STATUS OF THIS MEMO:

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt. The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html. Distribution of this memo is unlimited.

ABSTRACT

This is an informational document submitted in response to a request from the IETF Traffic Engineering Working Group (TEWG) for service provider uses, requirements, and desires for traffic engineering best current practices. As such, the work sets a direction for routing and traffic performance management in networks based on traffic engineering (TE) and QoS best current practices and operational experience, such as used in the AT&T dynamic routing/class-of-service network [A98]. Analysis models are used to demonstrate that these currently operational TE/QoS methods and best current practices are extensible to Internet TE and packet networks in general.

The draft describes, analyzes, and recommends TE methods which control a network's response to traffic demands and other stimuli, such as link failures or node failures. The functions discussed and recommendations made are consistent with the definition of TE employed by the TEWG:

Internet Traffic Engineering is concerned with the performance optimization of operational networks. It encompasses the measurement, modeling, characterization, and control of Internet traffic, and the application of techniques to achieve specific performance objectives, including the reliable and expeditious movement of traffic through the network, the efficient utilization of network resources, and the planning of network capacity.

This definition of TE methods is somewhat inconsistent with the telecom usage of the term "traffic engineering", which has more to do with network dimensioning and capacity planning. While these functions are encompassed by the above definition of TE, the scope of the analysis and recommendations
here goes well beyond dimensioning and includes call and connection routing, QoS resource management, routing table management, dynamic transport routing, capacity management, and operational requirements.

The recommended TE methods are meant to apply to IP-based, ATM-based, and TDM-based networks, as well as the interworking between these network technologies. Essentially all of the methods recommended are already widely applied in operational networks worldwide, particularly in PSTN networks employing TDM-based technology. However, the TE methods are shown to be extensible to packet-based technologies, that is, to IP-based and ATM-based technologies, and it is important that networks which evolve to employ these packet technologies have a sound foundation of TE methods to apply. Hence, it is the intent that the recommended TE methods in this draft be used as a basis for requirements for TE methods, and, as needed, for protocol development in IP-based, ATM-based, and TDM-based networks to implement the TE methods.

Hence the TE methods encompassed in this draft include:

* traffic management through control of routing functions, which include call routing (number/name translation to routing address), connection routing, QoS resource management, routing table management, and dynamic transport routing.

* capacity management through control of network design, including routing design.

* TE operational requirements for traffic management and capacity management, including forecasting, performance monitoring, and short-term network adjustment.

Results of analysis models are presented which illustrate the tradeoffs between various TE approaches. Based on the results of these studies as well as established practice and experience, TE methods are recommended for consideration in network evolution to IP-based, ATM-based, and/or TDM-based technologies.

*****************************************************************************
NOTE: A PDF VERSION OF THIS DRAFT (WITH FIGURES & TABLES) IS AVAILABLE AT http://www.research.att.com/~jrex/jerry/
*****************************************************************************

Ash <draft-ietf-tewg-qos-routing-02.txt> [Page MAIN-2]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001

TABLE OF CONTENTS

ABSTRACT

1.0 Introduction
2.0 Definitions
3.0 Traffic Engineering Model
4.0 Traffic Models
5.0 Traffic Management Functions
6.0 Capacity Management Functions
7.0 Traffic Engineering Operational Requirements
8.0 Traffic Engineering Modeling & Analysis
9.0 Conclusions/Recommendations
9.1 Conclusions/Recommendations on Call Routing & Connection Routing Methods (ANNEX 2)
9.2 Conclusions/Recommendations on QoS Resource Management (ANNEX 3)
9.3 Conclusions/Recommendations on Routing Table Management Methods & Requirements (ANNEX 4)
9.4 Conclusions/Recommendations on Dynamic Transport Routing Methods (ANNEX 5)
9.5 Conclusions/Recommendations on Capacity Management Methods (ANNEX 6)
9.6 Conclusions/Recommendations on TE Operational Requirements (ANNEX 7)
10.0 Recommended TE/QoS Methods for Multiservice Networks
10.1 Recommended Application-Layer IP-Network-Based Service-Creation Capabilities
10.2 Recommended Call/IP-Flow Control Layer Capabilities
10.3 Recommended Connection/Bearer Control Layer Capabilities
10.4 Recommended Transport Routing Capabilities
10.5 Recommended Network Operations Capabilities
10.6 Benefits of Recommended TE/QoS Methods for Multiservice Networks
11.0 Acknowledgements
12.0 Authors’ Addresses
13.0 Copyright Statement

ANNEX 1. Bibliography

ANNEX 2. Call Routing & Connection Routing Methods
2.1 Introduction
2.2 Call Routing Methods
2.3 Connection (Bearer-Path) Routing Methods
2.4 Hierarchical Fixed Routing (FR) Path Selection
2.5 Time-Dependent Routing (TDR) Path Selection
2.6 State-Dependent Routing (SDR) Path Selection
2.7 Event-Dependent Routing (EDR) Path Selection
2.8 Interdomain Routing
2.9 Modeling of Traffic Engineering Methods
2.9.1 Network Design Comparisons
2.9.2 Network Performance Comparisons
2.9.3 Single-Area Flat Topology versus Multi-Area Hierarchical Network Topology
2.9.4 Network Modeling Conclusions
2.10 Conclusions/Recommendations

ANNEX 3. QoS Resource Management Methods
3.1 Introduction
3.2 Class-of-Service Identification, Policy-Based Routing Table Derivation, & QoS Resource Management Steps
3.2.1 Class-of-Service Identification
3.2.2 Policy-Based Routing Table Derivation
3.2.3 QoS Resource Management Steps
3.3 Dynamic Bandwidth Allocation, Protection, and Reservation Principles
3.4 Per-Virtual-Network Bandwidth Allocation, Protection, and Reservation
3.4.1 Per-VNET Bandwidth Allocation/Reservation - Meshed Network Case
3.4.2 Per-VNET Bandwidth Allocation/Reservation - Sparse Network Case
3.5 Per-Flow Bandwidth Allocation, Protection, and Reservation
3.5.1 Per-Flow Bandwidth Allocation/Reservation - Meshed Network Case
3.5.2 Per-Flow Bandwidth Allocation/Reservation - Sparse Network Case
3.6 Packet-Level Traffic Control
3.7 Other QoS Resource Management Constraints
3.8 Interdomain QoS Resource Management
3.9 Modeling of Traffic Engineering Methods
3.9.1 Performance of Bandwidth Reservation Methods
3.9.2 Multiservice Network Performance: Per-VNET vs. Per-Flow Bandwidth Allocation
3.9.3 Multiservice Network Performance: Single-Area Flat Topology vs. Sparse Area Hierarchical Topology

Annex 4: Conclusions/Recommendations on Routing Table Management Methods & Requirements
Annex 5: Conclusions/Recommendations on Dynamic Transport Routing Methods
Annex 6: Conclusions/Recommendations on Capacity Management Methods
Annex 7: Conclusions/Recommendations on TE Operational Requirements
Annex 8: Recommended TE/QoS Methods for Multiservice Networks
Annex 9: Conclusions/Recommendations on TE Operational Requirements for Multiservice Networks
Annex 10: Benefits of Recommended TE/QoS Methods for Multiservice Networks
Annex 11: Acknowledgements
Annex 12: Authors’ Addresses
Annex 13: Copyright Statement

Annex 2: Call Routing & Connection Routing Methods
2.1 Introduction
2.2 Call Routing Methods
2.3 Connection (Bearer-Path) Routing Methods
2.4 Hierarchical Fixed Routing (FR) Path Selection
2.5 Time-Dependent Routing (TDR) Path Selection
2.6 State-Dependent Routing (SDR) Path Selection
2.7 Event-Dependent Routing (EDR) Path Selection
2.8 Interdomain Routing
2.9 Modeling of Traffic Engineering Methods
2.9.1 Network Design Comparisons
2.9.2 Network Performance Comparisons
2.9.3 Single-Area Flat Topology versus Multi-Area Hierarchical Network Topology
2.9.4 Network Modeling Conclusions
2.10 Conclusions/Recommendations

3.1 Introduction
3.2 Class-of-Service Identification, Policy-Based Routing Table Derivation, & QoS Resource Management Steps
3.2.1 Class-of-Service Identification
3.2.2 Policy-Based Routing Table Derivation
3.2.3 QoS Resource Management Steps
3.3 Dynamic Bandwidth Allocation, Protection, and Reservation Principles
3.4 Per-Virtual-Network Bandwidth Allocation, Protection, and Reservation
3.4.1 Per-VNET Bandwidth Allocation/Reservation - Meshed Network Case
3.4.2 Per-VNET Bandwidth Allocation/Reservation - Sparse Network Case
3.5 Per-Flow Bandwidth Allocation, Protection, and Reservation
3.5.1 Per-Flow Bandwidth Allocation/Reservation - Meshed Network Case
3.5.2 Per-Flow Bandwidth Allocation/Reservation - Sparse Network Case
3.6 Packet-Level Traffic Control
3.7 Other QoS Resource Management Constraints
3.8 Interdomain QoS Resource Management
3.9 Modeling of Traffic Engineering Methods
3.9.1 Performance of Bandwidth Reservation Methods
3.9.2 Multiservice Network Performance: Per-VNET vs. Per-Flow Bandwidth Allocation
3.9.3 Multiservice Network Performance: Single-Area Flat Topology vs. Sparse Area Hierarchical Topology
Multi-Area 2-Level Hierarchical Topology
3.9.4 Multiservice Network Performance: Need for MPLS & DiffServ
3.10 Conclusions/Recommendations

ANNEX 4. Routing Table Management Methods & Requirements
4.1 Introduction
4.2 Routing Table Management for IP-Based Networks
4.3 Routing Table Management for ATM-Based Networks
4.4 Routing Table Management for TDM-Based Networks
4.5 Signaling and Information Exchange Requirements
4.5.1 Call Routing (Number Translation to Routing Address) Information-Exchange Parameters
4.5.2 Connection Routing Information-Exchange Parameters
4.5.3 QoS Resource Management Information-Exchange Parameters
4.5.4 Routing Table Management Information-Exchange Parameters
4.5.5 Harmonization of Information-Exchange Standards
4.5.6 Open Routing Application Programming Interface (API)
4.6 Examples of Internetwork Routing
4.6.1 Internetwork E Uses a Mixed Path Selection Method
4.6.2 Internetwork E Uses a Single Path Selection Method
4.7 Modeling of Traffic Engineering Methods
4.8 Conclusions/Recommendations

ANNEX 5. Transport Routing Methods
5.1 Introduction
5.2 Dynamic Transport Routing Principles
5.3 Dynamic Transport Routing Examples
5.4 Reliable Transport Routing Design
5.4.1 Transport Link Design Models
5.4.2 Node Design Models
5.5 Modeling of Traffic Engineering Methods
5.5.1 Dynamic Transport Routing Capacity Design
5.5.2 Performance for Network Failures
5.5.3 Performance for General Traffic Overloads
5.5.4 Performance for Unexpected Overloads
5.5.5 Performance for Peak-Day Traffic Loads
5.6 Conclusions/Recommendations

ANNEX 6. Capacity Management Methods
6.1 Introduction
6.2 Link Capacity Design Models
6.3 Shortest Path Selection Models
6.4 Multihour Network Design Models
6.4.1 Discrete Event Flow Optimization (DEFO) Models
6.4.2 Traffic Load Flow Optimization (TLFO) Models
6.4.3 Virtual Trunking Flow Optimization (VTFO) Models
6.6 Day-to-day Load Variation Design Models
6.6 Forecast Uncertainty/Reserve Capacity Design Models
6.7 Meshed, Sparse, and Dynamic-Transport Design Models
6.8 Modeling of Traffic Engineering Methods
6.8.1 Per-Virtual-Network vs. Per-Flow Network Design
6.8.2 Integrated vs. Separate Voice/ISDN & Data Network Designs
6.8.3 Multilink vs. 2-Link Network Design
6.8.4 Single-area Flat vs. 2-Level Hierarchical Network Design
6.8.5 EDR vs. SDR Network Design
6.8.6 Dynamic Transport Routing vs. Fixed Transport Routing Network Design
6.9 Conclusions/Recommendations

ANNEX 7. Traffic Engineering Operational Requirements
1.0 Introduction

This is an informational document submitted in response to a request from the IETF Traffic Engineering Working Group (TEWG) for service provider uses, requirements, and desires for traffic engineering best current practices. As such, the work sets a direction for routing and traffic performance management in networks based on traffic engineering (TE) and QoS best current practices and operational experience, such as used in the AT&T dynamic routing/class-of-service network [A98]. Analysis models are used to demonstrate that these currently operational TE/QoS methods and best current practices are extensible to Internet TE and packet networks in general.

TE is an indispensable network function which controls a network’s response to traffic demands and other stimuli, such as network failures. TE encompasses

* traffic management through control of routing functions, which include number/name translation to routing address, connection routing, routing table management, QoS resource management, and dynamic transport routing.
* capacity management through control of network design.

Current and future networks are rapidly evolving to carry a multitude of voice/ISDN services and packet data services on Internet protocol (IP), asynchronous transfer mode (ATM), and time division multiplexing (TDM) networks. The long awaited data revolution is occurring, with the extremely rapid growth of data services such as IP-multimedia and frame-relay services. Within these categories of networks and services supported by IP, ATM, and TDM protocols have evolved various TE methods. The TE mechanisms
are covered in the draft, and a comparative analysis and performance evaluation of various TE alternatives is presented. Finally, operational requirements for TE implementation are covered.

We begin this document with a general model for TE functions, which include traffic management and capacity management functions responding to traffic demands on the network. We then present a traffic-variations model which these TE functions are responding to. Next we outline traffic management functions which include call routing (number/name translation to routing address), connection or bearer-path routing, QoS resource management, routing table management, and dynamic transport routing. These traffic management functions are further developed in ANNEXES 2, 3, 4, and 5. We then outline capacity management functions, which are further developed in ANNEX 6. Finally we briefly summarize TE operational requirements, which are further developed in ANNEX 7.

In ANNEX 2, we present models for call routing, which entails number/name translation to a routing address associated with service requests, and also compare various connection (bearer-path) routing methods. In ANNEX 3, we examine QoS resource management methods in detail, and illustrate per-flow versus per-virtual-network (or per-traffic-trunk or per-bandwidth-pipe) resource management and the realization of multiservice integration with priority routing services. In ANNEX 4, we identify and discuss routing table management approaches. This includes a discussion of TE signaling and information exchange requirements needed for interworking across network types, so that the information exchange at the interface is compatible across network types. In ANNEX 5 we describe methods for dynamic transport routing, which is enabled by the capabilities such as optical cross-connect devices, to dynamically rearrange transport network capacity. In ANNEX 6 we describe principles for TE capacity management, and in ANNEX 7 we present TE operational requirements.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

2.0 Definitions

Alternate Path Routing: a routing technique where multiple paths, rather than just the shortest path, between a source node and a destination node are utilized to route traffic, which is used to distribute load among multiple paths in the network;

Autonomous System: a routing domain which has a common administrative authority and consistent internal routing policy. An AS may employ multiple intradomain routing protocols and interfaces to other ASs via a common interdomain routing protocol;

Blocking: refers to the denial or non-admission of a call or connection-request, based for example on the lack of available resources on a particular link (e.g., link bandwidth or queuing resources);

Call: a generic term to describe the establishment, utilization, and release of a connection (bearer path) or data flow;

Call Routing: number (or name) translation to routing address(es), perhaps involving use of network servers or intelligent network (IN)
databases for service processing;

Circuit Switching
denotes the transfer of an individual set of
bits within a TDM time-slot over a
connection between an input port and an
output port within a given circuit-switching node
through the circuit-switching fabric (see Switching);

Class of Service
describes characteristics of a service such as
link capability requirements, QoS & traffic
threshold parameters;

Connection:
bearer path, label switched path, virtual circuit,
and/or virtual path established
by call routing and connection routing;

Connection Admission
a process by which it is determined whether a link
or a node has sufficient resources
to satisfy the QoS required for a connection or
flow. CAC is typically applied by each
node in the path of a connection or flow
during set-up to check local resource
availability;

Connection Routing: connection establishment through selection of one
path from path choices governed
by the routing table;

Crankback
a technique where a connection or flow setup is
backtracked along the
call/connection/flow path up to the first
node that can determine an alternative path to
the destination node;

Destination Node:
terminating node within a given network;

Flow:
bearer traffic associated with a given connection or
connectionless stream having the
same originating node, destination node, class of
service, and session identification;

GoS (grade of service)
a number of network design variables used to provide
a measure of adequacy
of a group of resources under specified
conditions (e.g., GoS variables may be
probability of loss, dial tone delay, etc.);

GoS standards
parameter values assigned as objectives for GoS
variables;

Integrated Services:
a model which allows for integration of services
with various QoS classes, such as
key-priority, normal-priority, & best-effort
priority services;

Link:
a bandwidth transmission medium between nodes that
is engineered as a unit;

Logical Link:
a bandwidth transmission medium of fixed bandwidth
(e.g., T1, DS3, OC3, etc.)
at the link layer (layer 2) between 2 nodes,
established on a path consisting of (possibly several)
physical transport links (at layer 1) which are
switched, for example, through several optical
cross-connect devices;

Node:
a network element (switch, router, exchange)
providing switching and routing
capabilities, or an aggregation of such
network elements representing a network;

Multiservice Network
a network in which various classes of service share
the transmission, switching,
queuing, management, and other resources of the network;

O-D pair: an originating node to destination node pair for a given connection/bandwidth-allocation request;

Originating Node: originating node within a given network;

Packet Switching denotes the transfer of an individual packet over a connection between an input port and an output port within a given packet-switching node through the packet-switching fabric (see Switching);

Path: a concatenation of links providing a connection/bandwidth-allocation between an O-D pair;

Physical Transport Link: a bandwidth transmission medium at the physical layer (layer 1) between 2 nodes, such as on an optical fiber system between terminal equipment used for the transmission of bits or packets (see transport);

Policy-Based Routing network function which involves the application of rules applied to input parameters to derive a routing table and its associated parameters;

QoS (quality of service) a set of service requirements to be met by the network while transporting a Connection or flow; the collective effect of service performance which determine the Degree of satisfaction of a user of the service;

QoS Resource Management network functions which include class-of-service identification, routing table derivation, connection admission, bandwidth allocation, bandwidth protection, Bandwidth reservation, priority routing, and priority queuing;

QoS Routing see QoS Resource Management;

QoS Variable any performance variable (such as congestion, delay, etc.) which is perceivable by a user;

Route: a set of paths connecting the same originating node-destination node pair;

Routing the process of determination, establishment, and use of routing tables to select paths between an input port at the ingress network edge and output port at the egress network edge; includes the process of performing both call routing and connection routing (see call routing and connection routing);

Routing Table: describes the path choices and selection rules to select one path out of the route for a connection/bandwidth-allocation request;

Switching denotes connection of an input port to an output port within a given node through the switching fabric;

Traffic Engineering encompasses traffic management, capacity management, traffic measurement and modeling, network modeling, and performance analysis;

Traffic Engineering Methods network functions which support traffic engineering and include call routing, connection routing, QoS resource management,
routing table management, and capacity management;

Traffic Stream: a class of connection requests with the same traffic characteristics;

Traffic Trunk: an aggregation of traffic flows of the same class which are routed on the same path (see logical link)

Transport refers to the transmission of bits or packets on the physical layer (layer 1) between 2 nodes, such as on an optical fiber system between terminal equipment (note that this definition is distinct from the IP-protocol terminology of transport as end-to-end connectivity at layer 4, such as with the Transport Control Protocol (TCP))

Via node: an intermediate node in a path within a given network;

3.0 Traffic Engineering Model

Figure 1.1 illustrates a model for network traffic engineering. The central box represents the network, which can have various architectures and configurations, and the routing tables used within the network. Network configurations could include metropolitan area networks, national intercity networks, and global international networks, which support both hierarchical and nonhierarchical structures and combinations of the two. Routing tables describe the path choices from an originating node to a terminating node, for a connection request for a particular service. Hierarchical and nonhierarchical traffic routing tables are possible, as are fixed routing tables and dynamic routing tables. Routing tables are used for a multiplicity of traffic and transport services on the telecommunications network.

The functions depicted in Figure 1.1 are consistent with the definition of TE employed by the Traffic Engineering Working Group (TEWG) within the Internet Engineering Task Force (IETF):

Internet Traffic Engineering is concerned with the performance optimization of operational networks. It encompasses the measurement, modeling, characterization, and control of Internet traffic, and the application of techniques to achieve specific performance objectives, including the reliable and expeditious movement of traffic through the network, the efficient utilization of network resources, and the planning of network capacity.

Terminology used in the draft, as illustrated in Figure 1.2, is that a link is a transmission medium (logical or physical) which connects two nodes, a path is a sequence of links connecting an origin and destination node, and a route is the set of different paths between the origin and destination that a call might be routed on within a particular routing discipline. Here a call is a generic term used to describe the establishment, utilization, and release of a connection, or data flow. In this context a call can refer to a voice call established perhaps using the SS7 signaling protocol, or to a web-based data flow session, established perhaps by the HTTP and associated IP-based protocols. Various implementations of routing tables are discussed in ANNEX 2.

Traffic engineering functions include traffic management, capacity management, and network planning. Traffic management ensures that network performance is maximized under all conditions including load shifts and failures. Capacity management ensures that the network is designed and
provisioned to meet performance objectives for network demands at minimum cost. Network planning ensures that node and transport capacity is planned and deployed in advance of forecasted traffic growth. Figure 1.1 illustrates traffic management, capacity management, and network planning as three interacting feedback loops around the network. The input driving the network ("system") is a noisy traffic load ("signal"), consisting of predictable average demand components added to unknown forecast error and load variation components. The load variation components have different time constants ranging from instantaneous variations, hour-to-hour variations, day-to-day variations, and week-to-week or seasonal variations. Accordingly, the time constants of the feedback controls are matched to the load variations, and function to regulate the service provided by the network through capacity and routing adjustments.

Traffic management functions include a) call routing, which entails number/name translation to routing address, b) connection or bearer-path routing methods, c) QoS resource management, d) routing table management, and e) dynamic transport routing. These functions can be a) decentralized and distributed to the network nodes, b) centralized and allocated to a centralized controller such as a bandwidth broker, or c) performed by a hybrid combination of these approaches.

Capacity management plans, schedules, and provisions needed capacity over a time horizon of several months to one year or more. Under exceptional circumstances, capacity can be added on a shorter-term basis, perhaps one to several weeks, to alleviate service problems. Network design embedded in capacity management encompasses both routing design and capacity design. Routing design takes account of the capacity provided by capacity management, and on a weekly or possibly real-time basis adjusts routing tables as necessary to correct service problems. The updated routing tables are provisioned (configured) in the switching systems either directly or via an automated routing update system. Network planning includes node planning and transport planning, operates over a multiyear forecast interval, and drives network capacity expansion over a multiyear period based on network forecasts.

