand calls were blocked in the AT&T switched network, almost all of them originating or terminating at the Nashville, TN switch.

The blocking at the Nashville switch is plotted at 5 minute intervals in the Figure 9, and reaches almost 25 percent blocking at its peak. It is noteworthy that the blocking in the network returned to zero after the 37,000 trunks were restored in the first 11 minutes, even though there were 30,000 trunks still out of service. RTNR was able to find paths on which to complete traffic even though there were far fewer trunks than normal even after the FASTAR restoration. Hence RTNR in combination with FASTAR provides a significant self-healing capability for the AT&T network, and helps provide quality service for AT&T's customers. It is also noteworthy that the remaining circuits were restored in about two hours later. Hence without RTNR and FASTAR, degradation of service would have lasted far longer than it did. As discussed earlier, RTNR provides priority routing for selected customers and services, which permits priority calls to be routed in preference to other calls. As illustrated in Figure 9, blocking of the priority services is essentially zero throughout this entire network event.

3.5 Results for International Traffic With Multiple Ingress/Egress Routing

Table 4 illustrates the performance of RTNR multiple ingress/egress routing in comparison to DNHR performance for international traffic between the United States and Taiwan, during the Chinese New Year Celebration over the three-day period of heavy calling for that holiday.

TABLE 4.NETWORK BLOCKING PERFORMANCE COMPARISON FOR DNHR AND RTNR MULTIPLE INGRESS/EGRESS ROUTING TAIWAN -- CHINESE NEW YEAR (2/3-5/91, 2/14-16/92)

Day of Chinese New Year	DNHR Network Performance (1991) (% blk.)	RTNR Multiple Ingress/Egress Network Performance (1992) (% blk.)
Day 1	73.2	31.6
Day 2	44.6	6.0
Day 3	6.0	0.0
Average	53.6	19.3

We see from these results the considerable improvement in the quality of service following the introduction of RTNR multiple ingress/egress routing for Chinese New Year traffic to Taiwan.

4. RTNR NETWORK OPERATIONS

In order to design, administer, and manage the RTNR network, we use the family of operations systems illustrated in Figure 10. The switch planning system determines the switch capacity needed in the network as demands change over time. Switch planning is done on a multi-month cycle. To design trunk capacity we use a trunk engineering system which performs the design function for the trunking portion of the network. The network servicing system analyzes the latest traffic data every week and then re-optimizes the bandwidth allocation and sizing of the network to accommodate on-going changes in demands from that engineered semi-annually. Real time operation is controlled from the network operations center in Bedminster, New Jersey, using the NEMOS system (Network Management Operations System). NEMOS gathers data from the RTNR network every 5 minutes, and allows network managers to put in manual controls to deal with such events as failures, overloads, and problems that cannot be foreseen in advance. All of these systems interface with the RTNR network to reliably collect data and download information to the control elements of the network, in some cases, in near real time.

With the introduction of RTNR, it was important to forecast and design the network to achieve the kind of efficiencies that are possible with adaptive routing. With the inception of RTNR, a network design model known as the RTNR unified algorithm was developed by Bell Laboratories. The unified algorithm is a simplified yet accurate and efficient technique to design the RTNR network. An important simplification introduced with the design of RTNR networks is that routing patterns are not calculated by

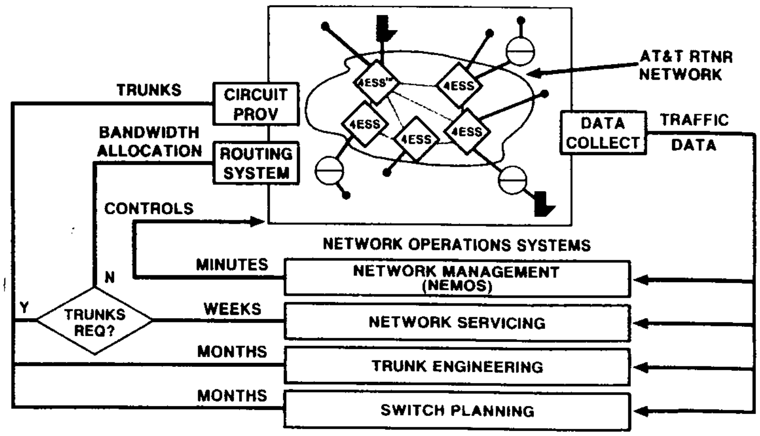

FIGURE 10. RTNR NETWORK OPERATIONS

the design algorithm, since these are computed in real-time in the switch. This led to simplifications introduced in the RTNR unified algorithm in that the routing patterns computed in DNHR no longer were needed. In addition to improving service quality, RTNR also brings reduced network capital requirements. That is, the RTNR network can carry the same traffic load as could the DNHR network, but with 2 to 3% fewer facilities in the network, based on the RTNR unified algorithm design.

RTNR introduces simplifications into the servicing of the network. Under DNHR, routing tables had to be periodically reoptimized in network servicing and downloaded into switching systems via the automated routing administration system. Reoptimizing and changing the routing tables in the network represented a large administrative effort involving millions of records. This function is eliminated by RTNR because RTNR does not have routing tables: the routing is generated in real-time for each call, and then discarded. Also since RTNR adapts to network conditions, less network demand servicing is required. This is one of the major operational advantages of RTNR; that is, to automatically adapt the traffic routing so as to move the traffic load to where the capacity is available in the network.

The management of the RTNR network is similar to DNHR in many respects, but also has some significant differences. The RTNR network, like the DNHR network, is managed as a single national entity. With RTNR, some new network displays and controls were implemented and are used at the network operations center, where the video displays are changed through software updates to the NEMOS system. Under DNHR, NEMOS automatically put in reroutes to solve blocking problems. The reroute algorithm in NEMOS was able to look everywhere in the network for additional available capacity, and add up to 7 additional routes to the 14 existing routes used by DNHR, on a 5 minute basis. RTNR replaces this automatic rerouting function that NEMOS performed under DNHR. That is, because RTNR examines all possible one and two-link paths, the NEMOS reroute function provides no additional two-link choices not examined by RTNR, and hence is not required. This is yet another operational simplification introduced under RTNR.

To summarize the impact of RTNR on network operations, RTNR allows reduced operations costs since with RTNR a number of operations have been simplified or eliminated. These simplifications include eliminating the storing of voluminous routing tables in the network switches, eliminating the calculation of routing tables in network design and forecasting, eliminating the administrative routing operations which required downloaded new routing information as part of network servicing, and eliminating the automatic rerouting function in network management; these functions are all unnecessary with RTNR. In addition to the cost and expense savings realized with these simplifications, RTNR network design with the unified algorithm allows capital cost reduction which can be attained through the more efficient operation of the RTNR network.

5. SUMMARY & CONCLUSION

We have described RTNR and why AT&T introduced this new dynamic routing strategy. RTNR brings benefits to AT&T customers in terms of new service flexibility and significantly higher service quality and reliability, at reduced cost. RTNR implements a service-based dynamic class-of-service routing feature for extending dynamic routing to emerging services, and provides a self-healing network capability to ensure a network-wide path selection and immediate adaptation to failure. We have presented the network results across a spectrum network conditions, which include high day loads on the Monday after Thanksgiving 1991, peak day loads on Thanksgiving and Christmas 1991, a cable cut near Austin, Indiana on September 25, 1991, a focused overload caused by Hurricane Bob on August 19, 1991, and the international network performance to Taiwan during the Chinese New Year three-day holiday celebration after the introduction of multiple ingress/egress routing to that country. We conclude that RTNR a) improves network performance by lowering blocking and call set up delay, b) simplifies the operations environment and lowers operations cost, c) simplifies routing data structures and their administration, d) allows sharing of voice and data network capacity in the deployment of ISDN, and e) provides a possible flexible routing direction for the globalization of dynamic routing.

REFERENCES

[1] Narendra, K. S., Wright, E. A., Mason, L. G., "Application of Learning Automata to Telephone Traffic Routing and Control," IEEE Transactions on Systems, Man, and Cybernetics, Volume SMC-7, Number 11, November 1977.

[2] Cameron, W. H., Galloy, P., Graham, W. J., "Report on the Toronto Advance Routing Concept Trial," Proceedings of the Networks Conference, Paris, France, 1980.

[3] Ash, G. R., Cardwell, R. H., Murray, R. P., "Design and Optimization of Networks with Dynamic Routing," Bell System Technical Journal, Volume 60, No. 8, October 1981.

[4] Ash, G. R., Kafker, A. H., Krishnan, K. R., "Servicing and Real-Time Control of Networks with Dynamic Routing," Bell System Technical Journal, Volume 60, No. 8, October 1981.