The scope of the TE methods includes the establishment of connections for narrowband, wideband, and broadband multimedia services within multiservice networks and between multiservice networks. Here a multiservice network refers to one in which various classes of service share the transmission, switching, management, and other resources of the network. These classes of services can include constant bit rate (CBR), variable bit rate (VBR), unassigned bit rate (UBR), and available bit rate (ABR) traffic classes. There are quantitative performance requirements that the various classes of service normally are required to meet, such as end-to-end blocking, delay, and/or delay-jitter objectives. These objectives are achieved through a combination of traffic management and capacity management.

Figure 1.3 illustrates the functionality for setting up a connection from an originating node in one network to a destination node in another network, using one or more routing methods across networks of various types. The Figure illustrates a multimedia connection between two PCs which carries traffic for a combination of voice, video, and image applications. For this purpose a logical point-to-point connection is established from the PC served by node al to the PC served by node c2. The connection could be a CBR ISDN connection across TDM-based network A and ATM-based network C, or it might be a VBR connection via IP-based network B. Gateway nodes a3, b1, b4, and c1 provide the interworking capabilities between the TDM-, ATM-, and IP-based networks. The actual multimedia connection might be routed, for
example, on a path consisting of nodes a1-a2-a3-b1-b4-c1-c2, or possibly on a different path through different gateway nodes.

We now briefly describe the traffic model, the traffic management functions, the capacity management functions, and the TE operational requirements, which are further developed in ANNEXES 2-7 of the draft.

4.0 Traffic Models

In this section we discuss load variation models which drive traffic engineering functions, that is traffic management, capacity management, and network planning. Table 1.1 summarizes examples of models that could be used to represent the different traffic variations under consideration. Traffic models for both voice and data traffic need to be reflected.

Work has been done on measurement and characterization of data traffic, such as web-based traffic [FGLRRT00, FGHW99, LTWW94]. Some of the analysis suggests that web-based traffic can be self-similar, or fractal, with very large variability and extremely long tails of the associated traffic distributions. Characterization studies of such data traffic have investigated various traditional models, such as the Markov modulated Poisson Process (MMPP), in which it is shown that MMPP with two parameters can suitably capture the essential nature of the data traffic [H99, BCHLL99].

Modeling work has been done to investigate the causes of the extreme variability of web-based traffic. In [HM00], the congestion-control mechanisms for web-based traffic, such as window flow control for transport-control-protocol (TCP) traffic appear to be at the root cause of its extreme variability over small time scales. [FGHW99] also shows that the variability over small time scales is impacted in a major way by the presence of TCP-like flow control algorithms which give rise to burstiness and clustering of IP packets. However, [FGHW99] also finds that the self-similar behavior over long time scales is almost exclusively due to user-related variability and not dependent on the underlying network-specific aspects.

Regarding the modeling of voice and data traffic in a multiservice model, [HM00] suggests that the regular flow control dynamics are more useful to model than the self-similar traffic itself. Much of the traffic to be modeled is VBR traffic subject to service level agreements (SLAs), which is subject to admission control based on equivalent bandwidth resource requirements and also to traffic shaping in which out-of-contract packets are marked for dropping in the network queues if congestion arises. Other VBR traffic, such as best-effort internet traffic, is not allocated any bandwidth in the admission of session flows, and all of its packets would be subject to dropping ahead of the CBR and VBR-SLA traffic. Hence, we can think of the traffic model consisting of two components:

* the CBR and VBR-SLA traffic that is not marked for dropping constitute less variable traffic subject to more traditional models

* the VBR best-effort traffic and the VBR-SLA traffic packets that are marked and subject to dropping constitute a much more variable, self-similar traffic component.

Considerable work has been done on modeling of broadband and other data traffic, in which two-parameter models that capture the mean and burstiness of the connection and flow arrival processes have proven to be quite
adequate. See [E.716] for a good reference on this. Much work has also been done on measurement and characterization of voice traffic, and two-parameter models reflecting mean and variance (the ratio of the variance to the mean is sometimes called the peakedness parameter) of traffic have proven to be accurate models. We model the large variability in packet arrival processes in an attempt to capture the extreme variability of the traffic.

Here we reflect the two-parameter, multiservice traffic models for connection and flow arrival processes, which are manageable from a modeling and analysis aspect and which attempt to capture essential aspects of data and voice traffic variability for purposes of traffic engineering and QoS methods. In ANNEX 2 we introduce the models of variability in the packet arrival processes.

Table 1.1
Traffic Models for Load Variations of Connection/Flow Arrival Processes

For instantaneous traffic load variations, the load is typically modeled as a stationary random process over a given period (normally within each hourly period) characterized by a fixed mean and variance. From hour to hour, the mean traffic loads are modeled as changing deterministically; for example, according to their 20-day average values. From day to day, for a fixed hour, the mean load can be modeled, for example, as a random variable having a gamma distribution with a mean equal to the 20-day average load. From week to week, the load variation is modeled as a random process in the network design procedure. The random component of the realized week-to-week load is the forecast error, which is equal to the forecast load minus the realized load. Forecast error is accounted for in short-term capacity management.

In traffic management, traffic load variations such as instantaneous variations, hour-to-hour variations, day-to-day traffic variations, and week-to-week variations are responded to in traffic management by appropriately controlling number translation/routing, path selection, routing table management, and/or QoS resource management. Traffic management provides monitoring of network performance through collection and display of traffic and performance data, and allows traffic management controls, such as destination-address per-connection blocking, per-connection gapping, routing table modification, and path selection reroute controls, to be inserted when circumstances warrant. For example, a focused overload might lead to application of connection gapping controls in which a connection request to a particular destination address or set of addresses is admitted only once every x seconds, and connections arriving after an accepted call are rejected for the next x seconds. In that way call gapping throttles the calls and prevents overloading the network to a particular focal point. Routing table modification and reroute control are illustrated in ANNEXES 2, 3, 5, and 7.

Capacity management must provide sufficient capacity to carry the expected traffic variations so as to meet end-to-end blocking/delay objective levels. Here the term blocking refers to the denial or non-admission of a call or connection request, based for example on the lack of available resources on a particular link (e.g., link bandwidth or queuing resources). Traffic load variations lead in direct measure to capacity increments and can be categorized as (1) minute-to-minute instantaneous variations and associated busy-hour traffic load capacity, (2) hour-to-hour variations and associated multihour capacity, (3) day-to-day variations and associated day-to-day capacity, and (4) week-to-week variations and associated reserve capacity.
Design methods within the capacity management procedure account for the mean and variance of the within-the-hour variations of the offered and overflow loads. For example, classical methods [e.g., Wil56] are used to size links for these two parameters of load. Multihour dynamic route design accounts for the hour-to-hour variations of the load and, hour-to-hour capacity can vary from zero to 20 percent or more of network capacity. Hour-to-hour capacity can be reduced by multihour dynamic routing design models such as the discrete event flow optimization, traffic load flow optimization, and virtual trunking flow optimization models described in ANNEX 6. As noted in Table 1.1, capacity management excludes non-recurring traffic such as caused by overloads (focused or general overloads), or failures. This process is described further in ANNEX 7.

It is known that some daily variations are systematic (for example, Monday morning business traffic is usually higher than Friday morning); however, in some day-to-day variation models these systematic changes are ignored and lumped into the stochastic model. For instance, the traffic load between Los Angeles and New Brunswick is very similar from one day to the next, but the exact calling levels differ for any given day. This load variation can be characterized in network design by a stochastic model for the daily variation, which results in additional capacity called day-to-day capacity. Day-to-day capacity is needed to meet the average blocking/delay objective when the load varies according to the stochastic model. Day-to-day capacity is nonzero due to the nonlinearities in link blocking and/or link queuing delay levels as a function of load. When the load on a link fluctuates about a mean value, because of day-to-day variation, the mean blocking/delay is higher than the blocking/delay produced by the mean load. Therefore, additional capacity is provided to maintain the blocking/delay probability grade-of-service objective in the presence of day-to-day load variation.

Typical day-to-day capacity required is 4--7 percent of the network cost for medium to high day-to-day variations, respectively. Reserve capacity, like day-to-day capacity, comes about because load uncertainties---in this case forecast errors---tend to cause capacity buildup in excess of the network design that exactly matches the forecast loads. Reluctance to disconnect and rearrange link and transport capacity contributes to this reserve capacity buildup. At a minimum, the currently measured mean load is used to adjust routing and capacity design, as needed. In addition, the forecast-error variance component in used in some models to build in so-called protective capacity. Reserve or protective capacity can provide a cushion against overloads and failures, and generally benefits network performance. However, provision for reserve capacity is not usually built into the capacity management design process, but arises because of sound administrative procedures. These procedures attempt to minimize total cost, including both network capital costs and operations costs. Studies have shown that reserve capacity in some networks to be in the range of 15 to 25 percent or more of network cost [FHH79]. This is further described in ANNEXES 5 and 6.

5.0 Traffic Management Functions

In ANNEXES 2-5, traffic management functions are discussed:

a) Call Routing Methods (ANNEX 2). Call routing involves the translation of a number or name to a routing address. We describe how number (or name) translation should result in the E.164 ATM end-system addresses (AESA), network routing addresses (NRAs), and/or IP addresses. These addresses are used for routing purposes and therefore must be carried in the connection-setup information element (IE).
b) Connection/Bearer-Path Routing Methods (ANNEX 2). Connection or bearer-path routing involves the selection of a path from the originating node to the destination node in a network. We discuss bearer-path selection methods, which are categorized into the following four types: fixed routing (FR), time-dependent routing (TDR), state-dependent routing (SDR), and event-dependent routing (EDR). These methods are associated with routing tables, which consist of a route and rules to select one path from the route for a given connection or bandwidth-allocation request.

c) QoS Resource Management Methods (ANNEX 3). QoS resource management functions include class-of-service derivation, policy-based routing table derivation, connection admission, bandwidth allocation, bandwidth protection, bandwidth reservation, priority routing, priority queuing, and other related resource management functions.

d) Routing Table Management Methods (ANNEX 4). Routing table management information, such as topology update, status information, or routing recommendations, is used for purposes of applying the routing table design rules for determining path choices in the routing table. This information is exchanged between one node and another node, such as between the ON and DN, for example, or between a node and a network element such as a bandwidth-broker processor (BBP). This information is used to generate the routing table, and then the routing table is used to determine the path choices used in the selection of a path.

e) Dynamic Transport Routing Methods (ANNEX 5). Dynamic transport routing combines with dynamic traffic routing to shift transport bandwidth among node pairs and services through use of flexible transport switching technology, such as optical cross-connects (OXC)s. Dynamic transport routing offers advantages of simplicity of design and robustness to load variations and network failures, and can provide automatic link provisioning, diverse link routing, and rapid link restoration for improved transport capacity utilization and performance under stress. OXC's can reconfigure logical transport capacity on demand, such as for peak day traffic, weekly redesign of link capacity, or emergency restoration of capacity under node or transport failure. MPLS control capabilities are proposed for the setup of layer 2 logical links through OXC's [ARDC99].

6.0 Capacity Management Functions

In ANNEX 6, we discuss capacity management methods, as follows:

a) Link Capacity Design Models. These models find the optimum tradeoff between traffic carried on a shortest network path (perhaps a direct link) versus traffic carried on alternate (longer, less efficient) network paths.

b) Shortest Path Selection Models. These models enable the determination of shortest paths in order to provide a more efficient and flexible routing plan.

c) Multihour Network Design Models. Three models are described including i) discrete event flow optimization (DEFO) models, ii) traffic load flow optimization (TLFO) models, and iii) virtual trunking flow optimization (VTFO) models. DEFO models have the advantage of being able to model traffic and routing methods of arbitrary complexity, for example, such as self-similar traffic.

d) Day-to-day Load Variation Design Models. These models describe techniques for handling day-to-day variations in capacity design.

e) Forecast Uncertainty/Reserve Capacity Design Models. These models describe the means for accounting for errors in projecting design traffic loads in the capacity design of the network.
7.0 Traffic Engineering Operational Requirements

In ANNEX 7, we discuss traffic engineering operational requirements, as follows:

a) Traffic Management. We discuss requirements for real-time performance monitoring, network control, and work center functions. The latter includes automatic controls, manual controls, code controls, cancel controls, reroute controls, peak-day controls, traffic management on peak days, and interfaces to other work centers.

b) Capacity Management - Forecasting. We discuss requirements for load forecasting, including configuration database functions, load aggregation, basing, and projection functions, and load adjustment cycle and view of business adjustment cycle. We also discuss network design, work center functions, and interfaces to other work centers.

c) Capacity Management - Daily and Weekly Performance Monitoring. We discuss requirements for daily congestion analysis, study-week congestion analysis, and study-period congestion analysis.

d) Capacity Management - Short-Term Network Adjustment. We discuss requirements for network design, work center functions, and interfaces to other work centers.

e) Comparison of off-line (TDR) versus on-line (SDR/EDR) TE methods. We contrast off-line TE methods, such as in a TDR-based network, with on-line TE methods, such as in an SDR- or EDR-based network.

8.0 Traffic Engineering Modeling & Analysis

In ANNEXES 2-6 we use network models to illustrate the traffic engineering methods developed in the draft. The details of the models are presented in each ANNEX in accordance with the TE functions being illustrated.

In the draft, a full-scale 135-node national network node model is used together with a multiservice traffic demand model to study various TE scenarios and tradeoffs. Typical voice/ISDN traffic loads are used to model the various network alternatives. These voice/ISDN loads are further segmented in the model into eight constant-bit-rate (CBR) virtual networks (VNETs), including business voice, consumer voice, international voice in and out, key-service voice, normal and key-service 64-kbps ISDN data, and 384-kbps ISDN data. The data services traffic model incorporates typical traffic load patterns and comprises three additional VNET load patterns. These include a) a variable bit rate real-time (VBR-RT) VNET, representing services such as IP-telephony and compressed voice, b) a variable bit rate and credit card check, and c) an unassigned bit rate (UBR) VNET, representing best-effort services such as email, voice mail, and file transfer multimedia applications. The cost model represents typical switching and transport costs, and illustrates the economies-of-scale for costs projected for high capacity network elements in the future.

Many different alternatives and tradeoffs are examined in the models, including:

1. centralized routing table control versus distributed control
2. off-line, pre-planned (e.g., TDR-based) routing table control versus on-line routing table control (e.g., SDR- or EDR-based)
3. per-flow traffic management versus per-virtual-network (or per-traffic-trunk or per-bandwidth-pipe) traffic management
4. sparse logical topology versus meshed logical topology
5. FR versus TDR versus SDR versus EDR path selection
6. multilink path selection versus two-link path selection
7. path selection using local status information versus global status

Ash <draft-ietf-tewg-qos-routing-02.txt> [Page MAIN-16]
8. Global status dissemination alternatives including status flooding, distributed query for status, and centralized status in a bandwidth-broker processor.

Table 1.2 summarizes brief comparisons and observations, based on the modeling, in each of the above alternatives and tradeoffs (further details are contained in ANNEXES 2-6).

Table 1.2
Tradeoff Categories and Comparisons
(Based on Modeling in ANNEXES 2-6)

9.0 Conclusions/Recommendations

Following is a summary of the main conclusions/recommendations reached in the draft.

9.1 Conclusions/Recommendations on Call Routing & Connections Routing Methods (ANNEX 2)

* TE methods are recommended to be applied, and in all cases of the TE methods being applied, network performance is always better and usually substantially better than when no TE methods are applied.

* Sparse-topology multilink-routing networks are recommended and provide better overall performance under overload than meshed-topology networks, but performance under failure may favor the 2-link STT-EDR/DC-SDR meshed-topology options with more alternate routing choices.

* Single-area flat topologies are recommended and exhibit better network performance and, as discussed and modeled in ANNEX 6, greater design efficiencies in comparison with multi-area hierarchical topologies. As illustrated in ANNEX 4, larger administrative areas can be achieved through use of EDR-based TE methods as compared to SDR-based TE methods.

* Event-dependent-routing (EDR) TE path selection methods are recommended and exhibit comparable or better network performance compared to state-dependent-routing (SDR) methods.

  a. EDR TE methods are shown to an important class of TE algorithms. EDR TE methods are distinct from the TDR and SDR TE methods in how the paths (e.g., MPLS label switched paths, or LSPs) are selected. In the SDR TE case, the available link bandwidth (based on LSA flooding of ALB information) is typically used to compute the path. In the EDR TE case, the ALB information is not needed to compute the path, therefore the ALB flooding does not need to take place (reducing the overhead).

  b. EDR TE algorithms are adaptive and distributed in nature and typically use learning models to find good paths for TE in a network. For example, in a success-to-the-top (STT) EDR TE method, if the LSR-A to LSR-B bandwidth needs to be modified, say increased by delta-BW, the primary LSP-p is tried first. If delta-BW is not available on one or more links of LSP-p, then the currently successful LSP-s is tried next. If delta-BW is not available on one or more links of LSP-s, then a new LSP is searched by trying additional candidate paths until a new successful LSP-n is found or the candidate paths are exhausted. LSP-n is then marked as the currently successful path for the next time bandwidth needs to be modified. The performance of distributed EDR TE methods is shown to be equal to or better
than SDR methods, centralized or distributed.

c. While SDR TE models typically use available-link-bandwidth (ALB) flooding for TE path selection, EDR TE methods do not require ALB flooding. Rather, EDR TE methods typically search out capacity by learning models, as in the STT method above. ALB flooding can be very resource intensive, since it requires link bandwidth to carry LSAs, processor capacity to process LSAs, and the overhead can limit area/autonomous system (AS) size. Modeling results show EDR TE methods can lead to a large reduction in ALB flooding overhead without loss of network throughput performance [as shown in ANNEX 4].

d. State information as used by the SDR options (such as with link-state flooding) provides essentially equivalent performance to the EDR options, which typically used distributed routing with crankback and no flooding.

e. Various path selection methods can interwork with each other in the same network, as required for multi-vendor network operation.

* Interdomain routing methods are recommended which extend the intradomain call routing and connection routing concepts, such as flexible path selection and per-class-of-service bandwidth selection, to routing between network domains.

9.2 Conclusions/Recommendations on QoS Resource Management Methods (ANNEX 3)

* QoS resource management is recommended and is shown to be effective in achieving connection-level and packet-level GoS objectives, as well as key service, normal service, and best effort service differentiation.

* Admission control is recommended and is the basis that allows for applying most of the other controls described in this draft.

* Per-VNET bandwidth allocation is recommended and is essentially equivalent to per-flow bandwidth allocation in network performance and efficiency. Because of the much lower routing table management overhead requirements, as discussed and modeled in ANNEX 4, per-VNET bandwidth allocation is preferred to per-flow allocation.

* Both MPLS QoS and bandwidth management and DiffServ priority queuing management are recommended and are important for ensuring that multiservice network performance objectives are met under a range of network conditions. Both mechanisms operate together to ensure QoS resource allocation mechanisms (bandwidth allocation, protection, and priority queuing) are achieved.

9.3 Conclusions/Recommendations on Routing Table Management Methods & Requirements (ANNEX 4)

* Per-VNET bandwidth allocation is recommended and is preferred to per-flow allocation because of the much lower routing table management overhead requirements. Per-VNET bandwidth allocation is essentially equivalent to per-flow bandwidth allocation in network performance and efficiency, as discussed in ANNEX 3.
EDR TE methods are recommended and can lead to a large reduction in ALB flooding overhead without loss of network throughput performance. While SDR TE methods typically use ALB flooding for TE path selection, EDR TE methods do not require ALB flooding. Rather, EDR TE methods typically search out capacity by learning models, as in the STT method. ALB flooding can be very resource intensive, since it requires link bandwidth to carry LSAs, processor capacity to process LSAs, and the overhead can limit area/autonomous system (AS) size.

EDR TE methods are recommended and lead to possible larger administrative areas as compared to SDR-based TE methods because of lower routing table management overhead requirements. This can help achieve single-area flat topologies which, as discussed in ANNEX 3, exhibit better network performance and, as discussed in ANNEX 6, greater design efficiencies in comparison with multi-area hierarchical topologies.

9.4 Conclusions/Recommendations on Transport Routing Methods (ANNEX 5)

Dynamic transport routing is recommended and provides greater network throughput and, consequently, enhanced revenue, and at the same time capital savings should result, as discussed in ANNEX 6.

a. Dynamic transport routing network design enhances network performance under failure, which arises from automatic inter-backbone-router and access logical-link diversity in combination with the dynamic traffic routing and transport restoration of logical links.

b. Dynamic transport routing network design is recommended and improves network performance in comparison with fixed transport routing for all network conditions simulated, which include abnormal and unpredictable traffic load patterns.

Traffic and transport restoration level design is recommended and allows for link diversity to ensure a minimum level of performance under failure.

Robust routing techniques are recommended, which include dynamic traffic routing, multiple ingress/egress routing, and logical link diversity routing; these methods improve response to node or transport failures.

9.5 Conclusions/Recommendations on Capacity Management Methods (ANNEX 6)

Discrete event flow optimization (DEFO) design models are recommended and are shown to be able to capture very complex routing behavior through the equivalent of a simulation model provided in software in the routing design module. By this means, very complex routing networks have been designed by the model, which include all of the routing methods discussed in ANNEX 2 (FR, TDR, SDR, and EDR methods) and the multiservice QoS resource allocation models discussed in ANNEX 3.

Sparse topology options are recommended, such as the multilink STT-EDR/DC-SDR/DP-SDR options, which lead to capital cost advantages, and more importantly to operation simplicity and cost reduction. Capital cost savings are subject to the particular switching and transport cost assumptions. Operational issues are further detailed in ANNEX 7.

Voice and data integration is recommended and...
Multilink routing methods are recommended and exhibit greater design efficiencies in comparison with 2-link routing methods. As discussed and modeled in ANNEX 3, multilink topologies exhibit better network performance under overloads in comparison with 2-link routing topologies; however the 2-link topologies do better under failure scenarios.

Single-area flat topologies are recommended and exhibit greater design efficiencies in termination and transport capacity, but higher cost, and, as discussed and modeled in ANNEX 3, better network performance in comparison with multi-area hierarchical topologies. As illustrated in ANNEX 4, larger administrative areas can be achieved through use of EDR-based TE methods as compared to SDR-based TE methods.

EDR methods are recommended and exhibit comparable design efficiencies to SDR. This suggests that there is not a significant advantage for employing link-state information in these network designs, especially given the high overhead in flooding link-state information in SDR methods.

Dynamic transport routing is recommended and achieves capital savings by concentrating capacity on fewer, high-capacity physical fiber links and, as discussed in ANNEX 5, achieves higher network throughput and enhanced revenue by their ability to flexibly allocate bandwidth on the logical links serving the access and inter-node traffic.

9.6 Conclusions/Recommendations on TE Operational Requirements (ANNEX 7)

Monitoring of traffic and performance data is recommended and is required for traffic management, capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment.

Traffic management is recommended and is required to provide monitoring of network performance through collection and display of real-time traffic and performance data and allow traffic management controls such as code blocks, connection request gapping, and reroute controls to be inserted when circumstances warrant.

Capacity management is recommended and is required for capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment.

Forecasting is recommended and is required to operate over a multiyear forecast interval and drive network capacity expansion.

Daily and weekly performance monitoring is recommended and is required to identify any service problems in the network. If service problems are detected, short-term network adjustment can include routing table updates and, if necessary, short-term capacity additions to alleviate service problems. Updated routing tables are sent to the switching systems.
either directly or via an automated routing update system.

* Short-term capacity additions are recommended and are required as needed, but only as an exception, whereas most capacity changes are normally forecasted, planned, scheduled, and managed over a period of months or a year or more.

* Network design, which includes routing design and capacity design, is recommended and is required within the capacity management function.

* Network planning is recommended and is required for longer-term node planning and transport network planning, and operates over a horizon of months to years to plan and implement new node and transport capacity.

10. Recommended TE/QoS Methods for Multiservice Networks

In summary, TE methods are recommended in this Section for consideration in network evolution. These recommendations are based on

* results of analysis models presented in ANNEXES 2-6, which illustrate the tradeoffs between various TE approaches,

* results of operational comparison studies presented in ANNEXES 2-6,

* established best current practices and experience.

10.1 Recommended Application-Layer IP-Network-Based Service-Creation Capabilities

As discussed in ANNEX 4, these capabilities are recommended for application-layer service-creation capabilities:

* Parlay API (application programming interface)
* call processing language (CPL) & common gateway interface (CGI)
* SIP/IN (intelligent network) interworking

10.2 Recommended Call/IP-Flow Control Layer Capabilities

As discussed in ANNEXES 2 and 4, these capabilities are recommended for name translation, call signaling, and split gateway control:

* ENUM/DNS-based name to IP-address translation
* SIP-based distributed call signaling (DCS)
* MGCP/MEGACO for split gateway control

10.3 Recommended Connection/Bearer Control Layer Capabilities

In this Section we summarize the findings in ANNEXES 2, 3, and 4 which give rise to a recommendation for a TE/QoS admission control method for connection/flow admission, which incorporates dynamic QoS routing connection/bearer layer control.

The analysis considered in ANNEXES 2, 3, and 4 investigates bandwidth allocation for the aggregated case ("per traffic-trunk" or per-VNET (virtual network)) versus the per-flow bandwidth allocation. The following
recommendations are made on QoS resource management, topology, and connection layer control:

* virtual-network traffic allocation for multiservice network
* MPLS-based virtual-network based QoS resource management & dynamic bandwidth reservation methods
* DiffServ-based priority queuing
* per-virtual-network (per-traffic-trunk) bandwidth allocation for lower routing table management overhead
* sparse-topology multilink routing for better performance & design efficiency
* single-area flat topology (as much as possible, while retaining edge-core architecture) for better performance & design efficiency
* MPLS and DiffServ functionality to meet TE/QoS requirements
* success-to-the-top (STT) event-dependent-routing (EDR) TE path selection methods for better performance & lower overhead

These TE admission control and dynamic QoS routing methods will ensure stable/efficient performance of TE methods and help manage resources for and differentiate key service, normal service, & best effort service, and are now briefly summarized. Figure 1.4 illustrates the recommended QoS resource management methods. As illustrated in the Figure, in the multi-service, QoS resource management network, bandwidth is allocated to the individual VNETs (high-priority key services VNETs, normal-priority services VNETs, and best-effort low-priority services VNETs). The Figure also illustrates the use of virtual-network traffic allocation for multiservice networks and the means to differentiate key service, normal service, & best effort service. High-priority and normal-priority traffic connections/flows are subject to admission control based on equivalent bandwidth allocation techniques. However, best-effort services are allocated no bandwidth, and all best-effort traffic is subject to dropping in the queuing/scheduling discipline under congestion conditions.

This allocated bandwidth is protected by bandwidth reservation methods, as needed, but otherwise shared. Each ON monitors VNET bandwidth use on each VNET CRLSP, and determines when VNET CRLSP bandwidth needs to be increased or decreased. Bandwidth changes in VNET bandwidth capacity are determined by ONs based on an overall aggregated bandwidth demand for VNET capacity (not on a per-connection demand basis). Based on the aggregated bandwidth demand, these ONs make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth on the CRLSPs constituting the VNET bandwidth capacity. For example, if connection requests are made for VNET CRLSP bandwidth that exceeds the current CRLSP bandwidth allocation, the ON initiates a bandwidth modification request on the appropriate CRLSP(s). For example, this bandwidth modification request may entail increasing the current CRLSP bandwidth allocation by a discrete increment of bandwidth denoted here as delta-bandwidth (DBW). DBW is a large enough bandwidth change so that modification requests are made relatively infrequently. Also, the ON periodically monitors CRLSP bandwidth use, such as once each minute, and if bandwidth use falls below the current CRLSP allocation the ON initiates a bandwidth modification request to
decrease the CRLSP bandwidth allocation by a unit of bandwidth such as DBW.

Therefore the recommendation is to do "per-VNET", or per traffic trunk, bandwidth allocation, and not call by call, or "per flow" allocation, as, as discussed in Sections 3.4 and 3.5. This kind of per-VNET bandwidth allocation also applies in the case of multi-area TE, as discussed in Sections 2.8 and 3.8. Therefore some telephony concepts, such as call-by-call set up, are not needed in VoIP/TE. That is, there are often good reasons not to make things look like the PSTN. On the other hand, some principles do still apply to VoIP/TE, but are not used as yet, and should be.

The main point about bandwidth reservation is related to both admission control and queue management. That is, if a flow is to be admitted on a longer path, that is, not the primary path (which is preferred and tried first, but let us assume did not have the available bandwidth on one or more links/queues), then there needs to be a minimum level of available bandwidth, call in RESBW (reserved bandwidth), available on each link and in each queue in addition to the requested bandwidth (REQBW). That is, one needs to have RESBW + BEWBW available on each link and queue before admitting the flow on the longer path. On the primary path RESBW is not required. The simulation results given in ANNEX 3 are for an MPLS network, and the results show the effect of using bandwidth reservation, and what happens if you do not use bandwidth reservation (see Tables 3.4 and 3.5). Bandwidth allocation and management is done according to the traffic priority (i.e., key, normal, and best effort), as described in ANNEX 3, and is an additional use of bandwidth reservation methods beyond the use in path selection, as in the example above. Bandwidth allocation in the queues is done according to traffic priority, as discussed in Section 3.6. These principles put forth in the draft do not depend on whether the underlying technology is IP/MPLS-based, ATM/PNNI-based, or TDM/E.351-based, they apply to all technologies, as is demonstrated by the models.

In the models the per-VNET method compares favorably with the per-flow method, which is all feasible within the current MPLS protocol specification and is therefore recommended for the TE admission control and dynamic QoS routing methods.

Furthermore, we find that a distributed event-dependent-routing (EDR)/STT method of LSP management works just as well or better than the state-dependent-routing (SDR) with flooding. An example of the EDR/STT method:

Figure 1.5 illustrates the recommended STT EDR path selection method and the use of a sparse, single-area topology.

The EDR/STT method is fully distributed, reduces flooding, and a larger perhaps even a single backbone area could be used as a result. Edge-router (ER) to backbone-router (BR) hierarchy is also modeled. We modeled an MPLS/Diffserv ER-BR resource management, although it is sometimes claimed that Diffserv alone would suffice on the ER-BR links. The problem there is what happens when bandwidth is exhausted for the connection-oriented voice, ISDN, IP-telephony, etc. services versus the best-effort services. One needs a TE admission control mechanism to reject connection requests when need be. In the ER/BR hierarchy modeled, there is a mesh of LSPs in the backbone, but separate LSPs ("big pipes") for each ER to the backbone BRs, that is, for each ER-BR area (i.e., there is no ER-ER LSP mesh in this case).
Recommended Transport Routing Capabilities

As discussed in ANNEX 5, the following recommendations are made for transport routing:

* dynamic transport routing for better performance & design efficiency
* traffic and transport restoration level design, which allows for link diversity to ensure a minimum level of performance under failure

Recommended Network Operations Capabilities

As discussed in ANNEXES 5 and 6, the following recommendations are made for network operations and design:

* monitor traffic & performance data for traffic management & capacity management

Figure 1.1 illustrates the monitoring of network traffic and performance data to support traffic management and capacity management functions.

* traffic management methods to provide monitoring of network performance and implement traffic management controls such as code blocks, connection request gapping, and reroute controls

Figure 1.7 illustrates the recommended traffic management functions.

* capacity management methods to include capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment

Figure 1.8 illustrates the recommended capacity management functions.

* discrete event flow optimization (DEFO) design models to capture complex routing behavior and design multiservice TE networks

Figure 1.9 illustrates the recommended DEFO design models. The greatest advantage of the DEFO model is its ability to capture very complex routing behavior through the equivalent of a simulation model provided in software in the routing design module. By this means, very complex routing networks have been designed by the model, which include all of the routing methods discussed in ANNEX 2, TDR, SDR, and EDR methods, and the multiservice QoS resource allocation models discussed in ANNEX 3. Complex traffic processes, such as self-similar traffic, can also be modeled with DEFO methods.

Benefits of Recommended TE/QoS Methods for Multiservice Integrated Networks

The benefits of recommended TE/QoS Methods for IP-based multiservice integrated network are as follows:

* IP-network-based service creation (Parlay API, CPL/CGI, SIP-IN)
* lower operations & capital cost
* improved performance
* simplified network management

The IP-network-based service creation capabilities are discussed in ANNEX 4.
the operations and capital cost impacts in ANNEXES 2 and 6, and improved performance impacts in ANNEXES 2 and 3.

Simplified network management comes about because of the following impacts of the recommended TE admission control and dynamic QoS routing methods:

* distributed control, as discussed in ANNEX 2
* eliminate available-link-bandwidth flooding, as discussed in ANNEX 4
* larger/fewer areas, as discussed in ANNEX 4
* automatic provisioning of topology database, as discussed in ANNEX 3
* fewer links/sparse network to provision, as discussed in ANNEX 2

11. Acknowledgements

The author is indebted to many people for much help and encouragement in the course of developing this work. In the IETF I’d like to especially thank Dan Awduche of Movaz, Jim Boyle of Level 3 Communications, Angela Chiu of Celion Networks, and Tom Scott of Vedatel for all their helpful comments, assistance, and encouragement. Within AT&T I’d like to especially thank Chuck Dvorak, Bur Goode, Wai Sum Lai, and Jennifer Rexford for the excellent support and on-going discussions. In the ITU I’d like to particularly thank Anne Elvidge of BT, Tommy Petersen of Ericsson, Bruce Pettitt of Nortel, and Jim Roberts of France Telecom for the valuable help throughout the course of this work. I’d also like to thank Professor Lorne Mason of INRS/University of Quebec for his many insights, discussions, and comments in the course of this work.

12. Authors’ Addresses

Gerald R. Ash
AT&T Labs
Room MT D5-2A01
200 Laurel Avenue
Middletown, NJ 07748
Phone: 732-420-4578
Fax: 732-368-8659
Email: gash@att.com

13. Full Copyright Statement

Copyright (C) The Internet Society (1998). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works.

However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.
The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Figures:

Figure 1.1 Traffic Engineering Model
Figure 1.2 Terminology
Figure 1.3 Example of Multimedia Connection Across TDM-, ATM-, and IP-Based Networks
Figure 1.4 Use MPLS/Diffserv/Virtual-Network-Based QoS Resource Management with Dynamic Bandwidth Reservation Methods
Figure 1.5 Use Success-to-the-Top (STT) Event-Dependent-Routing (EDR) TE Path Selection Methods in a Sparse, Single-Area Topology
Figure 1.6 Use Virtual-Network Traffic Allocation for Multiservice Network Differentiate Key Service, Normal Service, & Best Effort Service
Figure 1.7 Employ Traffic Management Methods to Provide Monitoring of Network Performance and Implement Traffic Management Controls (such as code blocks, connection request gapping, and reroute controls)

Ash <draft-ietf-tewg-qos-routing-02.txt> [Page MAIN-26]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001

ANNEX 1. Bibliography


[A01] Ashwood-Smith, P., et. al., Generalized Multi-Protocol Label Switching (GMPLS) Architecture, work in progress


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-1]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


Internet Draft  TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


ITU-T Recommendation, Number of circuits to be provided in automatic and/or semiautomatic operation, November 1988.

ITU-T Recommendation, Calculation of the number of circuits in a group carrying overflow traffic, November 1988.

ITU-T Recommendation, Number of circuits in a high-usage group, November 1988.


ITU-T Recommendation, Dimensioning a circuit group with multi-slot bearer services and no overflow inputs, March 1993.

ITU-T Recommendation, Dimensioning a circuit group with multi-slot bearer services and overflow traffic, March 2000.


ETSI Secretariat, Telecommunications and Internet Protocol Harmonization over Networks (TIPHON); Naming and Addressing; Scenario 2, DTS/TIPHON-04002 v1.1.64, 1998


Requirements for support of Diff-Serv-aware MPLS Traffic Engineering, work in progress.


[G.723.1] ITU-T Recommendation, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, March 1996.


[H.225.0] ITU-T Recommendation, Media Stream Packetization and Synchronization on Non-Guaranteed Quality of Service LANs, November 1996.


[HSMOA00] Huang, C., Sharma, V., Makam, S., Owens, K., A Path Protection/Restoration Mechanism for MPLS Networks, work in progress.


[HSM000] Huang, C., Sharma, V., Makam, S., Owens, K., A Path Protection/Restoration Mechanism for MPLS Networks, work in progress.

[HSMOA00] Huang, C., Sharma, V., Makam, S., Owens, K., Akyol, B., Extensions to RSVP-TE for MPLS Path Protection, work in progress.


[IYBKQ00] Isoyama, K., Yoshida, M., Brunner, M., Kind, A., Quittek, J., Policy Framework QoS Information Model for MPLS, work in progress.

[J00] Jamoussi, B., Editor, Constraint-Based LSP Setup using LDP, work in progress.


[K99] Kilkki, K., Differentiated Services for the Internet, Macmillan,
1999.


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-8]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


[LG00] Lee, C., Ganti, S., Path Request and Path Reply Message, work in progress.


[M98a] Ma, Q., Quality-of-Service Routing in Integrated Services Networks,


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-9]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


[NWRH99] Neilson, R., Wheeler, J., Reichmeyer, F., Hares, S., A Discussion
of Bandwidth Broker Requirements for Internet2 Qbone Deployment, August 1999.


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-10]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


[R99] Roberts, J. W., Engineering for Quality of Service, Chapter appearing in [PW00].


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-11]

Internet Draft TE & QoS Methods for IP,ATM,TDM-Based Networks August 2001


[RO00] Roberts, J. W., Oueslati-Boulahia, S., Quality of Service by Flow Aware Networking, work in progress.


[SAHG00] Slutsman, L., Ash, G., Haerens, F., Gurbani, V. K., Framework and Requirements for the Internet Intelligent Network (IIN), work in progress.


[ST98] Sikora, J., Teitelbaum, B., Differentiated Services for Internet2, Internet2: Joint Applications/Engineering QoS Workshop, Santa Clara, CA, May
1998.


[SX01] Strand, J., Xue, Y., Routing for Optical Networks With Multiple Routing Domains, oif2001.046 (for a copy send an email request to jls@research.att.com).


Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX1-12]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001


[V99] Villamizar, C., MPLS Optimized Multipath, work in progress.


[WHJ00] Wright, S., Herzog, S., Jaeger, R., Requirements for Policy Enabled MPLS, work in progress.


2.1 Introduction

In the draft we assume the separation of "call routing" and signaling for call establishment from "connection (or bearer-path) routing" and signaling for bearer-channel establishment. Call routing protocols primarily translate a number or a name, which is given to the network as part of a call setup, to a routing address needed for the connection (bearer-path) establishment. Call routing protocols are described for example in [Q.2761] for the Broadband ISDN Used Part (B-ISUP) call signaling, [ATM990048] for bearer-independent call control (BICC), or virtual trunking call signaling, [H.323] for H.323 call signaling, [GR99] for the media gateway control [RFC2805] call signaling, and in [HSSR99] for the session initiation protocol (SIP) call signaling. Connection routing protocols include for example [Q.2761] for B-ISUP signaling, [ATM960055] for PNNI signaling, [ATM960061] for UNI signaling, [DN99] for switched virtual path (SVP) signaling, and [J00] for MPLS constraint-based routing label distribution protocol (CRLDP) signaling.

A specific connection or bearer-path routing method is characterized by the routing table used in the method. The routing table consists of a set of paths and rules to select one path from the route for a given connection.
request. When a connection request arrives at its originating node (ON),
the ON implementing the routing method executes the path selection rules
associated with the routing table for the connection to determine a selected
path from among the path candidates in the route for the connection request.
In a particular routing method, the path selected for the connection request
is governed by the connection routing, or path selection, rules. Various
path selection methods are discussed: fixed routing (FR) path selection,
time-dependent routing (TDR) path selection, state-dependent routing (SDR)
path selection, and event-dependent routing (EDR) path selection.

2.2 Call Routing Methods

Call routing entails number (or name) translation to a routing address,
which is then used for connection establishment. Routing addresses can
consist, for example, of a) E.164 ATM end system addresses (AESAs) [E.191],
b) network routing addresses (NRAs) [E.353], and/or c) IP addresses [S94].
As discussed in ANNEX 4, a TE requirement is the need for carrying
E.164-AESA addresses, NRAs, and IP addresses in the connection-setup
information element (IE). In that case, E.164-AESA addresses, NRAs, and IP
addresses become the standard addressing method for interworking across IP-,
ATM-, and TDM-based networks. Another TE requirement is that a call
identification code (CIC) be carried in the call-control and bearer-control
connection-setup IEs in order to correlate the call-control setup with the
bearer-control setup [Q.1901, ATM990048]. Carrying these additional
parameters in the Signaling System 7 (SS7) ISDN User Part (ISUP)
connection-setup IEs is referred to as the bearer independent call control
(BICC) protocol.

Number (or name) translation, then, should result in the E.164-AESA
addresses, NRAs, and/or IP addresses. NRA formats are covered in [E.353],
and IP-address formats in [S94]. The AESA address has a 20-byte format as
shown in Figure 2.1a below [E.191].

The IDP is the initial domain part and the DSP is the domain specific part.
The IDP is further subdivided into the AFI and IDI. The IDI is the initial
domain identifier and can contain the 15-digit E.164 address if the AFI is
set to 45. AFI is the authority and format identifier and determines what
kind of addressing method is followed, and based on the 1 octet AFI value,
the length of the IDI and DSP fields can change. The E.164-AESA address is
used to determine the path to the destination endpoint. E.164-AESA
addressing for B-ISDN services is supported in ATM networks using PNNI,
through use of the above AESA format. In this case the E.164 part of the
AESA address occupies the 8 octet IDI, and the 11 octet DSP can be used at
the discretion of the network operator (perhaps for sub-addresses). The
above AESA structure also supports AESA DCC (data country code) and AESA ICD
(international code designator) addressing formats.

Within the IP network, routing is performed using IP addresses. Translation
databases, such as based on domain name system (DNS) technology [RFC2916],
are used to translate the E.164 numbers/names for calls to IP addresses for
routing over the IP network. The IP address is a 4-byte address structure
as shown below:

There are five classes of IP addresses. Different classes have different
field lengths for the network identification field. Classless inter-domain
routing (CIDR) allows blocks of addresses to be given to service providers
in such a manner as to provide efficient address aggregation. This is
accompanied by capabilities in the BGP4.0 protocol for efficient address
advertisements [RL00, S94].
2.3 Connection (Bearer-Path) Routing Methods

Connection routing is characterized by the routing table used in the method and rules to select one path from the route for a given connection or bandwidth-allocation request. When a connection/bandwidth-allocation request is initiated by an ON, the ON implementing the routing method executes the path selection rules associated with the routing table for the connection/bandwidth-allocation to find an admissible path from among the paths in the route that satisfies the connection/bandwidth-allocation request. In a particular routing method, the selected path is determined according to the rules associated with the routing table. In a network with originating connection/bandwidth-allocation control, the ON maintains control of the connection/bandwidth-allocation request. If crankback/bandwidth-not-available is used, for example, at a via node (VN), the preceding node maintains control of the connection/bandwidth-allocation request even if the request is blocked on all the links outgoing from the VN.

Here we are discussing network-layer connection routing (sometimes referred to as "layer-3" routing), as opposed to the link-layer logical-transport-link ("layer-2") routing or physical-layer ("layer-1") routing. In the draft the term "link" will normally mean "logical-link." In ANNEX 5 we address logical-link routing.

The network-layer (layer-3) connection routing methods addressed include those discussed in

* Open Shortest Path First (OSPF), Border Gateway Protocol (BGP), and Multiprotocol Label Switching (MPLS) for IP-based routing methods,
* User-to-Network Interface (UNI), Private Network-to-Network Interface (PNNI), ATM Inter-Network Interface (AINI), and Bandwidth Modify for ATM-based routing methods, and

In an IP network, logical links called traffic trunks can be defined which consist of MPLS label switched paths (LSPs) between the IP nodes. Traffic trunks are used to allocate the bandwidth of the logical links to various node pairs. In an ATM network, logical links called virtual paths (VPs) (the equivalent of traffic trunks) can be defined between the ATM nodes, and VPs can be used to allocate the bandwidth of the logical links to various node pairs. In a TDM network, the logical links consist of trunk groups between the TDM nodes.

A sparse logical link network is typically used with IP and ATM technology, as illustrated in Figure 2.2, and FR, TDR, SDR, and EDR can be used in combination with multilink shortest path selection.

A meshed logical-link network is typically used with TDM technology, but can be used also with IP or ATM technology as well, and selected paths are normally limited to 1 or 2 logical links, or trunk groups, as illustrated in Figure 2.3.

Paths may be set up on individual connections (or "per flow") for each call request, such as on a switched virtual circuits (SVC). Paths may also be set up for bandwidth-allocation requests associated with "bandwidth pipes" or traffic trunks, such as on switched virtual paths (SVPs) in ATM-based networks or constraint-based routing label switched paths (CRLSPs) in
IP-based networks. Paths are determined by (normally proprietary) algorithms based on the network topology and reachable address information. These paths can cross multiple peer groups in ATM-based networks, and multiple autonomous systems (ASs) in IP-based networks. An ON may select a path from the routing table based on the routing rules and the QoS resource management criteria, described in ANNEX 3, which must be satisfied on each logical-link in the path. If a link is not allowed based on the QoS criteria, then a release with crankback/bandwidth-not-available parameter is used to signal that condition to the ON in order to return the connection/bandwidth-allocation request to the ON, which may then select an alternate path. In addition to controlling bandwidth allocation, the QoS resource management procedures can check end-to-end transfer delay, delay variation, and transmission quality considerations such as loss, echo, and noise.