[5] Field, F. A., "The Benefits of Dynamic Nonhierarchical Routing in Metropolitan Traffic Networks," Proceedings of the Tenth International Teletraffic Congress, Montreal, June 1983.

[6] Ash, G. R., Kafker, A. H., Krishnan, K. R., "Intercity Dynamic Routing Architecture and Feasibility," Proceedings of the Tenth International Teletraffic Congress, Montreal, June 1983.

[7] Carroll, J. J., DiCarlo-Cottone, M. J., Kafker, A. H., "4ESS™ Switch Implementation of Dynamic Nonhierarchical Routing," Proceedings of the IEEE Global Telecommunications Conference, San Diego, Ca., December 1983.

[8] Ash, G. R., "Use of a Trunk Status Map for Real-Time DNHR," Proceedings of the Eleventh International Teletraffic Congress, Kyoto, Japan, September 1985.

[9] Gibbens, R. J. "Some Aspects of Dynamic Routing in Circuit-Switched Telecommunications Networks," Statistics Laboratory, University of Cambridge, January 1986.

[10] Chemouil, P., Filipiak, J., Gauthier, P., "Analysis and Control of Traffic Routing in Circuit-Switched Networks," Computer Networks and ISDN Systems, Volume 11, Number 3, March 1986.

[11] Hurley, B. R., Seidl, C. J. R., Sewell, W. F., "A Survey of Dynamic Routing Methods for Circuit-Switched Traffic," IEEE Communications Magazine, Volume 25, 1987.

[12] Gauthier, P., Chemouil, P., Klein, M., "STAR: A System to Test Adaptive Routing in France," Proceedings of the IEEE Global Telecommunications Conference, Tokyo, Japan, 1987.

[13] Stacey, R. R., Songhurst, D. J., "Dynamic Alternative Routing in the British Telecom Trunk Network," Proceedings of the International Switching Symposium, Phoenix, Arizona, March 1987.

[14] Watanabe, Y., Mori, H., "Dynamic Routing Schemes for International ISDNs," Proceedings of the Fifth ITC Specialists Seminar: Traffic Engineering for ISDN Design and Planning, Lake Como, Italy, May 1987.

[15] Caron, F. "Results of the Telecom Canada High Performance Routing Trial," Proceedings of the Twelfth International Teletraffic Congress, Torino, Italy, June 1988.

[16] Ash, G. R., Blake, B. M., Schwartz, S. D., "Integrated Network Routing and Design," Proceedings of the Twelfth International Teletraffic Congress, Torino, Italy, June 1988.

[17] Krishnan, K. R., Ott, T. J., "State-Dependent Routing for Telephone Traffic: Theory and Results," Proceedings of the 25th Conference on Decision and Control, Athens, Greece, December 1988.

[18] Ash, G. R., Schwartz, S. D., "Network Routing Evolution," Network Management and Control Workshop, Tarrytown, N.Y., September 1989.

[19] Ash, G. R., Chemouil, P., Kashper, A. N., Katz, S. S., Yamazaki, K., Watanabe, Y., "Robust Design and Planning of a Worldwide Intelligent Network," IEEE Journal on Selected Areas in Communications, Vol. 7, No. 8, October 1989.

[20] Ash, G. R., Oberer, E., "Dynamic Routing in the AT&T Network - Improved Service Quality at Lower Cost," Proceedings of the IEEE Global Telecommunications Conference, Dallas, Texas, November 1989.

[21] Mitra, D., Seery, J. B., "Comparative Evaluations of Randomized and Dynamic Routing Strategies for Circuit-Switched Networks," IEEE Transactions on Communications, January 1991.

[22] Ash, G. R., Chen, J-S., Frey, A. E., Huang, B. D., "Real-Time Network Routing in a Dynamic Class-of-Service Network," Proceedings of the Thirteenth International Teletraffic Congress, Copenhagen, Denmark, June 1991.

[23] Langlois, F., Regnier, J., "Dynamic Congestion Control in Circuit-Switched Telecommunications Networks," Proceedings of the Thirteenth International Teletraffic Congress, Copenhagen, Denmark, June 1991.

[24] Mitra, D., Gibbens, R. J., Huang, B. D., "Analysis and Optimal Design of Aggregated-Least-Busy-Alternative Routing on Symmetric Loss Networks with Trunk Reservation," Proceedings of the Thirteenth International Teletraffic Congress, Copenhagen, Denmark, June 1991.