When source routing is used, setup of a connection/bandwidth-allocation request is achieved by having the ON identify the entire selected path including all VNs and DN in the path in a designated-transit-list (DTL) or explicit-route (ER) parameter in the connection-setup IE. If the QoS or traffic parameters cannot be realized at any of the VNs in the connection setup request, then the VN generates a crankback (CBK)/bandwidth-not-available (BNA) parameter in the connection-release IE which allows a VN to return control of the connection request to the ON for further alternate routing. In ANNEX 4, the DTL/ER and CBK/BNA elements are identified as being required for interworking across IP-, ATM-, and TDM--based networks.

As noted earlier, connection routing, or path selection, methods are categorized into the following four types: fixed routing (FR), time-dependent routing (TDR), state-dependent routing (SDR), and event-dependent routing (EDR). We discuss each of these methods in the following paragraphs. Examples of each of these path selection methods are illustrated in Figures 2.4a and 2.4b and discussed in the following sections.

Dynamic routing allows routing tables to be changed dynamically, either in an off-line, preplanned, time-varying manner, as in TDR, or on-line, in real time, as in SDR or EDR. With off-line, pre-planned TDR path selection methods, routing patterns contained in routing tables might change every hour or at least several times a day to respond to measured hourly shifts in traffic loads, and in general TDR routing tables change with a time constant normally greater than a call/traffic-flow holding time. A typical TDR routing method may change routing tables every hour, which is longer than a typical voice call/traffic-flow holding time of a few minutes. Three implementations of TDR dynamic path selection are illustrated in Figure 2.4a, which shows multilink path routing, 2-link path routing, and progressive routing.

TDR routing tables are preplanned, preconfigured, and recalculated perhaps each week within the capacity management network design function. Real-time dynamic path selection does not depend on precalculated routing tables. Rather, the node or centralized bandwidth broker senses the immediate traffic load and if necessary searches out new paths through the network possibly on a per-traffic-flow basis. With real-time path selection methods, routing tables change with a time constant on the order of or less than a call/traffic-flow holding time. As illustrated in Figure 2.4b, on-line, real-time path selection methods include EDR and SDR.

2.4 Hierarchical Fixed Routing (FR) Path Selection
Hierarchical fixed routing (FR) is an important routing topology employed in all types of networks, including IP-, ATM-, and TDM-based networks. In IP-based networks, there is often a hierarchical relationship among different "areas", or sub-networks. Hierarchical multi-domain (or multi-area or multi-autonomous-system) topologies are normally used with IP routing protocols (OSPF, BGP) and ATM routing protocols (PNNI), as well as within almost all TDM-based network routing topologies.

For example, in Figure 2.4c, BB1 and BB2 could be backbone nodes in a "backbone area", and AN1 and AN2 could be access nodes in separate "access areas" distinct from the backbone area. Routing between the areas follows a hierarchical routing pattern, while routing within an area follows an interior gateway protocol (IGP), such as OSPF plus MPLS. Similarly, in ATM-based networks the same concept exists, but here the "areas" are called "peer-groups", and for example, the IGP used within peer-groups could be PNNI. In TDM-based networks, the routing between sub-networks, for example, metropolitan-area-networks and long-distance networks, is normally hierarchical, as in IP- and ATM-based networks, and the IGP in TDM-based networks could be either hierarchical or dynamic routing. We now discuss more specific attributes and methods for hierarchical FR path selection.

In a FR method, a routing pattern is fixed for a connection request. A typical example of fixed routing is a conventional, TDM-based, hierarchical alternate routing pattern where the route and route selection sequence are determined on a preplanned basis and maintained over a long period of time. Hierarchical FR is illustrated below. FR is more efficiently applied, however, when the network is nonhierarchical, or flat, as compared to the hierarchical structure [A98].

The aim of hierarchical fixed routing is to carry as much traffic as is economically feasible over direct links between pairs of nodes low in the hierarchy. This is accomplished by application of routing procedures to determine where sufficient load exists to justify high-usage logical-links, and then by application of alternate-routing principles that effectively pool the capacities of high-usage links with those of final links, to the end that all traffic is carried efficiently.

The routing of connection requests in a hierarchical network involves an originating ladder, a terminating ladder, and links interconnecting the two ladders. In a two-level network, for example, the originating ladder is the final link from lower level-1 node to the upper level-2 node, and the terminating ladder is the final link from upper level-2 node to the lower level-1 node. Links AN1-BB2, AN2-BB1, and BB1--BB2 in Figure 2.4c are examples of interladder links.

The identification of the proper interladder link for the routing of a given connection request identifies the originating ladder "exit" point and the terminating ladder "entry" point. Once these exit and entry points are identified and the intraladder links are known, a first-choice path from originating to terminating location can be determined.

Various levels of traffic concentration are used to achieve an appropriate balance between transport and switching. The generally preferred routing sequence for the AN1 to AN2 connections is

1. A connection request involving no via nodes: path AN1-AN2 (if the link existed).
2. A connection request involving one via node: path AN1-BB2-AN2, AN1-BB1-AN2, in that order.

3. A connection request involving two via nodes: path AN1-BB1-BB2-AN2

This procedure provides only the first-choice interladder link from AN1 to AN2. Connection requests from AN2 to AN1 often route differently. To determine the AN2-to-AN1 route requires reversing the diagram, making AN2-BB2 the originating ladder and AN1-BB1 the terminating ladder. In Figure 2.4c the preferred path from AN2 to AN1 is AN2-AN1, AN2-BB1-AN1, AN2-BB2-AN1, and AN2-BB2-BB1-AN1, in that order. The alternate path for any high-usage link is the path the node-to-node traffic load between the nodes would follow if the high-usage link did not exist. In Figure 2.4c, this is AN2-BB1-AN1.

2.5 Time-Dependent Routing (TDR) Path Selection

TDR methods are a type of dynamic routing in which the routing tables are altered at a fixed point in time during the day or week. TDR routing tables are determined on an off-line, preplanned basis and are implemented consistently over a time period. The TDR routing tables are determined considering the time variation of traffic load in the network, for example based on measured hourly load patterns. Several TDR time periods are used to divide up the hours on an average business day and weekend into contiguous routing intervals sometimes called load set periods. Typically, the TDR routing tables used in the network are coordinated by taking advantage of noncoincidence of busy hours among the traffic loads.

In TDR, the routing tables are preplanned and designed off-line using a centralized bandwidth broker, which employs a TDR network design model. Such models are discussed in ANNEX 6. The off-line computation determines the optimal routes from a very large number of possible alternatives, in order to maximize network throughput and/or minimize the network cost. The designed routing tables are loaded and stored in the various nodes in the TDR network, and periodically recomputed and updated (e.g., every week) by the bandwidth broker. In this way an ON does not require additional network information to construct TDR routing tables, once the routing tables have been loaded. This is in contrast to the design of routing tables on-line in real time, such as in the SDR and EDR methods described below. Paths in the TDR routing table may consist of time varying routing choices and use a subset of the available paths. Paths used in various time periods need not be the same.

Paths in the TDR routing table may consist of the direct link, a 2-link path through a single VN, or a multiple-link path through multiple VNs. Path routing implies selection of an entire path between originating and terminating nodes before a connection is actually attempted on that path. If a connection on one link in a path is blocked (e.g., because of insufficient bandwidth), the connection request then attempts another complete path. Implementation of such a routing method can be done through control from the originating node, plus a multiple-link crankback capability to allow paths of two, three, or more links to be used. Crankback is an information-exchange message capability that allows a connection request blocked on a link in a path to return to the originating node for further alternate routing on other paths. Path-to-path routing is nonhierarchical and allows the choice of the most economical paths rather than being restricted to hierarchical paths.

Path selection rules employed in TDR routing tables, for example,
consist of simple sequential routing. In the sequential method all traffic in a given time period is offered to a single route, and lets the first path in the route overflow to the second path which overflows to the third path, and so on. Thus, traffic is routed sequentially from path to path, and the route is allowed to change from hour to hour to achieve the preplanned dynamic, or time varying, nature of the TDR method.

Other TDR path selection rules can employ probabilistic techniques to select each path in the route and thus influence the realized flows. One such method of implementing TDR multilink path selection is to allocate fractions of the traffic to routes and to allow the fractions to vary as a function of time. One approach is cyclic path selection, illustrated in Figure 2.4a, which has as its first route \((i, 2, \ldots, M)\), where the notation \((i, j, k)\) means that all traffic is offered first to path \(i\), which overflows to path \(j\), which overflows to path \(k\). The second route of a cyclic route choice is a cyclic permutation of the first route: \((2, 3, \ldots, M, 1)\). The third route is likewise \((3, 4, \ldots, M, 1, 2)\), and so on. This approach has computational advantages because its cyclic structure requires considerably fewer calculations in the design model than does a general collection of paths.

The route congestion level of cyclic routes are identical; what varies from route to route is the proportion of flow on the various links.

Two-link TDR path selection is illustrated in Figure 2.4a. An example implementation is 2-link sequential TDR (2S-TDR) path selection. By using the crankback signal, 2S-TDR limits path connections to at most two links, and, in meshed network topologies, such TDR 2-link sequential path selection allows nearly as much network utilization and performance improvement as TDR multilink path selection. This is because in the design of multilink path routing in meshed networks, about 98 percent of the traffic is routed on one- and 2-link paths, even though paths of greater length are allowed.

Because of switching costs, paths with one or two links are usually less expensive than paths with more links. Therefore, as illustrated in Figure 2.4a, 2-link path routing uses the simplifying restriction that paths can have only one or two links, which requires only single-link crankback to implement and uses no common links as is possible with multilink path routing. Alternative 2-link path selection methods include the cyclic routing method described above and sequential routing.

In sequential routing, all traffic in a given hour is offered to a single route, and the first path is allowed to overflow to the second path, which overflows to the third path, and so on. Thus, traffic is routed sequentially from path to path with no probabilistic methods being used to influence the realized flows. The reason that sequential routing works well is that permuting path order provides sufficient flexibility to achieve desired flows without the need for probabilistic routing. In 2S-TDR, the sequential route is allowed to change from hour to hour. The TDR nature of the dynamic path selection method is achieved by introducing several route choices, which consist of different sequences of paths, and each path has one or, at most, two links in tandem.

Paths in the routing table are subject to depth-of-search (DoS) restrictions for QoS resource management, which is discussed in ANNEX 3. DoS requires that the bandwidth capacity available on each link in the path be sufficient to meet a DoS bandwidth threshold level, which is passed to each node in the path in the setup message. DoS restrictions prevent connections that path on the first-choice or primary (often the shortest) ON-DN path, for example, from being swamped by alternate routed multiple-link connections.

A TDR connection set-up example is now given. The first step is for the ON
to identify the DN and routing table information to the DN. The ON then
tests for spare capacity on the first or shortest path, and in doing this
supplies the VNs and DN on this path, along with the DoS parameter, to all
nodes in the path. Each VN tests the available bandwidth capacity on each
link in the path against the DoS threshold. If there is sufficient
capacity, the VN forwards the connection setup to the next node, which
performs a similar function. If there is insufficient capacity, the VN
sends a release message with crankback/bandwidth-not-available parameter
back to the ON, at which point the ON tries the next path in the route as
determined by the routing table rules. As described above, the TDR routes
are preplanned off-line, and then loaded and stored in each ON.

Allocating traffic to the optimum path choice during each time period leads
to design benefits due to the noncoincidence of loads. Since in many network
applications traffic demands change with time in a reasonably predictable
manner, the routing also changes with time to achieve maximum link
utilization and minimum network cost. Several TDR routing time periods are
used to divide up the hours on an average business day and weekend into
contiguous routing intervals. The network design is performed in an
off-line, centralized computation in the bandwidth broker that determines the
optimal routing tables from a very large number of possible alternatives
in order to minimize the network cost. In TDR path selection, rather than
determine the optimal routing tables based on real-time information, a
centralized bandwidth broker design system employs a design model, such as
described in ANNEX 6. The effectiveness of the design depends on how
accurately we can estimate the traffic load on the network. Forecast errors
are corrected in the short-term capacity management process, which allows
routing table updates to replace link augments whenever possible, as
described in ANNEX 7.

2.6 State-Dependent Routing (SDR) Path Selection

In SDR, the routing tables are altered automatically according to the state
of the network. For a given SDR method, the routing table rules are
implemented to determine the path choices in response to changing network
status, and are used over a relatively short time period. Information on
network status may be collected at a central bandwidth broker processor or
distributed to nodes in the network. The information exchange may be
performed on a periodic or on-demand basis. SDR methods use the principle
of routing connections on the best available path on the basis of network
state information. For example, in the least loaded routing (LLR) method,
the residual capacity of candidate paths is calculated, and the path having
the largest residual capacity is selected for the connection. Various
relative levels of link occupancy can be used to define link load states,
such as lightly-loaded, heavily-loaded, or bandwidth-not-available states.
Methods of defining these link load states are discussed in ANNEX 3. In
general, SDR methods calculate a path cost for each connection request based
on various factors such as the load-state or congestion state of the links
in the network.

In SDR, the routing tables are designed on-line by the ON or a central
bandwidth broker processor (BBP) through the use of network status and
topology information obtained through information exchange with other nodes
and/or a centralized BBP. There are various implementations of SDR
distinguished by

a) whether the computation of the routing tables is distributed among
the network nodes or centralized and done in a centralized BBP, and
b) whether the computation of the routing tables is done periodically or connection by connection.

This leads to three different implementations of SDR:

a) centralized periodic SDR (CP-SDR) -- here the centralized BBP obtains link status and traffic status information from the various nodes on a periodic basis (e.g., every 10 seconds) and performs a computation of the optimal routing table on a periodic basis. To determine the optimal routing table, the BBP executes a particular routing table optimization procedure such as LLR and transmits the routing tables to the network nodes on a periodic basis (e.g., every 10 seconds).

b) distributed periodic SDR (DP-SDR) -- here each node in the SDR network obtains link status and traffic status information from all the other nodes on a periodic basis (e.g., every 5 minutes) and performs a computation of the optimal routing table on a periodic basis (e.g., every 5 minutes). To determine the optimal routing table, the ON executes a particular routing table optimization procedure such as LLR.

c) distributed connection-by-connection (DC-SDR) SDR -- here an ON in the SDR network obtains link status and traffic status information from the DN, and perhaps from selected VNs, on a connection by connection basis and performs a computation of the optimal routing table for each connection. To determine the optimal routing table, the ON executes a particular routing table optimization procedure such as LLR.

In DP-SDR path selection, nodes may exchange status and traffic data, for example, every five minutes, between traffic management processors, and based on analysis of this data, the traffic management processors can dynamically select alternate paths to optimize network performance. This method is illustrated in Figure 2.4b. Flooding is a common technique for distributing the status and traffic data, however other techniques with less overhead are also available, such as a query-for-status method, as discussed in ANNEX 4.

Figure 2.4b illustrates a CP-SDR path selection method with periodic updates based on periodic network status. CP-SDR path selection provides near-real-time routing decisions by having an update of the idle bandwidth on each link sent to a network database every five seconds. Routing tables are determined from analysis of the status data using a path selection method which provides that the shortest path choice is used if the bandwidth is available. If the shortest path is busy (e.g., bandwidth is unavailable on one or more links), the second path is selected from the list of feasible paths on the basis of having the greatest level of idle bandwidth at the time; the current second path choice becomes the third, and so on. This path update is performed, for example, every five seconds. The CP-SDR model uses dynamically activated bandwidth reservation and other controls to automatically modify routing tables during network overloads and failures. CP-SDR requires the use of network status and routing recommendation information-exchange messages.

Figure 2.4b also illustrates an example of a DC-SDR path selection method. In DC-SDR, the routing computations are distributed among all the nodes in the network. DC-SDR uses real-time exchange of network status information, such as with query and status messages, to determine an optimal path from a very large number of possible choices. With DC-SDR, the originating node first tries the primary path and if it is not available finds an optimal alternate path by querying the destination node and perhaps several via
nodes through query-for-status network signaling for the busy-idle load status of all links connected on the alternate paths to the destination node. The originating node then finds the least loaded alternate path to route the connection request. DC-SDR computes required bandwidth allocations by virtual network from node-measured traffic flows and uses this capacity allocation to reserve capacity when needed for each virtual network. Any excess traffic above the expected flow is routed to temporarily idle capacity borrowed from capacity reserved for other loads that happen to be below their expected levels. Idle link capacity is communicated to other nodes via the query-status information-exchange messages, as illustrated in Figure 2.4b, and the excess traffic is dynamically allocated to the set of allowed paths that are identified as having temporarily idle capacity. DC-SDR controls the sharing of available capacity by using dynamic bandwidth reservation, as described in ANNEX 3, to protect the capacity required to meet expected loads and to minimize the loss of traffic for classes-of-service which exceed their expected load and allocated capacity.

Paths in the SDR routing table may consist of the direct link, a 2-link path through a single VN, or a multiple-link path through multiple VNs. Paths in the routing table are subject to DoS restrictions on each link.

2.7 Event-Dependent Routing (EDR) Path Selection

In EDR, the routing tables are updated locally on the basis of whether connections succeed or fail on a given path choice. In the EDR learning approaches, the path last tried, which is also successful, is tried again until blocked, at which time another path is selected at random and tried on the next connection request. EDR path choices can also be changed with time in accordance with changes in traffic load patterns. Success-to-the-top (STT) EDR path selection, illustrated in Figure 2.4b, is a decentralized, on-line path selection method with update based on random routing. STT-EDR uses a simplified decentralized learning method to achieve flexible adaptive routing. The primary path path-p is used first if available, and a currently successful alternate path path-s is used until it is blocked. In the case that path-s is blocked (e.g., bandwidth is not available on one or more links), a new alternate path path-n is selected at random as the alternate path choice for the next connection request overflow from the primary path. As described in ANNEX 3, dynamically activated bandwidth reservation is used under congestion conditions to protect traffic on the primary path. STT-EDR uses crankback when an alternate path is blocked at a via node, and the connection request advances to a new random path choice. In STT-EDR, many path choices can be tried by a given connection request before the request is blocked.

In the EDR learning approaches, the current alternate path choice can be updated randomly, cyclically, or by some other means, and may be maintained as long as a connection can be established successfully on the path. Hence the routing table is constructed with the information determined during connection setup, and no additional information is required by the ON. Paths in the EDR routing table may consist of the direct link, a 2-link path through a single VN, or a multiple-link path through multiple VNs. Paths in the routing table are subject to DoS restrictions on each link. Note that for either SDR or EDR, as in TDR, the alternate path for a connection request may be changed in a time-dependent manner considering the time-variation of the traffic load.

2.8 Interdomain Routing
In current practice, interdomain routing protocols generally do not incorporate standardized path selection or per class-of-service resource management. For example, in IP-based networks BGP [RL00] is used for interdomain routing but does not incorporate per class-of-service resource allocation as described in this Section. Also, MPLS techniques have not yet been addressed for interdomain applications. Extensions to interdomain routing methods discussed in this Section therefore can be considered to extend the call routing and connection routing concepts to routing between network domains.

Many of the principles described for intradomain routing can be extended to interdomain routing. As illustrated in Figure 2.5, interdomain routing paths can be divided into three types:

* a primary shortest path between the originating domain and destination domain,
* alternate paths with all nodes in the origination domain and destination domain, and
* alternate or transit paths through other transit domains.

Interdomain routing can support a multiple ingress/egress capability, as illustrated in Figure 2.5 in which a connection request is routed either on the shortest path or, if not available, via an alternate path through any one of the other nodes from an originating node to a gateway node.

Within an originating network, a destination network could be served by more than one gateway node, such as OGN1 and OGN2 in Figure 2.5, in which case multiple ingress/egress routing is used. As illustrated in Figure 2.5, with multiple ingress/egress routing, a connection request from the originating node N1 destined for the destination gateway node DGN1 tries first to access the links from originating gateway node OGN2 to DGN1. In doing this it is possible that the connection request could be routed from N1 to OGN2 directly or via N2. If no bandwidth is available from OGN2 to DGN1, the control of the connection request can be returned to N1 with a crankback/bandwidth-not-available indicator, after which the connection request is routed to OGN1 to access the OGN1-to-DGN1 bandwidth. If the connection request cannot be completed on the link connecting gateway node OGN1 to DGN1, the connection request can return to the originating node N1 through use of a crankback/band-not-available indicator for possible further routing to another gateway node (not shown). In this manner all ingress/egress connectivity is utilized to a connecting network, maximizing connection request completion and reliability.

Once the connection request reaches an originating gateway node (such as OGN1 or OGN2), this node determines the routing to the destination gateway node DGN1 and routes the connection request accordingly. In completing the connection request to DGN1, an originating gateway node can dynamically select a direct shortest path, an alternate path through an alternate node in the destination network, or perhaps an alternate path through an alternate node in another network domain. Hence, with interdomain routing, connection requests are routed first to a shortest primary path between the originating and destination domain, then to a list of alternate paths through alternate nodes in the terminating network domain, then to a list of alternate paths through alternate nodes in the originating network domain (e.g., OGN1 and OGN2 in Figure 2.5), and finally to a list of alternate paths through nodes in other transit network domains.

Examples of alternate paths which might be selected through a transit network domain are N1-OGN1-VGN1-DGN1, N1-OGN1-VGN2-DGN1, or
N1-N2-OGN2-VGN2-DGN1 in Figure 2.5. Such paths through transit network
domains may be tried last in the example network configuration in the
Figure. For example, flexible interdomain routing may try to find an
available alternate path based on link load states, where known, and
connection request completion performance, where it can be inferred. That
is, the originating gateway node (e.g., node OGN1 in Figure 2.5) may use its
link status to a via node in a transit domain (e.g., links OGN1-VGN1 and
OGN1-VGN2) in combination with the connection request completion performance
from the candidate via node to the destination node in the destination
network domain, in order to find the most available path to route the
connection request over. For each path, a load state and a completion state
are tracked. The load state indicates whether the link bandwidth from the
gateway node to the via node is lightly loaded, heavily loaded, reserved, or
busy. The completion state indicates whether a path is achieving
above-average completion, average completion, or below-average completion.
The selection of a via path, then, is based on the load state and completion
state. Alternate paths in the same destination network domain and in a
transit network domain are each considered separately. During times of
congestion, the link bandwidth to a candidate via node may be in a reserved
state, in which case the remaining link bandwidth is reserved for traffic
routing directly to the candidate via node. During periods of no congestion,
capacity not needed by one virtual network is made available to other
virtual networks that are experiencing loads above their allocation.

Completion rate is tracked on the various via paths (such as the path
through via node VGN1 or VGN2 to destination node DGN1 in Figure 2.5) by
taking account of the information relating either the successful completion
or non-completion of a connection request through the via node. A
non-completion, or failure, is scored for the connection request if a
signaling release message is received from the far end after the connection
request seizes an egress link, indicating a network in-completion cause
value. If no such signaling release message is received after the connection
request seizes capacity on the egress link, then the connection request is
scored as a success. Each gateway node keeps a connection request
completion history of the success or failure, for example, of the last 10
connection requests using a particular via path, and it drops the oldest
record and adds the connection request completion for the newest connection
request on that path. Based on the number of connection request completions
relative to the total number of connection requests, a completion state is
computed.

Based on the completion states, connection requests are normally routed on
the first path with a high completion state with a lightly loaded egress
link. If such a path does not exist, then a path having an average
completion state with a lightly loaded egress link is selected, followed by
a path having a low completion state with a lightly loaded egress link. If
no path with a lightly loaded egress link is available, and if the search
depth permits the use of a heavily loaded egress link, the paths with
heavily loaded egress links are searched in the order of high completion,
average completion, and low completion. If no such paths are available,
paths with reserved egress links are searched in the same order, based on
the connection request completion state, if the search depth permits the use
of a reserved egress link.
The rules for selecting primary shortest paths and alternate paths for a connection request are governed by the availability of shortest path bandwidth and node-to-node congestion. The path sequence consists of the primary shortest path, lightly loaded alternate paths, heavily loaded alternate paths, and reserved alternate paths. Alternate paths are first selected which include nodes only in the originating and destination domains, and then selected through transit domains if necessary.

Thus we have illustrated that interdomain routing methods can be considered to extend the intradomain call routing and connection routing concepts, such as flexible path selection and per-class-of-service bandwidth selection, to routing between network domains.

2.9 Modeling of Traffic Engineering Methods

In the draft, a full-scale national network node model is used together with a multiservice traffic demand model to study various TE scenarios and tradeoffs. The 135-node national model is illustrated in Figure 2.6.

Typical voice/ISDN traffic loads are used to model the various network alternatives, which are based on 72 hours of a full-scale national network loading. Tables 2.1a, 2.1b, and 2.1c summarize the multiservice traffic model used for the TE studies. Three levels of traffic priority -- key, normal, and best-effort -- are given to the various class-of-service categories, or virtual networks (VNETs), illustrated in Tables 2.1a-c. Class-of-service, traffic priority, and QoS resource management are all discussed further in ANNEX 3.

The voice/ISDN loads are further segmented in the model into eight constant-bit-rate (CBR) VNETs, including business voice, consumer voice, international voice in and out, key-service voice, normal and key-service 64-kbps ISDN data, and 384-kbps ISDN data. For the CBR voice services, the mean data rate is assumed to be 64 kbps for all VNETs except the 384-kbps ISDN data VNET-8, for which the mean data rate is 384 kbps.

Table 2.1a
Virtual Network (VNET) Traffic Model used for TE Studies

Table 2.1b
Virtual Network (VNET) Traffic Model used for TE Studies
Average Number of Flows by Network Busy Hours (CST)

Table 2.1c
Virtual Network (VNET) Traffic Model used for TE Studies
Average Data Volume (Mbps) by Network Busy Hours (CST)

The data services traffic model incorporates typical traffic load patterns and comprises three additional VNET load patterns. These data services VNETs include

* variable bit rate real-time (VBR-RT) VNET-9, representing services such as IP-telephony and compressed voice,
* variable bit rate non-real-time (VBR-NRT) VNET-10, representing services such as WWW multimedia and credit card check, and
* unassigned bit rate (UBR) VNET-11, representing services such as email, voice mail, and file transfer multimedia applications.

For the VBR-RT connections, the data rate varies from 6.4 to 51.2 kbps with
a mean of 25.6 kbps. The VBR-RT connections are assumed to be interactive and delay sensitive. For the VBR-NRT connections, the data rate varies from 38.4 to 64 kbps with a mean of 51.2 kbps, and the VBR-NRT flows are assumed to be non-delay sensitive. For the UBR connections, the data rate varies from 6.4 to 3072 kbps with a mean of 1536 kbps. The UBR flows are assumed to be best-effort priority and non-delay sensitive. For modeling purposes, the service and link bandwidth is segmented into 6.4 kbps slots, that is, 10 slots per 64 kbps channel.

Here the traffic loads are dynamically varying and tracked by the exponential smoothing models discussed in ANNEX 3. Table 2.1b gives the average number of flows for each class of service (VNET) in various network busy hours, and Table 2.1c gives the average data volume in Mbps for each class of service (VNET) in various network busy hours. We can see that the voice/isdn traffic (i.e., VNETs 1-8 in Tables 2.1a-c) has a majority of the flows (approximately 75%) of the total in Monday busy hours, compared to the various "data" traffic sources (i.e., VNETs 9-11 in Tables 2.1a-c). However the voice/isdn traffic has a minority of the total traffic data volume (approximately 70-80%) of the total Mbps demand, compared to the various "data" traffic sources (i.e., VNETs 9-11 in Tables 2.1a-c). The model is based on traffic projections for the "data" traffic and actual voice/isdn traffic levels, wherein the data traffic dominating the voice/isdn traffic is a realistic scenario under many traffic projections.

Note also the time variation of the various classes of service. The business voice traffic (VNET-1) peaks in the Monday daytime busy hours and is considerably lower in intensity in the Sunday and Monday evening hours. The consumer voice traffic (VNET-2) follows an opposite pattern. The best effort traffic (VNET-11) follows a pattern similar to business voice, peaking in the Monday daytime hours, but of course is much larger in data volume (Mbps) compared to the voice traffic. On the other hand, the IP multimedia traffic (VNET-10) follows a pattern similar to the consumer voice traffic. These noncoincident traffic patterns make it more efficient to carry them on the same multiservice network, wherein capacity can be used more efficiently because it is shared across different time periods. There are classes of service with large data volumes and small data volumes, which all must be combined in the multiservice network, and meet their QoS requirements.

The cost model represents typical switching and transport costs, and illustrates the economies-of-scale for costs projected for high capacity network elements in the future. Table 2.2 gives the model used for average switching and transport costs allocated per 64 kbps unit of bandwidth, as follows:

Table 2.2
Cost Assumptions (average cost per equivalent 64 kbps bandwidth)

A discrete event network design model, described in ANNEX 6, is used in the design and analysis of 5 connection routing methods with TE methods applied: 2-STT-EDR path routing in a meshed logical network, 2-link DC-SDR routing in a meshed logical network, and multilink STT-EDR, DC-SDR, and DP-SDR routing, as might be supported for example by MPLS TE in a sparse logical network. We also model the case where no TE call and connection routing methods are applied.

The network models for the 2-link STT-EDR/DC-SDR, and multilink STT-EDR/DC-SDR/DP-SDR networks are now described. In the 2-link STT-EDR and DC-SDR models, we assume 135 packet-switched-nodes (MPLS- or PNNI-based).
Synchronous to asynchronous conversion (SAC) is assumed to occur at the packet-switched-nodes for link connections from circuit-switched-nodes. Links in these 2-link STT-EDR/DC-SDR models are assumed to provide more fine-grained (1.536 mbpsT1-level) logical link bandwidth allocation, and a meshed network topology design results among the nodes, that is, links exist between most (90 percent or more) of the nodes. In the 2-link STT-EDR/DC-SDR models, one and 2-link routing with crankback is used throughout the network. Two-link path selection is modeled both with both STT path selection and distributed connection-by-connection SDR (DC-SDR) path selection. Packet-switched-nodes use 2-link STT-EDR or 2-link DC-SDR routing to all other nodes. Quality-of-service priority queuing is modeled in the performance analyses, in which the key-services are given the highest priority, normal services the middle priority, and best-effort services the lowest priority in the queuing model. This queuing model quantifies the level of delayed traffic for each virtual network. In routing a connection with 2-link STT-EDR routing, the ON checks the equivalent bandwidth and allowed DoS first on the direct path, then on the current successful 2-link via path, and then sequentially on all candidate 2-link paths. In routing a connection with 2-link DC-SDR, the ON checks the equivalent bandwidth and allowed DoS first on the direct path, and then on the least-loaded path that meets the equivalent bandwidth and DoS requirements. Each VN checks the equivalent bandwidth and allowed DoS provided in the setup message, and uses crankback to the ON if the equivalent bandwidth or DoS are not met.

In the multilink STT-EDR/DC-SDR/DP-SDR model, we assume 135 packet-switched-nodes. Because high rate OC3/12/48 links provide highly aggregated link bandwidth allocation, a sparse network topology design results among the packet-switched-nodes, that is, high rate OC3/12/48 links exist between relatively few (10 to 20 percent) of the packet-switched-nodes. Secondly, multilink shortest path selection with crankback is used throughout the network. For example, the STT EDR TE algorithm used is adaptive and distributed in nature and uses learning models to find good paths for TE in a network. With STT EDR, if the LSR-A to LSR-B bandwidth needs to be modified, say increased by delta-BW, the primary LSP-p is tried first. If delta-BW is not available on one or more links of LSP-p, then the currently successful LSP-s is tried next. If delta-BW is not available on one or more links of LSP-s, then a new LSP is searched by trying additional candidate paths until a new successful LSP-n is found or the candidate paths are exhausted. LSP-n is then marked as the currently successful path for the next time bandwidth needs to be modified.

Quality-of-service priority queuing is modeled in the performance analyses, in which the key-services are given the highest priority, normal services the middle priority, and best-effort services the lowest priority in the queuing model. This queuing model quantifies the level of delayed traffic for each virtual network. The multilink path selection options are modeled with STT path selection, DC-SDR path selection, and distributed periodic path selection (DP-SDR). In the model of DP-SDR, the status updates, which are made with flooding link status updates every 10 seconds. Note that the multilink DP-SDR performance results should also be comparable to the performance of multilink centralized-periodic SDR (CP-SDR), in which status updates and path selection updates are made every 10 seconds, respectively, to and from a bandwidth-broker processor.

In routing a connection with multilink shortest path selection with 2-link STT-EDR routing, for example, the ON checks the equivalent bandwidth and allowed DoS first on the first choice path, then on current successful alternate path, and then sequentially on all candidate alternate paths. Again, each VN checks the equivalent bandwidth and allowed DoS provided in
the setup message, and uses crankback to the ON if the equivalent bandwidth or DoS are not met.

In the models the logical network design is optimized for each routing alternative, while the physical transport links and node locations are held fixed. We examine the performance and network design tradeoffs of

* logical topology design (sparse or mesh), and
* routing method (2-link, multilink, fixed, dynamic, SDR, EDR, hierarchical, nonhierarchical, etc.)

Generally the meshed logical topologies are optimized by 1- and 2-link routing, while the sparse logical topologies are optimized by multilink shortest path routing. Modeling results include

* designs for dynamic 2-link routing (SDR, EDR) and multilink routing (SDR, EDR),
* designs for voice/ISDN-only traffic (VNETs 1-8 in Table 2.1) and data-only traffic (VNETs 9-11)
* designs for integrated voice/ISDN and data traffic (VNETs 1-11)
* designs for fixed hierarchical routing
* designs where all voice traffic is compressed (VNETs 1-5 and VNET 9 all use the IP-telephony traffic characteristics of VNET 9)
* performance analyses for overloads and failures

2.9.1 Network Design Comparisons

Illustrative network design costs for the voice/ISDN-only designs (VNETs 1-8 in Table 2.1), for the data-only designs (VNETs 9-11 in Table 2.1), and for the integrated voice/ISDN and data designs (VNETs 1-11 in Table 2.1), are given in Figure 2.7, 2.8, and 2.9, respectively. These design costs and details are discussed further in ANNEX 6.

The design results show that the 2-link STT-EDR and 2-link DC-SDR logical mesh networks are highly connected (90%+), while the multilink MPLS-based and PNNI-based networks are sparsely connected (10-20%). The network cost comparisons illustrate that the sparse MPLS and PNNI networks achieve a small cost advantage, since they take advantage of the greater cost efficiencies of high bandwidth logical links (up to OC48), as reflected in Table 2.2. However, these differences in cost may not be significant, and can change as equipment costs evolve and as the relative cost of switching and transport equipment changes. Sensitivities of the results to different cost assumptions were investigated. For example, if the relative cost of transport increases relative to switching, then the 2-link meshed networks can appear to be more efficient than the sparse multilink networks. These results are consistent with those presented in other studies of meshed and sparse logical networks, as a function of relative switching and transport costs, see for example [A98].

Comparing the results of the separate voice/ISDN and data designs and the integrated voice/ISDN and data designs shows that integration does achieve some capital cost advantage of about 5 to 20 percent. The larger range of integration design efficiencies is achieved as a result of the economies of scale of larger capacity network elements, as reflected in cost assumptions given in Table 2.2. However, probably more significant are the operational savings of integration which result from operating a single network rather than two or more networks. In addition, the performance of an integrated voice and data network leads to advantages in capacity sharing, especially when different traffic classes having different routing priorities, such as
key service and best-effort service, are integrated and share capacity on
the same network. These performance results are reported below. A study of
voice compression for all voice traffic, such as might occur if IP-telephony
is widely deployed, shows that network capital costs might be reduced by as
much as 10% if this evolutionary direction is followed. An analysis of
fixed hierarchical routing versus dynamic routing illustrates that more than
20% reduction in network capital costs can be achieved with dynamic routing.
In addition, operation savings also result from simpler provisioning of the
dynamic routing options.

2.9.2  Network Performance Comparisons

The performance analyses for overloads and failures include connection
request admission control with QoS resource management. As discussed in
ANNEX 3, in the example QoS resource management approach, we distinguish the
key services, normal services, and best-effort services. Performance
comparisons are presented in Tables 2.3, 2.4, and 2.5 for various TE
methods, including 2-link and multilink EDR and SDR approaches, and a
baseline case of no TE methods applied. Table 2.3 gives performance results
for a 30% general overload, Table 2.4 gives performance results for a
six-times overload on a single network node, and Table 2.5 gives performance
results for a single logical-link failure.

Table 2.3
Performance Comparison for Various Connection-Routing TE Methods & No TE
Methods
30% General Overload (% Lost/Delayed Traffic)
(135-Node Multiservice Network Model)

<table>
<thead>
<tr>
<th>TE Methods</th>
<th>Ash</th>
<th>&lt;draft-ietf-tewg-qos-routing-02.txt&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>No TE</td>
<td></td>
<td>[Page ANNEX2-16]</td>
</tr>
</tbody>
</table>

Internet Draft  TE & QoS Methods for IP, ATM, TDM-Based Networks  August 2001

Table 2.4
Performance Comparison for Various Connection-Routing TE Methods & No TE
Methods
6X Focused Overload on OKBK (% Lost/Delayed Traffic)
(135-Node Multiservice Network Model)

Table 2.5
Performance Comparison for Various Connection-Routing TE Methods & No TE
Methods
Failure on CHCG-NYCM Link (% Lost/Delayed Traffic)
(135-Node Multiservice Network Model)

In all cases of the TE methods being applied, the performance is always
better and usually substantially better than when no TE methods are applied.
The performance analysis results show that the multilink
STT-EDR/DC-SDR/DP-SDR options (in sparse topologies) perform somewhat better
under overloads than the 2-link STT-EDR/DC-SDR options (in meshed
topologies), because of greater sharing of network capacity. Under failure,
the 2-link STT-EDR/DC-SDR options perform better for many of the virtual
network categories than the multilink STT-EDR/DC-SDR/CP-SDR options, because
they have a richer choice of alternate routing paths and are much more
highly connected than the multilink STT-EDR/DC-SDR/DP-SDR networks. Loss of
a link in a sparsely connected multilink STT-EDR/DC-SDR/DP-SDR network can
have more serious consequences than in more highly connected logical
networks. The performance results illustrate that capacity sharing of CBR,
VBR, and UBR traffic classes, when combined with QoS resource management
and priority queuing, leads to efficient use of bandwidth with minimal traffic
delay and loss impact, even under overload and failure scenarios. These QoS
resource management trends are further examined in ANNEX 3.
The STT and SDR path selection methods are quite comparable for the 2-link, meshed-topology network scenarios. However, the STT path selection method performs somewhat better than the SDR options in the multilink, sparse-topology case. In addition, the DC-SDR path selection option performs somewhat better than the CP-DCR option in the multilink case, which is a result of the 10-second old status information causing misdirected paths in some cases. Hence, it can be concluded that frequently-updated, available-link-bandwidth (ALB) state information does not necessarily improve performance in all cases, and that if ALB state information is used, it is sometimes better that it is very recent status information.

2.9.3 Single-Area Flat Topology vs. Multi-Area 2-Level Hierarchical Network Topology

We also investigate the performance of hierarchical network designs, which represent the topological configuration to be expected with multi-area (or multi-autonomous-system (multi-AS), or multi-domain) networks. In Figure 2.10 we show the model considered, which consists of 135 edge nodes each homed onto one of 21 backbone nodes. Typically, the edge nodes may be grouped into separate areas or autonomous systems, and the backbone nodes into another area or autonomous system. Within each area a flat routing topology exists, however between edge areas and the backbone area a hierarchical routing relationship exists. This routing hierarchy is modeled in ANNEX 3 for both the per-flow and per-virtual-network bandwidth allocation examples, here the results are given for the per-flow allocation case in Tables 2.6 to 2.8 for the 30% general overload, 6-times focused overload, and link failure examples, respectively. We can see that the performance of the hierarchical network case is substantially worse than the flat network model, which models a single area or autonomous system consisting of 135 nodes.

Table 2.6
Performance of Single-Area Flat vs. Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under 30% General Overload
(Multilink STT-EDR Routing; 135-Node Network Model)

Table 2.7
Performance of Single-Area Flat vs. Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under 6X Focused Overload on OKBK
(Multilink STT-EDR Routing; 135-Node Network Model)

Table 2.8
Performance of Single-Area Flat vs. Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under Failure on CHCG-NYCM Link
(Multilink STT-EDR Routing; 135-Node Network Model)

2.9.4 Network Modeling Conclusions

The TE modeling conclusions are summarized as follows:

1. Capital cost advantages may be attributed to the sparse topology options, such as the multilink STT-EDR/DC-SDR/DP-SDR options, but may not be significant compared to operational costs, and are subject to the particular switching and transport cost assumptions. Capacity design models are further detailed in ANNEX 6 and operational issues in ANNEX 7.

2. In all cases of the TE methods being applied, the performance is
always better and usually substantially better than when no TE methods are applied.

3. The sparse-topology multilink-routing networks provide better overall performance under overload, but performance under failure may favor the 2-link STT-EDR/DC-SDR options with more alternate routing choices. One item of concern in the sparse-topology multilink-routing networks is with post dial delay, in which perhaps 5 or more links may need to be connected for an individual connection request.

4. Single-area flat topologies exhibit better network performance and, as discussed and modeled in ANNEX 6, greater design efficiencies in comparison with multi-area hierarchical topologies. As illustrated in ANNEX 4, larger administrative areas can be achieved through use of EDR-based TE methods as compared to SDR-based TE methods.

5. State information as used by the 2-link and multilink SDR options provides only a small network capital cost advantage, and essentially equivalent performance to the 2-link STT-EDR options, as illustrated in the network performance results.

6. Various path selection methods can interwork with each other in the same network, which is required for multi-vendor network operation.

7. QoS resource management, as further described in ANNEX 3, is shown to be effective in achieving key service, normal service, and best effort service differentiation.

8. Voice and data integration can provide capital cost advantages, but may be more important in achieving operational simplicity and cost reduction.

9. If IP-telephony takes hold and a significant portion of voice calls use voice compression technology, this could lead to more efficient networks.

Overall the packet-based (e.g., MPLS/TE) multilink, sparse-topology routing strategies offer several advantages. The sparse logical topology with the high-speed switching and transport links may have economic benefit due to lower cost network designs achieved by the economies of scale of higher rate network elements. The sparse, high-bandwidth, logical-link networks have been shown to have better response to overload conditions than logical mesh networks, due to greater sharing of network capacity. The packet-based routing protocols have capabilities for automatic provisioning of links, nodes, and reachable addresses, which provide operational advantages for such networks. Because the sparse high-bandwidth-link network designs have dramatically fewer links to provision compared to mesh network designs (10-20% connected versus 90% or more connected for mesh networks), there is less provisioning work to perform. In addition to having fewer links to provision, sparse high-bandwidth-link network designs use larger increments of capacity on individual links and therefore capacity additions would need to occur less frequently than in highly connected mesh networks, which would have much smaller increments of capacity on the individual links. The sparse-topology, multilink-routing methods are synergistic with evolution of data network services which implement these protocols, and such routing methods have been in place for many years in data networks. Should a service provider pursue integration of the voice/ISDN and data services networks, these factors will help support such an integration direction.

2.10 Conclusions/Recommendations
We have discussed call routing and connection routing methods employed in TE functions. Several connection routing alternatives were discussed, which include FR, TDR, EDR, and SDR methods. Models were presented to illustrate the network design and performance tradeoffs between the many TE approaches explained in the ANNEX, and conclusions were drawn on the advantages of various routing and topology options in network operation. Overall the packet-based (e.g., MPLS/TE) multilink, sparse-topology routing strategies were found to offer several advantages.

The following conclusions/recommendations are reached in the ANNEX:

* TE methods are recommended to be applied, and in all cases of the TE methods being applied, network performance is always better and usually substantially better than when no TE methods are applied.

* Sparse-topology multilink-routing networks are recommended and provide better overall performance under overload than meshed-topology networks, but performance under failure may favor the 2-link STT-EDR/DC-SDR meshed-topology options with more alternate routing choices.

* Single-area flat topologies are recommended and exhibit better network performance and, as discussed and modeled in ANNEX 6, greater design efficiencies in comparison with multi-area hierarchical topologies. As illustrated in ANNEX 4, larger administrative areas can be achieved through use of EDR-based TE methods as compared to SDR-based TE methods.

* Event-dependent-routing (EDR) TE path selection methods are recommended and exhibit comparable or better network performance compared to state-dependent-routing (SDR) methods.

a. EDR TE methods are shown to an important class of TE algorithms. EDR TE methods are distinct from the TDR and SDR TE methods in how the paths (e.g., MPLS label switched paths, or LSPs) are selected. In the SDR TE case, the available link bandwidth (based on LSA flooding of ALB information) is typically used to compute the path. In the EDR TE case, the ALB information is not needed to compute the path, therefore the ALB flooding does not need to take place (reducing the overhead).

b. EDR TE algorithms are adaptive and distributed in nature and typically use learning models to find good paths for TE in a network. For example, in a success-to-the-top (STT) EDR TE method, if the LSR-A to LSR-B bandwidth needs to be modified, say increased by delta-BW, the primary LSP-p is tried first. If delta-BW is not available on one or more links of LSP-p, then the currently successful LSP-s is tried next. If delta-BW is not available on one or more links of LSP-s, then a new LSP is searched by trying additional candidate paths until a new successful LSP-n is found or the candidate paths are exhausted. LSP-n is then marked as the currently successful path for the next time bandwidth needs to be modified. The performance of distributed EDR TE methods is shown to be equal to or better than SDR methods, centralized or distributed.

c. While SDR TE models typically use available-link-bandwidth (ALB) flooding for TE path selection, EDR TE methods do not require ALB flooding. Rather, EDR TE methods typically search out capacity by learning models, as in the STT method above. ALB flooding can be very resource intensive, since it requires link bandwidth to carry LSAs, processor capacity to process LSAs, and the overhead can limit area/autonomous system (AS) size. Modeling results show EDR TE methods can lead to a large
reduction in ALB flooding overhead without loss of network throughput performance [as shown in ANNEX 4].

d. State information as used by the SDR options (such as with link-state flooding) provides essentially equivalent performance to the EDR options, which typically used distributed routing with crankback and no flooding.

e. Various path selection methods can interwork with each other in the same network, as required for multi-vendor network operation.

* Interdomain routing methods are recommended which extend the intradomain call routing and connection routing concepts, such as flexible path selection and per-class-of-service bandwidth selection, to routing between network domains.

Figures:

Figure 2.1a AESA Address Structure
Figure 2.1b IP Address Structure
Figure 2.2 Sparse Logical Network Topology with Connections Routed on Multilink Paths
Figure 2.3 Mesh Logical Network Topology with Connections Routed on 1- and 2-Link Paths
Figure 2.4a TDR Dynamic Path Selection Methods
Figure 2.4b EDR & SDR Dynamic Path Selection Methods
Figure 2.4c Hierarchical Fixed Routing Path Selection Methods (2-Level Hierarchical Network)
Figure 2.5 Multiple Ingress/Egress Interdomain Routing
Figure 2.6 135-Node National Network Model
Figure 2.7 Voice/ISDN Network Design Cost (includes traffic for VNET-1 to VNET-8 in Table 2.1)

Figure 2.8 Data Network Design Cost (includes traffic for VNET-9 to VNET-11 in Table 2.1)
Figure 2.9 Integrated Voice/ISDN & Data Network Design Cost (includes traffic for VNET-1 to VNET-11 in Table 2.1)
Figure 2.10 Hierarchical Network Model

ANNEX 3
QoS Resource Management Methods

Traffic Engineering & QoS Methods for IP-, ATM-, & TDM-Based Multiservice Networks

3.1 Introduction

QoS resource management (sometimes called QoS routing) functions include class-of-service identification, routing table derivation, connection admission, bandwidth allocation, bandwidth protection, bandwidth reservation, priority routing, priority queuing, and other related resource management functions. QoS resource management methods have been applied successfully in TDM-based networks [A98], and are being extended to IP-based and ATM-based networks. In an illustrative QoS resource management method,
Bandwidth is allocated in discrete changes to each of several virtual networks (VNETs), which are each assigned a priority corresponding to either high-priority key services, normal-priority services, or best-effort low-priority services. Examples of services within these VNET categories include:

* high-priority key services such as defense voice communication,
* normal-priority services such as constant rate, interactive, delay-sensitive voice; variable rate, interactive, delay-sensitive IP-telephony; and variable rate, non-interactive, non-delay-sensitive WWW file transfer, and
* low-priority best-effort services such as variable rate, non-interactive, non-delay-sensitive voice mail, email, and file transfer.

Bandwidth changes in VNET bandwidth capacity can be determined by edge nodes on a per-flow (per-connection) basis, or based on an overall aggregated bandwidth demand for VNET capacity (not on a per-connection demand basis). In the latter case of per-VNET bandwidth allocation, based on the aggregated bandwidth demand, edge nodes make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth, such as on the constraint-based routing label switched paths (CRLSPs) constituting the VNET bandwidth capacity.

In the illustrative QoS resource management method, which we assume is MPLS-based, the bandwidth allocation control for each VNET CRLSP is based on estimated bandwidth needs, bandwidth use, and status of links in the CRLSP. The edge node, or originating node (ON), determines when VNET bandwidth needs to be increased or decreased on a CRLSP, and uses an illustrative MPLS CRLSP bandwidth modification procedure to execute needed bandwidth allocation changes on VNET CRLSPs. In the bandwidth allocation procedure the constraint-based routing label distribution protocol (CRLDP) [J00] or the resource reservation protocol (RSVP-TE) [AGBLSS00] could be used, for example, to specify appropriate parameters in the label request message a) to request bandwidth allocation changes on each link in the CRLSP, and b) to determine if link bandwidth can be allocated on each link in the CRLSP. If a link bandwidth allocation is not allowed, a notification message with an illustrative crankback parameter allows the ON to search out possible bandwidth allocation on another CRLSP. In particular, we illustrate an optional depth-of-search (DoS) parameter in the label request message to control the bandwidth allocation on individual links in a CRLSP. In addition, we illustrate an optional modify parameter in the label request message to allow dynamic modification of the assigned traffic parameters (such as peak data rate, committed data rate, etc.) of an already existing CRLSP. Finally, we illustrate a crankback parameter in the notification message to allow an edge node to search out additional alternate CRLSPs when a given CRLSP cannot accommodate a bandwidth request.

QoS resource management therefore can be applied on a per-flow (or per-call or per-connection-request) basis, or can be beneficially applied to traffic trunks (also known as "bandwidth pipes" or "virtual trunking") in the form of CRLSPs in IP-based networks or SVPs in ATM-based networks.

QoS resource management provides integration of services on a shared network, for many classes-of-service such as:

* CBR services including voice, 64-, 384-, and 1,536-kbps N-ISDN switched digital data, international switched transit, priority defense communication, virtual private network, 800/free-phone, fiber preferred, and other services.
Real-time VBR services including IP-telephony, compressed video, and other services.
Non-real-time VBR services including WWW file transfer, credit card check, and other services.
UBR services including voice mail, email, file transfer, and other services.

We now illustrate the principles of QoS resource management, which includes integration of many traffic classes, as discussed above.

3.2 Class-of-Service Identification, Policy-Based Routing Table Derivation, & QoS Resource Management Steps

QoS resource management functions include class-of-service identification, routing table derivation, connection admission, bandwidth allocation, bandwidth protection, bandwidth reservation, priority routing, and priority queuing. In this Section we discuss class-of-service identification and routing-table derivation.

3.2.1 Class-of-Service Identification

QoS resource management entails identifying class-of-service and class-of-service parameters, which may include, for example:

* service identity (SI),
* virtual network (VNET),
* link capability (LC), and
* QoS and traffic threshold parameters.

The SI describes the actual service associated with the call. The VNET describes the bandwidth allocation and routing table parameters to be used by the call. The LC describes the link hardware capabilities such as fiber, radio, satellite, and digital circuit multiplexing equipment (DCME), that the call should require, prefer, or avoid. The combination of SI, VNET, and LC constitute the class-of-service, which together with the network node number is used to access routing table data.

In addition to controlling bandwidth allocation, the QoS resource management procedures can check end-to-end transfer delay, delay variation, and transmission quality considerations such as loss, echo, and noise, as discussed in Section 3.7 below.

Ash              <draft-ietf-tewg-qos-routing-02.txt>       [Page ANNEX3-2]

Internet Draft  TE & QoS Methods for IP,ATM,TDM-Based Networks August 2001

Determination of class-of-service begins with translation at the originating node. The number or name is translated to determine the routing address of the destination node. If multiple ingress/egress routing is used, multiple destination node addresses are derived for the call. Other data derived from call information, such as link characteristics, Q.931 message information elements, Information Interchange digits, and network control point routing information, are used to derive the class-of-service for the call.

3.2.2 Policy-Based Routing Table Derivation

Class-of-service parameters are derived through application of policy-based routing. Policy-based routing involves the application of rules applied to input parameters to derive a routing table and its associated parameters. Input parameters for applying policy-based rules to derive SI, VNET, and LC could include numbering plan, type of origination/destination network, and type of service. Policy-based routing rules may then be applied to the
derived SI, VNET, and LC to derive the routing table and associated parameters.

Hence policy-based routing rules are used in SI derivation, which for example uses the type of origin, type of destination, signaling service type, and dialed number/name service type to derive the SI. The type of origin can be derived normally from the type of incoming link to the connected network domain, connecting either to a directly connected (also known as nodal) customer equipment location, a switched access local exchange carrier, or an international carrier location. Similarly, based on the dialed numbering plan, the type of destination network is derived and can be a directly connected (nodal) customer location if a private numbering plan is used (for example, within a virtual private network), a switched access customer location if a National Numbering Plan (NNP) number is used to the destination, or an international customer location if the international E.164 numbering plan is used. Signaling service type is derived based on bearer capability within signaling messages, information digits in dialed digit codes, numbering plan, or other signaling information and can indicate long-distance service (LDS), virtual private network (VPN) service, ISDN switched digital service (SDS), and other service types. Finally, dialed number service type is derived based on special dialed number codes such as 800 numbers or 900 numbers and can indicate 800 (FreePhone) service, 900 (Mass-Announcement) service, and other service types. Type of origin, type of destination, signaling service type, and dialed number service type are then all used to derive the SI.

The following are examples of the use of policy-based routing rules to derive class-of-service parameters. A long-distance service (LDS) SI, for example, is derived from the following information:

1. The type of origination network is a switched access local exchange carrier, because the call originates from a local exchange carrier node.
2. The type of destination network is a switched access local exchange carrier, based on the NNP dialed number.
3. The signaling service type is long-distance service, based on the numbering plan (NNP).
4. The dialed number service type is not used to distinguish long-distance service SI.

An 800 (FreePhone) service SI, for example, is derived from similar information, except that:

4. The dialed number service type is based on the 800 dialed "freephone" number to distinguish the 800 service SI.

A VPN service SI, for example, is derived from similar information, except that:

3. The signaling service type is based on the originating customer having access to VPN intelligent network (IN)-based services to derive the VPN service SI.

A service identity mapping table uses the above four inputs to derive the service identity. This policy-based routing table is changeable by administrative updates, in which new service information can be defined without software modifications to the node processing. From the SI and
bearer-service capability the SI/bearer-service-to-virtual network mapping table is used to derive the VNET.

Table 2.1 in ANNEX 2 illustrates the VNET mapping table. Here the SIs are mapped to individual virtual networks. Routing parameters for priority or key services are discussed further in the sections below.

Link capability selection allows calls to be routed on links that have the particular characteristics required by these calls. A call can require, prefer, or avoid a set of link characteristics such as fiber transmission, radio transmission, satellite transmission, or compressed voice transmission. Link capability requirements for the call can be determined by the SI of the call or by other information derived from the signaling message or from the routing number. The routing logic allows the call to skip those links that have undesired characteristics and to seek a best match for the requirements of the call.

3.2.3 QoS Resource Management Steps

The illustrative QoS resource management method consists of the following steps:

1. At the ON, the destination node (DN), SI, VNET, and QoS resource management information are determined through the number/name translation database and other service information available at the ON.
2. The DN and QoS resource management information are used to access the appropriate VNET and routing table between the ON and DN.
3. The connection request is set up over the first available path in the routing table with the required transmission resource selected based on the QoS resource management data.

In the first step, the ON translates the dialed digits to determine the address of the DN. If multiple ingress/egress routing is used, multiple destination node addresses are derived for the connection request. Other data derived from connection request information includes link characteristics, Q.931 message information elements, information interchange (II) digits, and service control point (SCP) routing information, and are used to derive the QoS resource management parameters (SI, VNET, LC, and QoS/traffic thresholds). SI describes the actual service associated with the connection request, VNET describes the bandwidth allocation and routing table parameters to be used by the connection request, and the LC describes the link characteristics including fiber, radio, satellite, and voice compression, that the connection request should require, prefer, or avoid. Each connection request is classified by its SI. A connection request for an individual service is allocated an equivalent bandwidth equal to EQBW and routed on a particular VNET. For CBR services the equivalent bandwidth EQBW is equal to the average or sustained bit rate. For VBR services the equivalent bandwidth EQBW is a function of the sustained bit rate, peak bit rate, and perhaps other parameters. For example, EQBW equals 64 kbps of bandwidth for CBR voice connections, 64 kbps of bandwidth for CBR ISDN switched digital 64-kbps connections, and 384-kbps of bandwidth for CBR ISDN switched digital 384-kbps connections.

In the second step, the SI value is used to derive the VNET. In the multi-service, QoS resource management network, bandwidth is allocated to individual VNETs which is protected as needed but otherwise shared. Under normal non-blocking/delay network conditions, all services fully share all available bandwidth. When blocking/delay occurs for VNET i, bandwidth reservation acts to prohibit alternate-routed traffic and traffic from other
VNETs from seizing the allocated capacity for VNET i. Associated with each VNET are average bandwidth ($BW_{avg}$) and maximum bandwidth ($BW_{max}$) parameters to govern bandwidth allocation and protection, which are discussed further in the next Section. As discussed, LC selection allows connection requests to be routed on specific transmission links that have the particular characteristics required by a connection request.

In the third step, the VNET routing table determines which network capacity is allowed to be selected for each connection request. In using the VNET routing table to select network capacity, the ON selects a first choice path based on the routing table selection rules. Whether or not bandwidth can be allocated to the connection request on the first choice path is determined by the QoS resource management rules given below. If a first choice path cannot be accessed, the ON may then try alternate paths determined by FR, TDR, SDR, or EDR path selection rules outlined in ANNEX 2. Whether or not bandwidth can be allocated to the connection request on the alternate path again is determined by the QoS resource management rules now described.

3.3 Dynamic Bandwidth Allocation, Protection, and Reservation Principles

QoS resource management functions include class-of-service identification, routing table derivation, connection admission, bandwidth allocation, bandwidth protection, bandwidth reservation, priority routing, and priority queuing. In this Section we discuss connection admission, bandwidth allocation, bandwidth protection, and bandwidth reservation.

This Section specifies the resource allocation controls and priority mechanisms, and the information needed to support them. In the illustrative QoS resource management method, the connection/bandwidth-allocation admission control for each link in the path is performed based on the status of the link. The ON may select any path for which the first link is allowed according to QoS resource management criteria. If a subsequent link is not allowed, then a release with crankback/bandwidth-not-available is used to return to the ON and select an alternate path. This use of an EDR path selection, which entails the use of the release with crankback/bandwidth-not-available mechanism to search for an available path, is an alternative to SDR path selection, which may entail flooding of frequently changing link state parameters such as available-cell-rate. The tradeoffs between EDR with crankback and SDR with link-state flooding are further discussed in ANNEX 6. In particular, when EDR path selection with crankback is used in lieu of SDR path selection with link-state flooding, the reduction in the frequency of such link-state parameter flooding allows for larger peer group sizes. This is because link-state flooding can consume substantial processor and link resources, in terms of message processing by the processors and link bandwidth consumed by messages on the links.

Two cases of QoS resource management are considered in this ANNEX: per-virtual-network (per-VNET) management and per-flow management. In the per-VNET method, such as illustrated for IP-based MPLS networks, aggregated LSP bandwidth is managed to meet the overall bandwidth requirements of VNET service needs. Individual flows are allocated bandwidth within the CRLSPs accordingly, as CRLSP bandwidth is available. In the per-flow method, bandwidth is allocated to each individual flow, such as in SVC set-up in an ATM-based network, from the overall pool of bandwidth, as the total pool bandwidth is available. A fundamental principle applied in these bandwidth allocation methods is the use of bandwidth reservation techniques. We first review bandwidth reservation principles and then discuss per-VNET and per-flow QoS resource allocation.
Bandwidth reservation (the TDM-network terminology is "trunk reservation") gives preference to the preferred traffic by allowing it to seize any idle bandwidth in a link, while allowing the non-preferred traffic to only seize bandwidth if there is a minimum level of idle bandwidth available, where the minimum-bandwidth threshold is called the reservation level. P. J. Burke [Bur61] first analyzed bandwidth reservation behavior from the solution of the birth–death equations for the bandwidth reservation model. Burke’s model showed the relative lost-traffic level for preferred traffic, which is not subject to bandwidth reservation restrictions, as compared to non-preferred traffic, which is subject to the restrictions. Figure 3.1 illustrates the percent lost traffic of preferred and non-preferred traffic on a typical link with 10 percent traffic overload. It is seen that the preferred traffic lost traffic is near zero, whereas the non-preferred lost traffic is much higher, and this situation is maintained across a wide variation in the percentage of the preferred traffic load. Hence, bandwidth reservation protection is robust to traffic variations and provides significant dynamic protection of particular streams of traffic.

Bandwidth reservation is a crucial technique used in nonhierarchical networks to prevent "instability," which can severely reduce throughput in periods of congestion, perhaps by as much as 50 percent of the traffic-carrying capacity of a network [E.525]. The phenomenon of instability has an interesting mathematical solution to network flow equations, which has been presented in several studies [NaM73, Kru82, Aki84]. It is shown in these studies that nonhierarchical networks exhibit two stable states, or bistability, under congestion and that networks can transition between these stable states in a network congestion condition that has been demonstrated in simulation studies. A simple explanation of how this bistable phenomenon arises is that under congestion, a network is often not able to complete a connection request on the primary shortest path, which consist in this example of a single link. If alternate routing is allowed, such as on longer, multiple-link paths, which are assumed in this example to consist of two links, then the connection request might be completed on a two-link path selected from among a large number of two-link path choices, only one of which needs sufficient idle bandwidth on both links to be used to route the connection. Because this two-link connection now occupies resources that could perhaps otherwise be used to complete two one-link connections, this is a less efficient use of network resources under congestion. In the event that a large fraction of all connections cannot complete on the direct link but instead occupy two-link paths, the total network throughput capacity is reduced by one-half because most connections take twice the resources needed. This is one stable state; that is, most or all connections use two links. The other stable state is that most or all connections use one link, which is the desired condition.

Bandwidth reservation is used to prevent this unstable behavior by having the preferred traffic on a link be the direct traffic on the primary, shortest path, and the non-preferred traffic, subjected to bandwidth reservation restrictions as described above, be the alternate-routed traffic on longer paths. In this way the alternate-routed traffic is inhibited from selecting longer alternate paths when sufficient idle trunk capacity is not available on all links of an alternate-routed connection, which is the likely condition under network and link congestion. Mathematically, the studies of bistable network behavior have shown that bandwidth reservation used in this manner to favor primary shortest connections eliminates the bistability problem in nonhierarchical networks and allows such networks to maintain efficient utilization under congestion by favoring connections completed on the shortest path. For this reason, dynamic bandwidth
reservation is universally applied in nonhierarchical TDM-based networks [E.529], and often in hierarchical networks [Mum76].

There are differences in how and when bandwidth reservation is applied, however, such as whether the bandwidth reservation for connections routed on the primary path is in place at all times or whether it is dynamically triggered to be used only under network or link congestion. This is a complex network throughput trade-off issue, because bandwidth reservation can lead to some loss in throughput under normal, low-congestion conditions. This loss in throughput arises because if bandwidth is reserved for connections on the primary path, but these connection requests do not arrive, then the capacity is needlessly reserved when it might be used to complete alternate-routed traffic that might otherwise be blocked. However, under network congestion, the use of bandwidth reservation is critical to preventing network instability, as explained above [E.525, E.529, E.731].

It is beneficial for bandwidth reservation techniques be included in IP-based and ATM-based routing methods, in order to ensure the efficient use of network resources especially under congestion conditions. Currently recommended path-selection methods, such as methods for optimized multipath for traffic engineering in IP-based MPLS networks [V99], or path selection in ATM-based PNNI networks [ATM96055], give no guidance on the necessity for using bandwidth-reservation techniques. Such guidance is essential for acceptable network performance [E.737].

Examples are given in this ANNEX for dynamically triggered bandwidth reservation techniques, where bandwidth reservation is triggered only under network congestion. Such methods are shown to be effective in striking a balance between protecting network resources under congestion and ensuring that resources are available for sharing when conditions permit. In Section 3.6 the phenomenon of network instability is illustrated through simulation studies, and the effectiveness of bandwidth reservation in eliminating the instability is demonstrated. Bandwidth reservation is also shown to be an effective technique to share bandwidth capacity among services integrated on a primary path, where the reservation in this case is invoked to prefer link capacity on the primary path for one particular class-of-service as opposed to another class-of-service when network and link congestion are encountered. These two aspects of bandwidth reservation, that is, for avoiding instability and for sharing bandwidth capacity among services, are illustrated in Section 3.4.
VNET CRLSP, and determines when VNET CRLSP bandwidth needs to be increased or decreased. Bandwidth changes in VNET bandwidth capacity are determined by ONs based on an overall aggregated bandwidth demand for VNET capacity (not on a per-connection demand basis). Based on the aggregated bandwidth demand, these ONs make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth on the CRLSPs constituting the VNET bandwidth capacity. For example, if connection requests are made for VNET CRLSP bandwidth that exceeds the current CRLSP bandwidth allocation, the ON initiates a bandwidth modification request on the appropriate CRLSP(s). For example, this bandwidth modification request may entail increasing the current CRLSP bandwidth allocation by a discrete increment of bandwidth denoted here as delta-bandwidth (DBW). DBW is a large enough bandwidth change so that modification requests are made relatively infrequently. Also, the ON periodically monitors CRLSP bandwidth use, such as once each minute, and if bandwidth use falls below the current CRLSP allocation the ON initiates a bandwidth modification request to decrease the CRLSP bandwidth allocation by a unit of bandwidth such as DBW.

In making a VNET bandwidth allocation modification, the ON determines the QoS resource management parameters including the VNET priority (key, normal, or best-effort), VNET bandwidth-in-use, VNET bandwidth allocation thresholds, and whether the CRLSP is a first choice CRLSP or alternate CRLSP. These parameters are used to access a VNET depth-of-search (DoS) table to determine a DoS load state threshold (Pi), or the "depth" to which network capacity can be allocated for the VNET bandwidth modification request. In using the DoS threshold to allocate VNET bandwidth capacity, the ON selects a first choice CRLSP based on the routing table selection rules.

Path selection in this IP network illustration may use open shortest path first (OSPF) for intra-domain routing. In OSPF-based layer 3 routing, as illustrated in Figure 3.3, ON A determines a list of shortest paths by using, for example, Dijkstra's algorithm.

This path list could be determined based on administrative weights of each link, which are communicated to all nodes within the autonomous system (AS) domain. These administrative weights may be set, for example, to \(1 + \epsilon \times \text{distance}\), where \(\epsilon\) is a factor giving a relatively smaller weight to the distance in comparison to the hop count. The ON selects a path from the list based on, for example, FR, TDR, SDR, or EDR path selection, as discussed in ANNEX 2.

For example, in using the first CRLSP A-B-E in Figure 3.3, ON A sends an MPLS label request message to VN B, which in turn forwards the label request message to DN E. VN B and DN E are passed in the explicit routing (ER) parameter contained in the label request message. Each node in the CRLSP reads the ER information, and passes the label request message to the next node listed in the ER parameter. If the first path is blocked at any of the links in the path, an MPLS notification message with a crankback parameter is returned to ON A, which can then attempt the next path. If FR is used, then this path is the next path in the shortest path list, for example path A-C-D-E. If TDR is used, then the next path is the next path in the routing table for the current time period. If SDR is used, OSPF implements a distributed method of flooding link status information, which is triggered either periodically and/or by crossing load state threshold values. This method of distributing link status information can be resource intensive and may not be any more efficient than simpler path selection methods such as EDR. If EDR is used, then the next path is the last successful path, and if that path is unsuccessful another alternate path is searched out according to the EDR path selection method.
Hence in using the selected CRLSP, the ON sends the explicit route, the requested traffic parameters (peak data rate, committed data rate, etc.), a DoS-parameter, and a modify-parameter in the MPLS label request message to each VN and the DN in the selected CRLSP. Whether or not bandwidth can be allocated to the bandwidth modification request on the first choice CRLSP is determined by each VN applying the QoS resource management rules. These rules entail that the VN determine the CRLSP link states, based on bandwidth use and bandwidth available, and compare the link load state to the DoS threshold Pi sent in the MPLS signaling parameters, as further explained below. If the first choice CRLSP cannot admit the bandwidth change, a VN or DN returns control to the ON through the use of the crankback-parameter in the MPLS notification message. At that point the ON may then try an alternate CRLSP. Whether or not bandwidth can be allocated to the bandwidth modification request on the alternate path again is determined by the use of the DoS threshold compared to the CRLSP link load state at each VN.

Priority queuing is used during the time the CRLSP is established, and at each link the queuing discipline is maintained such that the packets are given priority according to the VNET traffic priority.

Hence determination of the CRLSP link load states is necessary for QoS resource management to select network capacity on either the first choice CRLSP or alternate CRLSPs. Four link load states are distinguished: lightly loaded (LL), heavily loaded (HL), reserved (R), and busy (B). Management of CRLSP capacity uses the link state model and the DoS model to determine if a bandwidth modification request can be accepted on a given CRLSP. The allowed DoS load state threshold Pi determines if a bandwidth modification request can be accepted on a given link to an available bandwidth "depth." In setting up the bandwidth modification request, the ON encodes the DoS load state threshold allowed on each link in the DoS-parameter Pi, which is carried in the MPLS label request. If a CRLSP link is encountered at a VN in which the idle link bandwidth and link load state are below the allowed DoS load state threshold Pi, then the VN sends an MPLS notification message with the crankback-parameter to the ON, which can then route the bandwidth modification request to an alternate CRLSP choice. For example, in Figure 3.3, CRLSP A-B-E may be the first path tried where link A-B is in the LL state and link B-E is in the R state. If the DoS load state allowed is Pi=HL or better, then the CRLSP bandwidth modification request in the MPLS label request message is routed on link A-B but will not be admitted on link B-E, wherein the CRLSP bandwidth modification request will be cranked back in the MPLS notification message to the originating node A to try alternate CRLSP A-C-D-E. Here the CRLSP bandwidth modification request succeeds since all links have a state of HL or better.

3.4.1 Per-VNET Bandwidth Allocation/Reservation - Meshed Network Case

For purposes of bandwidth allocation reservation, two approaches are illustrated: one applicable to meshed network topologies and the other applicable to sparse topologies. In meshed networks, a greater number of logical links results in less traffic carried per link, and functions such as bandwidth reservation need to be more carefully controlled than in a sparse network. In a sparse network the traffic is concentrated on much larger, and many fewer logical links, and here bandwidth reservation does not have to be as carefully managed. Hence in the meshed network case, functions such as automatically triggering of bandwidth reservation on and off, dependent on the link/network congestion level, are beneficial to use. In the sparse network case, however, the complexity of such automatic triggering is not essential and bandwidth reservation may be permanently enabled without performance degradation.
Here we discuss a meshed network example of bandwidth allocation/reservation and in Section 3.4.2 we discuss the sparse network case.

The DoS load state threshold is a function of bandwidth-in-progress, VNET priority, and bandwidth allocation thresholds, as follows:

Table 3.1
Determination of Depth-of-Search (DoS) Load State Threshold
(Per-VNET Bandwidth Allocation, Meshed Network)

Note that BWIP, BWavg, and BWmax are specified per ON-DN pair, and that the QoS resource management method provides for a key priority VNET, a normal priority VNET, and a best effort VNET. Key services admitted by an ON on the key VNET are given higher priority routing treatment by allowing greater path selection DoS than normal services admitted on the normal VNET. Best effort services admitted on the best effort VNET are given lower priority routing treatment by allowing lesser path selection DoS than normal. 

The DoS load state threshold is determined by the following formula:

\[
\text{BWavgi}(w) = 0.5 \times \text{BWavgi}(w-1) + 0.5 \times [\text{BWIPavgi}(w) + \text{BWOVavgi}(w)]
\]

where all variables are specified per ON-DN pair, and where BWIPi and BWOVi are averaged across various load set periods, such as morning, afternoon, and evening averages for weekday, Saturday, and Sunday, to obtain BWIPavgi and BWOVavgi.

Table 3.2
Determination of Link Load State
(Meshed Network)

QoS resource management implements bandwidth reservation logic to favor connections routed on the first choice CRLSP in situations of link congestion. If link congestion (or blocking/delay) is detected, bandwidth reservation is immediately triggered and the reservation level N is set for the link according to the level of link congestion. In this manner bandwidth allocation requests attempting to alternate-path over a congested link are subject to bandwidth reservation, and the first choice CRLSP requests are favored for that link. At the same time, the LL and HL link state thresholds are raised accordingly in order to accommodate the reserved bandwidth capacity N for the VNET. Figure 3.4 illustrates bandwidth allocation and the mechanisms by which bandwidth is protected through bandwidth reservation. Under normal bandwidth allocation demands bandwidth is fully shared, but under overloaded bandwidth allocation demands, bandwidth is protected through the reservation mechanisms wherein each VNET can use its allocated bandwidth. Under failure, however, the reservation mechanisms operate to give the key VNET its allocated bandwidth before the normal priority VNET gets its bandwidth allocation. As noted on Table 3.1, the best effort low-priority VNET is not allocated bandwidth nor is bandwidth reserved for the best effort VNET. Further illustrations are given in Section 3.9 of the robustness of dynamic bandwidth reservation in protecting the preferred bandwidth requests across wide variations in
traffic conditions.

The reservation level $N$ (for example, $N$ may have 1 of 4 levels), is calculated for each link $k$ based on the link blocking/delay level of bandwidth allocation requests. The link blocking/delay level is equal to the total requested but rejected (or overflow) link bandwidth allocation (measured in total bandwidth), divided by the total requested link bandwidth allocation, over the last periodic update interval, which is, for example, every three minutes. That is

\[
\text{LBL}_k = \frac{\text{BWOW}_k}{\text{BWOF}_k}
\]

If $\text{LBL}_k$ exceeds a threshold value, the reservation level $N$ is calculated accordingly. The reserved bandwidth and link states are calculated based on the total link bandwidth required on link $k$, $\text{TRBW}_k$, which is computed on-line, for example every 1-minute interval $m$, and approximated as follows:

\[
\text{TRBW}_k(m) = 0.5 \times \text{TRBW}_k(m-1) + 0.5 \times \left[ 1.1 \times \text{TBWIP}_k(m) + \text{TBWOV}_k(m) \right]
\]

\[
\text{TBWIP}_k = \sum \text{BWIP}_i \text{ for all VNETs } i \text{ for bandwidth requests on their first choice CRLSP over link } k
\]

\[
\text{TBWOV}_k = \sum \text{BWOW}_i \text{ for all VNETs } i \text{ for bandwidth requests on their first choice CRLSP over link } k
\]

Therefore the reservation level and load state boundary thresholds are proportional to the estimated required bandwidth load, which means that the bandwidth reserved and the bandwidth required to constitute a lightly loaded link rise and fall with the bandwidth load, as, intuitively, they should.

\[
\begin{array}{l}
\text{BWOW}_k = \text{total requested bandwidth allocation rejected (or overflow) on link } k \\
\text{BWOF}_k = \text{total requested or offered bandwidth allocation on link } k \\
\text{LBL}_k = \text{link blocking/delay level on link } k = \frac{\text{BWOW}_k}{\text{BWOF}_k}
\end{array}
\]

3.4.2 Per-VNET Bandwidth Allocation/Reservation – Sparse Network Case

Here we discuss a sparse network example of bandwidth allocation/reservation. For the sparse network case of bandwidth reservation, a simpler method is illustrated which takes advantage of the concentration of traffic onto fewer, higher capacity backbone links. A small, fixed level of bandwidth reservation is used and permanently enabled on each link, as follows:

The DoS load state threshold again is a function of bandwidth-in-progress, VNET priority, and bandwidth allocation thresholds, however only the reserved (R) and non-reserved (NR) states are used, as follows:

\[
\text{Table 3.3}
\]

| Determination of Depth-of-Search (DoS) Load State Threshold (Per-VNET Bandwidth Allocation, Sparse Network) |

The corresponding load state table for the sparse network case is as follows:

\[
\text{Table 3.4}
\]
Determination of Link Load State (Sparse Network)

Note that reservation level is fixed and not dependent on any link blocking level (LBL) calculation or total required bandwidth (TRBW) calculation. Therefore LBL and TRBW monitoring are not required in this example bandwidth allocation/protection method.

3.5 Per-Flow Bandwidth Allocation, Protection, and Reservation

Per-flow QoS resource management methods have been applied successfully in TDM-based networks, where bandwidth allocation is determined by edge nodes based on bandwidth demand for each connection request. Based on the bandwidth demand, these edge nodes make changes in bandwidth allocation using for example an SVC-based QoS resource management approach illustrated in this Section. Again, the determination of the link load states is used for QoS resource management in order to select network capacity on either the first choice path or alternate paths. Also the allowed DoS load state threshold determines if an individual connection request can be admitted on a given link to an available bandwidth "depth." In setting up each connection request, the ON encodes the DoS load state threshold allowed on each link in the connection-setup IE. If a link is encountered at a VN in which the idle link bandwidth and link load state are below the allowed DoS load state threshold, then the VN sends a crankback/bandwidth-not-available IE to the ON, which can then route the connection request to an alternate path choice. For example, in Figure 3.3, path A-B-E may be the first path tried where link A-B is in the LL state and link B-E is in the R state. If the DoS load state allowed is HL or better, then the connection request is routed on link A-B but will not be admitted on link B-E, wherein the connection request will be cranked back to the originating node A to try alternate path A-C-D-E. Here the connection request succeeds since all links have a state of HL or better.

3.5.1 Per-Flow Bandwidth Allocation/Reservation - Meshed Network Case

Here again, two approaches are illustrated for bandwidth allocation/reservation: one applicable to meshed network topologies and the other applicable to sparse topologies. In a meshed network, a greater number of links results in less traffic carried per link, and functions such as bandwidth reservation need to be more carefully controlled than in a sparse network. In a sparse network the traffic is concentrated on much larger, and many fewer links, and here bandwidth reservation does not have to be as carefully management (such as automatically triggering bandwidth reservation on and off, dependent on the link/network congestion level).

Here we discuss a meshed network example of bandwidth allocation/reservation and in Section 3.5.2 we discuss the sparse network case.

The illustrative DoS load state threshold is a function of bandwidth-in-progress, service priority, and bandwidth allocation thresholds, as follows:

| Table 3.5 |
| Determination of Depth-of-Search (DoS) Load State Threshold |
| (Per-Flow Bandwidth Allocation, Meshed Network) |

Note that all parameters are specified per ON-DN pair, and that the QoS resource management method provides for key service and best effort service. Key services are given higher priority routing treatment by allowing greater path selection DoS than normal services. Best effort services are given
lower priority routing treatment by allowing lesser path selection DoS than normal. The quantities BWavgi are computed periodically, such as every week \( w \), and can be exponentially averaged over a several week period, as follows:

\[
\text{BWavgi}(w) = 0.5 \times \text{BWavgi}(w-1) + 0.5 \times [\text{BWIPavgi}(w) + \text{BWOVavgi}(w)]
\]

\[
\text{BWIPavgi} = \text{average bandwidth-in-progress across a load set period on VNET i}
\]

\[
\text{BWOVavgi} = \text{average bandwidth overflow across a load set period}
\]

where BWIPi and BVOVi are averaged across various load set periods, such as morning, afternoon, and evening averages for weekday, Saturday, and Sunday, to obtain BWIPavgi and BWOVavgi. Illustrative values of the thresholds to determine link load states are given in Table 3.2.

The illustrative QoS resource management method implements bandwidth reservation logic to favor connections routed on the first choice path in situations of link congestion. If link blocking/delay is detected, bandwidth reservation is immediately triggered and the reservation level \( N \) is set for the link according to the level of link congestion. In this manner traffic attempting to alternate-route over a congested link is subject to bandwidth reservation, and the first choice path traffic is favored for that link. At the same time, the LL and HL link state thresholds are raised accordingly in order to accommodate the reserved bandwidth capacity for the VNET. The reservation level \( N \) (for example, \( N \) may have 1 of 4 levels), is calculated for each link \( k \) based on the link blocking/delay level and the estimated link traffic. The link blocking/delay level is equal to the equivalent bandwidth overflow count divided by the equivalent bandwidth peg count over the last periodic update interval, which is typically three minutes. That is

\[
\text{BWOVk} = \frac{\text{equivalent bandwidth overflow count on link } k}{\text{equivalent bandwidth peg count on link } k}
\]

\[
\text{LBLk} = \frac{\text{link blocking/delay level on link } k}{\text{BWOVk}/\text{BWPCk}}
\]

Therefore the reservation level and load state boundary thresholds are proportional to the estimated required bandwidth traffic load, which means that the bandwidth reserved and the bandwidth required to constitute a lightly loaded link rise and fall with the traffic load, as, intuitively, they should.
3.5.2 Per-Flow Bandwidth Allocation/Reservation - Sparse Network Case

Here we discuss a sparse network example of bandwidth allocation/reservation. For the sparse network case of bandwidth reservation, a simpler method is illustrated which takes advantage of the concentration of traffic onto fewer, higher capacity backbone links. A small, fixed level of bandwidth reservation is used on each link, as follows:

The DoS load state threshold again is a function of bandwidth-in-progress, VNET priority, and bandwidth allocation thresholds, however only the reserved (R) and non-reserved (NR) states are used, as follows:

Table 3.6
Determination of Depth-of-Search (DoS) Load State Threshold
(Per-Flow Bandwidth Allocation, Sparse Network)

The corresponding load state table for the sparse network case is as follows:

Table 3.7
Determination of Link Load State (Sparse Network)

Note that reservation level is fixed and not dependent on any link blocking level (LBL) calculation or total required bandwidth (TRBW) calculation. Therefore LBL and TRBW monitoring are not required in this example.

3.6 Packet-Level Traffic Control

QoS resource management functions include class-of-service identification, routing table derivation, connection admission, bandwidth allocation, bandwidth protection, bandwidth reservation, priority routing, traffic shaping, packet priority, and queue management. Packet level traffic control encompasses the control procedures which allow packet level GoS objectives to be fulfilled [E.736]. Once a flow is admitted through the connection admission control (CAC) functions, packet level control a) ensures through traffic shaping that the user in fact emits traffic in conformity with the declared traffic parameters, b) ensures through packet priority and queue management that the network provides the requested quality of service in conformity with the declared traffic and allocated resources, c) defines performance evaluation methods and traffic control methods enabling a network to meet objectives for packet level network performance, and d) considers control actions such as usage parameter control to achieve a high level of consistency between the various control capabilities. Traffic controls may be distinguished according to whether their function is to enable quality of service guarantees at packet level (e.g. packet loss ratio) or at connection level (e.g. connection blocking probability).

As discussed in Sections 6 - 9, CAC determines if a link or path connection is capable or not of handling the requested connection. In a connection-oriented network, each connection is defined by a set of traffic parameters and QoS requirements. When the establishment of a new connection is requested, the network must decide if it has sufficient resources to accept it without infringing packet level GoS requirements for all established connections as well as the new connection. This decision can sometimes be made by allocating resources to specific connections or groups of connections and refusing new requests when insufficient resources are available. Note that the allocation is generally logical: no particular physical resources are attributed to a specific connection. The resources in
question are typically bandwidth and buffer space. It is assumed that resources are allocated independently for each link or path connection with a separate decision made for each transmission direction of a connection. A connection will be established only if resources are available on every link of its path, in both directions. Admission control could be applied for peak rate allocation wherein a network operator may choose to apply an overbooking factor. It is possible to base admission control on an equivalent packet rate: connection i is attributed an equivalent packet rate epri and connections are admitted while $S \cdot epri < c$, where $c$ is the link rate. If the only performance requirement consists in guaranteeing a minimum packet rate, epri may be set equal to this minimum rate: $epri = mpri$.

A connection request is specified by a traffic descriptor, delay variation tolerance and QoS requirements. The source traffic descriptor is a list of traffic parameters which should a) be understandable and conformance should be possible, b) be used in resource allocation meeting network performance requirements, and c) be enforceable by the usage parameter control and network parameter control. The traffic parameters may relate explicitly to connection traffic characteristics such as the peak data rate or implicitly define these characteristics by reference to a service type. End-to-end packet level QoS criteria use the following performance parameters: a) transfer delay; b) delay variation; c) loss ratio. End-to-end performance objectives relevant to traffic engineering are as follows: a) maximum end-to-end queuing delay; b) mean queuing delay; c) packet loss ratio. These performance objectives must be apportioned to the various network elements contributing to the performance degradation of a given connection so that the end-to-end QoS criteria are satisfied.

If the connection is accepted, there is implicitly defined a traffic contract whereby the network operator provides the requested quality of service on condition that the user emits traffic in conformity with the declared traffic descriptor; this is the role of usage parameter control. When more than one network is involved in a connection, it is also incumbent on each network to verify that the traffic it receives from the neighboring network conforms; this is network parameter control. One of the requirements on traffic parameters is that they be enforceable by the usage parameter control and network parameter control. This has led to a definition of traffic parameters: peak packet rate, sustainable packet rate and intrinsic burst tolerance allowing user conformance to be determined by the generic packet rate algorithm. In this method, supplementary packet delays may be introduced to shape the characteristics of a given flow. By smoothing packet rate variations, shaping generally allows an increase in the utilization of network resources leading to greater multiplexing gains. On the other hand, shaping may introduce non-negligible delays and a part of the end-to-end GoS objective must be allocated to the shaper. Shaping may be performed by the user to ensure compliance with declared traffic parameters and delay variation tolerance. The network operator may employ shaping at the network entrance, within the network or at the network egress (to meet constraints on output traffic characteristics). Shaping is an option for users and networks. A particular example of shaping is the reduction of delay variation by means of packet spacing. The spacer tries to produce a packet stream with a time between consecutive packets at least equal to the inverse of the peak data rate by imposing a variable delay on each packet.

We now discuss priority queuing as an illustrative traffic scheduling method, and further assume that a traffic shaper function is employed as discussed above, such as a leaky-bucket model, to determine out-of-contract...
traffic behavior and appropriately mark packets for possible dropping under congestion. These scheduling and shaping mechanisms compliment the connection admission mechanisms described in the previous Sections to appropriately allocate bandwidth on links in the network.

Note that priority queuing is used as an illustrative scheduling mechanism, whereas other methods may be used. DiffServ does not require that a particular queuing mechanism be used to achieve EF, AF, etc. QoS. Therefore the queuing implementation used for DiffServ could be weighted fair queuing (WFQ), priority queuing (PQ), or other queuing mechanism, depending on the choice in the implementation. In the analysis PQ is used for illustration, however the same or comparable results would be obtained with WFQ or other queuing mechanisms.

In addition to the QoS bandwidth management procedure for bandwidth allocation requests, a QoS priority of service queuing capability is used during the time connections are established on each of the three VNETs. At each link, a queuing discipline is maintained such that the packets being served are given priority in the following order: key VNET services, normal VNET services, and best effort VNET services. Following the MPLS CR-LSP bandwidth allocation setup and the application of QoS resource management rules, the priority of service parameter and label parameter need to be sent in each IP packet, as illustrated in Figure 3.5. The priority of service parameter may be included in the type of service (ToS), or differentiated services (DiffServ) [RFC2475, LDVKCH00, ST98], parameter already in the IP packet header. Another possible alternative is that the priority of service parameter can be associated with the MPLS label appended to the IP packet [LDVKCH00]. In either case, from the priority of service parameters, the IP node can determine the QoS treatment based on the QoS resource management (priority queuing) rules for key VNET packets, normal VNET packets, and best effort VNET packets. From the label parameter, the IP node can determine the next node to route the IP packet to as defined by the MPLS protocol. In this way, the backbone nodes can have a very simple per-packet processing implementation to implement QoS resource management and MPLS routing.

3.7 Other QoS Resource Management Constraints

Other QoS routing constraints are taken into account in the QoS resource management and route selection methods in addition to bandwidth allocation, bandwidth protection, and priority routing. These include end-to-end transfer delay, delay variation [G99a], and transmission quality considerations such as loss, echo, and noise [D99, G99a, G99b]. Additionally, link capability (LC) selection allows connection requests to be routed on specific transmission media that have the particular characteristics required by these connection requests. In general, a connection request can require, prefer, or avoid a set of transmission characteristics such as fiber optic or radio transmission, satellite or terrestrial transmission, or compressed or uncompressed transmission. The routing table logic allows the connection request to skip links that have undesired characteristics and to seek a best match for the requirements of the connection request. For any SI, a set of LC selection preferences is specified for the connection request. LC selection preferences can override the normal order of selection of paths. If a LC characteristic is required, then any path with a link that does not have that characteristic is skipped. If a characteristic is preferred, paths having all links with that characteristic are used first. Paths having links without the preferred characteristic will be used next. A LC preference is set for the presence or absence of a characteristic. For example, if fiberoptic transmission is required, then only paths with links having Fiberoptic=Yes are used. If we
prefer the presence of fiberoptic transmission, then paths having all links with Fiberoptic=Yes are used first, then paths having some links with Fiberoptic=No.

3.8 Interdomain QoS Resource Management

In current practice, interdomain routing protocols generally do not incorporate standardized path selection or per class-of-service QoS resource management. For example, in IP-based networks BGP [RL00] is used for interdomain routing but does not incorporate per class-of-service resource allocation as described in this Section. Also, MPLS techniques have not yet been addressed for interdomain applications. Extensions to interdomain routing methods discussed in this Section therefore can be considered to extend the call routing and connection routing concepts to routing between network domains.

Interdomain routing can also apply class-of-service routing concepts described in Section 3.2 and increased routing flexibility for interdomain routing. Principles discussed in Section 3.2 for class-of-service derivation and policy-based routing table derivation also apply in the case of interdomain QoS resource management. As described in ANNEX 2, interdomain routing works synergistically with multiple ingress/egress routing and alternate routing through transit domains. Interdomain routing can use link status information in combination with call completion history to select paths and also use dynamic bandwidth reservation techniques, as discussed in Sections 3.3 to 3.7.

Interdomain routing can use the virtual network concept that enables service integration by allocating bandwidth for services and using dynamic bandwidth reservation controls. These virtual network concepts have been described in this ANNEX, and can be extended directly to interdomain routing. For example, the links connected to the originating domain gateway nodes, such as links OCN1-DGN1, OGN2-DGN1, OGN1-VGN1, OGN1-VGN2, and OGN2-VGN2 in Figure 2.5, can define VNET bandwidth allocation, protection, reservation, and routing methods, exactly as described in Sections 3.3 to 3.7. In that way, bandwidth can be fully shared among virtual networks in the absence of congestion. When a certain virtual network encounters congestion, bandwidth is reserved to ensure that the virtual network reaches its allocated bandwidth. Interdomain routing can employ class-of-service routing capabilities including key service protection, directional flow control, link selection capability, automatically updated time-variable bandwidth allocation, and alternate routing capability through the use of overflow paths and control parameters such as interdomain routing load set periods. Link selection capability allows specific link characteristics, such as fiber transmission, to be preferentially selected. Thereby interdomain routing can improve performance and reduce the cost of the interdomain network with flexible routing capabilities, such as described in ANNEX 2 (Section 2.8).

Similar to intradomain routing, interdomain routing may include the following steps for call establishment:

1. At the originating gateway node (OGN), the destination gateway node (DGN), SI, VNET, and QoS resource management information are determined through the number/name translation database and other service information available at the OGN.
2. The DGN and QoS resource management information are used to access the appropriate VNET and routing table between the OGN and DGN.
3. The connection request is set up over the first available path in
the routing table with the required transmission resource selected based on
the QoS resource management data.

The rules for selecting the interdomain primary path and alternate paths for
a call can be governed by the availability of primary path bandwidth,
ode-to-node congestion, and link capability, as described in Sections 3.3
to 3.7. The path sequence consists of the primary shortest path, lightly
loaded alternate paths, heavily loaded alternate paths, and reserved
alternate paths, where these load states are further refined by combining
link load state information with path congestion state information, as
described in Section 2.7. Interdomain alternate paths which include nodes in
the originating domain and terminating domain are selected before alternate
paths which include transit domain nodes are selected. As described in
Sections 3.4 and 3.5, greater path selection depth is allowed if congestion
is detected to the destination network domain, because more alternate path
choices serve to reduce the congestion. During periods of no congestion,
capacity not needed by one virtual network is made available to other
virtual networks that are experiencing loads above their allocation.

The gateway node, for example, may automatically compute the bandwidth
allocations once a week and may use a different allocation for various load
set periods, for example each of 36 two-hour load set periods: 12 weekday,
12 Saturday, and 12 Sunday. The allocation of the bandwidth can be based on
a rolling average of the traffic load for each of the virtual networks, to
each destination node, in each of the load set periods. Under normal network
conditions in which there is no congestion, all virtual networks fully share
all available capacity. Under call overload, however, link bandwidth is
reserved to ensure that each virtual network gets the amount of bandwidth
allotted. This dynamic bandwidth reservation during times of overload
results in network performance that is analogous to having the link
bandwidth allocation between the two nodes dedicated for each VNET.

3.9 Modeling of Traffic Engineering Methods

In this Section, we again use the full-scale national network model
developed in ANNEX 2 to study various TE scenarios and tradeoffs. The
135-node national model is illustrated in Figure 2.6, the multiservice
traffic demand model is summarized in Table 2.1, and the cost model is
summarized in Table 2.2.

3.9.1 Performance of Bandwidth Reservation Methods

As discussed in Section 3.3, dynamic bandwidth reservation can be used to
favor one category of traffic over another category of traffic. A simple
example of the use of this method is to reserve bandwidth in order to prefer
traffic on the shorter primary paths over traffic using longer alternate
paths. This is most efficiently done by using a method which reserves
bandwidth only when congestion exists on links in the network. We now give
illustrations of this method, and compare the performance of a network in
which bandwidth reservation is used under congestion to the case when
bandwidth reservation is not used.

In the example, traffic is first routed on the shortest path, and then
allowed to alternate route on longer paths if the primary path is not
available. In the case where bandwidth reservation is used, five percent of
the link bandwidth is reserved for traffic on the primary path when
congestion is present on the link.

Table 3.4 illustrates the performance of bandwidth reservation methods for a
high-day network load pattern. This is the case for multilink path routing being used in to set up per-flow CRLSPs in a sparse network topology.

Table 3.4
Performance of Dynamic Bandwidth Reservation Methods for CRLSP Setup
Percent Lost/Delayed Traffic under Overload
(Per-Flow Multilink Path Routing in Sparse Network Topology; 135-Node Multiservice Network Model)

We can see from the results of Table 3.4 that performance improves when bandwidth reservation is used. The reason for the poor performance without bandwidth reservation is due to the lack of reserved capacity to favor traffic routed on the more direct primary paths under network congestion conditions. Without bandwidth reservation nonhierarchical networks can exhibit unstable behavior in which essentially all connections are established on longer alternate paths as opposed to shorter primary paths, which greatly reduces network throughput and increases network congestion [Aki84, Kru82, NaM73]. If we add the bandwidth reservation mechanism, then performance of the network is greatly improved.

Another example is given in Table 3.5, where 2-link SDR is used in a meshed network topology. In this case, the average business day loads for a 65-node national network model were inflated uniformly by 30 percent [A98]. The Table gives the average hourly lost traffic due to blocking of connection admissions in hours 2, 3, and 5, which correspond to the two early morning busy hours and the afternoon busy hour.

Table 3.5
Performance of Dynamic Bandwidth Reservation Methods
Percent Lost Traffic under 30% Overload
(Per-Flow 2-link SDR in Meshed Network Topology; 65-Node Network Model)

Again, we can see from the results of Table 3.5 that performance dramatically improves when bandwidth reservation is used. A clear instability arises when bandwidth reservation is not used, because under congestion a network state in which virtually all traffic occupies 2 links instead of 1 link is predominant. When bandwidth reservation is used, flows are much more likely to be routed on a 1-link path, because the bandwidth reservation mechanism makes it less likely that a 2-link path can be found in which both links have idle capacity in excess of the reservation level.

A performance comparison is given in Table 3.6 for a single link failure in a 135-node design averaged over 5 network busy hours, for the case without bandwidth reservation and with bandwidth reservation. Clearly the use of bandwidth reservation protects the performance of each virtual network class-of-service category.

Table 3.6
Performance of Dynamic Bandwidth Reservation Methods
Percent Lost/Delayed Traffic under 50% General Overload
(Multilink STT-EDR; 135-Node Network Model)

3.9.2 Multiservice Network Performance: Per-VNET vs. Per-Flow Bandwidth Allocation

Here we use the 135-node model to compare the per-virtual-network methods of QoS resource, as described in Section 3.3.2, and the per-flow methods described in Section 3.3.3. We look at these two cases in Figure 3.6, which illustrates the case of per-virtual-network CRLSP bandwidth allocation the
case of per-flow CRLSP bandwidth allocation. The two figures compare the performance in terms of lost or delayed traffic under a focused overload scenario on the Oakbrook (OKBK), IL node (such as might occur, for example, with a radio call-in give-away offer). The size of the focused overload is varied from the normal load (1X case) to a 10 times overload of the traffic to OKBK (10X case). Here a fixed routing (FR) CRLSP bandwidth allocation is used for both the per-flow CRLSP bandwidth allocation case and the per-virtual-network bandwidth allocation case. The results show that the per-flow and per-virtual-network bandwidth allocation performance is similar; however, the improved performance of the key priority traffic and normal priority traffic in relation to the best-effort priority traffic is clearly evident.

The performance analyses for overloads and failures for the per-flow and per-virtual-network bandwidth allocation are now examined in which event dependent routing (EDR) with success-to-the-top (STT) path selection are used. Again the simulations include call admission control with QoS resource management, in which we distinguish the key services, normal services, and best-effort services as indicated in the tables below. Table 3.7 gives performance results for a 30% general overload, Table 3.8 gives performance results for a six-times overload on a single network node, and Table 3.9 gives performance results for a single transport link failure. Performance analysis results show that the multilink STT-EDR per-flow bandwidth allocation and per-virtual-network bandwidth allocation options perform similarly under overloads and failures.

Table 3.7
Performance of Per-Flow & Per-Virtual-Network Bandwidth Allocation
Percent Lost/Delayed Traffic under 30% General Overload
(Single-Area Flat Network Topology; Multilink STT-EDR Routing; 135-Node Network Model)

Table 3.8
Performance of Per-Flow & Per-Virtual-Network Bandwidth Allocation
Percent Lost/Delayed Traffic under 6X Focused Overload on OKBK
(Single-Area Flat Network Topology; Multilink STT-EDR Routing; 135-Node Network Model)

Table 3.9
Performance of Per-Flow & Per-Virtual-Network Bandwidth Allocation
Percent Lost/Delayed Traffic under Failure on CHCG-NYCM Link
(Single-Area Flat Network Topology; Multilink STT-EDR Routing; 135-Node Network Model)

3.9.3 Multiservice Network Performance: Single-Area Flat Topology vs. Multi-Area 2-Level Hierarchical Flat Topology

We also investigate the performance of hierarchical network designs, which represent the topological configuration to be expected with multi-area (or multi-autonomous-system (multi-AS), or multi-domain) networks. In Figure 2.10 we show the model considered, which consists of 135 edge nodes each homed onto one of 21 backbone nodes. Typically, the edge nodes may be grouped into separate areas or autonomous systems, and the backbone nodes into another area or autonomous system. Within each area a flat routing topology exists, however between edge areas and the backbone area a hierarchical routing relationship exists. This routing hierarchy is modeled for both the per-flow and per-virtual-network bandwidth allocation examples, and the results are given in Tables 3.10 to 3.12 for the 30% general
overload, 6-times focused overload, and link failure examples, respectively. We can see that the performance of the hierarchical network case is substantially worse than the flat network model, which models a single area or autonomous system consisting of 135 nodes.

Table 3.10
Performance of Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under 30% General Overload
Per-Flow & Per-Virtual-Network Bandwidth Allocation
(Multilink STT-EDR Routing; 135-Node Network Model)

Table 3.11
Performance of Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under 6X Focused Overload on OKBK
Per-Flow & Per-Virtual-Network Bandwidth Allocation
(Multilink STT-EDR Routing; 135-Node Network Model)

Table 3.12
Performance of Multi-Area 2-Level Hierarchical Network Topology
Percent Lost/Delayed Traffic under Failure on CHCG-NYCM Link
Per-Flow & Per-Virtual-Network Bandwidth Allocation
(Multilink STT-EDR Routing; 135-Node Network Model)

3.9.4 Multiservice Network Performance: Need for MPLS & DiffServ

We illustrate the operation of MPLS and DiffServ in the multiservice network model with some examples. First suppose there is 10 mbps of normal-priority traffic and 10 mbps of best-effort priority traffic being carried in the network between node A and node B. Best-effort traffic is treated in the model as UBR traffic and is not allocated any bandwidth. Hence the best-effort traffic does not get any CRLSP bandwidth allocation, and is not treated as MPLS forward equivalence class (FEC) traffic. As such, the best-effort traffic would be routed by the interior gateway protocol, or IGP, such as OSPF. Hence the best-effort traffic cannot be denied bandwidth allocation as a means to throttle back such traffic at the edge router, which can be done with the normal-priority and key-priority traffic (i.e., normal and key traffic could be denied bandwidth allocation). The only way that the best-effort traffic gets dropped/lost is to drop it at the queues, therefore it is essential that the traffic that is allocated bandwidth on the CRLSPs have higher priority at the queues than the best-effort traffic. Therefore in the model the three classes of traffic get these DiffServ markings: best-effort get no-DiffServ marking, which ensures that it will get best-effort priority queuing treatment. Normal-priority traffic gets the assured forwarding (AF) DiffServ marking, which is a middle priority level of queuing treatment, and key-priority traffic gets the expedited forwarding (EF) DiffServ marking, which is the highest priority queuing level.

Now suppose that there is 30 mbps of bandwidth available between A and B and that all the normal-priority and best-effort traffic is getting through. Now suppose that the traffic for both the normal-priority and best-effort traffic increases to 20 mbps. The normal-priority traffic requests and gets a CRLSP bandwidth allocation increase to 20 mbps on the A to B CRLSP. However, the best-effort traffic, since it has no CRLSP assigned and therefore no bandwidth allocation, is just sent into the network at 20 mbps. Since there is only 30 mbps of bandwidth available from A to B, the network must drop 10 mbps of best-effort traffic in order to leave room for the 20 mbps of normal-priority traffic. The way this is done in the model is through the queuing mechanisms governed by the DiffServ priority settings on
each category of traffic. Through the DiffServ marking, the queuing mechanisms in the model discard about 10 mbps of the best-effort traffic at the priority queues. If the DiffServ markings were not used, then the normal-priority and best-effort traffic would compete equally on the first-in/first-out (FIFO) queues, and perhaps 15 mbps of each would get through, which is not the desired situation.

Taking this example further, if the normal-priority and best-effort traffic both increase to 40 mbps, then the normal-priority traffic tries to get a CRLSP bandwidth allocation increase to 40 mbps. However, the most it can get is 30 mbps, so 10 mbps is denied for the normal-priority traffic in the MPLS constraint-based routing procedure. By having the DiffServ markings of AF on the normal-priority traffic and none on the best-effort traffic, essentially all the best-effort traffic is dropped at the queues since the normal-priority traffic is allocated and gets the full 30 mbps of A to B bandwidth. If there were no DiffServ markings, then again perhaps 15 mbps of both normal-priority and best-effort get through. Or in this case, perhaps a greater amount of best-effort traffic is carried than normal-priority traffic, since 40 mbps of best-effort traffic is sent into the network and only 30 mbps of normal-priority traffic is sent into the network, and the FIFO queues will receive more best-effort pressure than normal-priority pressure.

Some general observations on the operation of MPLS and DiffServ in the multiservice TE models include the following:

1. In a multiservice network environment, with best-effort priority traffic (WWW traffic, email, ..), normal-priority traffic (CBR voice, IP-telephony voice, switched digital service, ..), and key-priority traffic (800-gold, incoming international, ..) sharing the same network, MPLS bandwidth allocation plus DiffServ/priority-queuing are both needed. In the models the normal-priority and key-priority traffic use MPLS to receive bandwidth allocation while the best-effort traffic gets no bandwidth allocation. Under congestion (e.g., from overloads or failures), the DiffServ/priority-queuing mechanisms push out the best-effort priority traffic at the queues so that the normal-priority and key-priority traffic can get through on the MPLS-allocated CRLSP bandwidth.

2. In a multiservice network where the normal-priority and key-priority traffic use MPLS to receive bandwidth allocation and there is no best-effort priority traffic, then the DiffServ/priority queuing becomes less important. This is because the MPLS bandwidth allocation more-or-less assures that the queues will not overflow, and perhaps therefore DiffServ would not be needed as much.

3. As bandwidth gets more and more plentiful/lower-cost, the point at which the MPLS and DiffServ mechanisms have a significant effect under traffic overload goes to a higher and higher threshold. For example, the models show that the overload factor at which congestion occurs gets larger as the bandwidth modules get larger (i.e., OC3 to OC12 to OC48 to OC192, etc.). However, the congestion point will always be reached with failures and/or large-enough overloads necessitating the MPLS/DiffServ mechanisms.

3.10 Conclusions/Recommendations

The conclusions/recommendations reached in this ANNEX are as follows:

* QoS resource management is recommended and is shown to be effective in achieving connection-level and packet-level GoS objectives, as well as
key service, normal service, and best effort service differentiation.

* Admission control is recommended and is the basis that allows for applying most of the other controls described in this draft.

* Bandwidth reservation is recommended and is critical to the stable and efficient performance of TE methods in a network, and to ensure the proper operation of multiservice bandwidth allocation, protection, and priority treatment.

* Per-VNET bandwidth allocation is recommended and is essentially equivalent to per-flow bandwidth allocation in network performance and efficiency. Because of the much lower routing table management overhead requirements, as discussed and modeled in ANNEX 4, per-VNET bandwidth allocation is preferred to per-flow allocation.

* Both MPLS QoS and bandwidth management and DiffServ priority queuing management are recommended and are important for ensuring that multiservice network performance objectives are met under a range of network conditions. Both mechanisms operate together to ensure QoS resource allocation mechanisms (bandwidth allocation, protection, and priority queuing) are achieved.

Figures:

Figure 3.1 Dynamic Bandwidth Reservation Performance under 10% Overload
Figure 3.2 Virtual Network (VNET) Bandwidth Management
Figure 3.3 Label Switched Path Selection for Bandwidth Modification Request
Figure 3.4 Bandwidth Allocation, Protection, and Priority Routing
Figure 3.5 IP Packet Structure under MPLS Switching
Figure 3.6 Performance under Focused Overload on OKBR Node

4.1 Introduction

Routing table management typically entails the automatic generation of routing tables based on network topology and other information such as status. Routing table management information, such as topology update, status information, or routing recommendations, is used for purposes of applying the routing table design rules for determining path choices in the routing table. This information is exchanged between one node and another node, such as between the originating node (ON) and destination node (DN), for example, or between a node and a network element such as a bandwidth-broker processor (BBP). This information is used to generate the routing table, and then the routing table is used to determine the path choices used in the selection of a path.

This automatic generation function is enabled by the automatic exchange of link, node, and reachable address information among the network nodes. In order to achieve automatic update and synchronization of the topology database, which is essential for routing table management, IP- and ATM-based networks already interpret HELLO protocol mechanisms to identify links in the network. For topology database synchronization the link state advertisement (LSA) is used in IP-based networks, and the PNNI
topology-state-element (PTSE) exchange is used in ATM-based networks, to automatically provision nodes, links, and reachable addresses in the topology database. Use of a single peer group/autonomous system for topology update leads to more efficient routing and easier administration, and is best achieved by minimizing the use of topology state (LSA and PTSE) flooding for dynamic topology state information. It is required in Section 4.5 that a topology state element (TSE) be developed within TDM-based networks. When this is the case, then the HELLO and LSA/TSE/PTSE parameters will become the standard topology update method for interworking across IP-, ATM-, and TDM-based networks.

Status update methods are required for use in routing table management within and between network types. In TDM-based networks, status updates of link and/or node status are used [E.350, E.351]. Within IP- and ATM-based networks, status updates are provided by a flooding mechanism. It is required in Section 4.5 that a routing status element (RSE) be developed within TDM-based networks, which will be compatible with the PNNI topology state element (PTSE) in ATM-based networks and the link state advertisement (LSA) element in IP-based networks. When this is the case, then the RSE/PTSE/LSA parameters will become the standard status update method for interworking across TDM-, ATM-, and IP-based networks.

Routing recommendation methods are proposed for use in routing table management within and between network types. For example, such methods provide for a database, such as a BGP, to advertise recommended paths to network nodes based on status information available in the database. Such routing recommendation methods are provided in TDM-based networks [E.350, E.351]. It is required in Section 4.5 that a routing recommendation element (RRE) be developed within ATM-based and IP-based networks. When this is the case, then the RRE parameters will become the standard query for status method for interworking across TDM-, ATM-, and IP-based networks.

4.2 Routing Table Management for IP-Based Networks

IP networks typically run the OSPF protocol for intra-domain routing [RFC2328, S95] and the BGP protocol for inter-domain routing [RL00, S95]. OSPF and BGP are designed for routing of datagram packets carrying multimedia internet traffic. Within OSPF, a link-state update topology exchange mechanism is used by each IP node to construct its own shortest path routing tables. Through use of these routing tables, the IP nodes match the destination IP address to the longest match in the table and thereby determine the shortest path to the destination for each IP packet. In current OSPF operation, this shortest path remains fixed unless a link is added or removed (e.g., fails), and/or an IP node enters or leaves the network. However the protocol allows for possibly more sophisticated dynamic routing mechanisms to be implemented. MPLS is currently being developed as a means by which IP networks may provide connection oriented, QoS-routing services, such as with ATM layer-2 switching technology [RVC99], and differentiated services (DiffServ) [RFC2475, ST98] is being developed to provide priority queuing control in IP-based networks. MPLS and DiffServ both provide essential capabilities for QoS resource management, as
These IP-based protocols provide for a) exchange of node and link status information, b) automatic update and synchronization of topology databases, and c) fixed and/or dynamic route selection based on topology and status information. For topology database synchronization, each node in an IP-based OSPF/BGP network exchanges HELLO packets with its immediate neighbors and thereby determines its local state information. This state information includes the identity and group membership of the node’s immediate neighbors, and the status of its links to the neighbors. Each node then bundles its state information in LSAs, which are reliably flooded throughout the autonomous system (AS), or group of nodes exchanging routing information and using a common routing protocol, which is analogous to the PNNI peer group used in ATM-based networks. The LSAs are used to flood node information, link state information, and reachability information. As in PNNI, some of the topology state information is static and some is dynamic. In order to allow larger AS group sizes, a network can use OSPF in such a way so as to minimize the amount of dynamic topology state information flooding, such as available link bandwidth, by setting thresholds to values that inhibit frequent updates.

IP-based routing of connection/bandwidth-allocation requests and QoS-routing support are in the process of standardization primarily within the MPLS and DiffServ activities in the IETF. IGPs such as OSPF are still applicable to determine routing in an MPLS architecture, but are only one of many proposed capabilities to implement traffic engineering (TE) with MPLS. The TE framework document [ACEWX00] calls for many TE mechanisms -- distributed, centralized, off-line, on-line, time-dependent, state-dependent, event-dependent, some but not all of which would involve interior gateway protocols such as OSPF.

As described in the TE framework document, a number of enhancements are needed to conventional link state IGPs, such as OSPF and IS-IS, to allow them to distribute additional state information required for constraint-based routing. Essentially, these enhancements require the propagation of additional information in link state advertisements. Specifically, in addition to normal link-state information, an enhanced IGP is required to propagate topology state information needed for constraint-based routing. Some of the additional topology state information include link attributes such as reservable bandwidth and link resource class attribute (an administratively specified property of the link). Deployment of MPLS for traffic engineering applications has commenced in some service provider networks. One operational scenario is to deploy MPLS in conjunction with an IGP (IS-IS-TE or OSPF-TE) that supports the traffic engineering extensions, in conjunction with constraint-based routing for explicit route computations, and a signaling protocol (e.g. RSVP-TE or CRFPD) for LSP instantiation.

In contemporary MPLS traffic engineering contexts, network administrators specify and configure link attributes and resource constraints such as maximum reservable bandwidth and resource class attributes for links (interaces) within the MPLS domain. A link state protocol that supports TE extensions (IS-IS-TE or OSPF-TE) is used to propagate information about network topology and link attribute to all routers in the routing area. Network administrators also specify all the LSPs that are to originate at each router. For each LSP, the network administrator specifies the destination node and the attributes of the LSP which indicate the requirements that must be satisfied during the path selection process. Each router then uses a local constraint-based routing process to compute explicit paths for all LSPs originating from it. Subsequently, a signaling
protocol is used to instantiate the LSPs. By assigning proper bandwidth values to links and LSPs, congestion caused by uneven traffic distribution can generally be avoided or mitigated. In order to perform constraint-based routing on a per-class basis for LSPs, the conventional IGPs (e.g., IS-IS and OSPF) should provide extensions to propagate per-class resource information [LC.

There are also proposals for using more centralized policy models to support TE implementation [WHj00, IYBKQ00]. As described in the TE Framework document, off-line (and on-line) TE considerations would be of limited utility if the network could not be controlled effectively to implement the results of TE decisions and to achieve desired network performance objectives. Capacity augmentation is a coarse grained solution to traffic engineering issues. However, it is simple and may be advantageous if bandwidth is abundant and cheap or if the current or expected network workload demands it. However, bandwidth is not always abundant and cheap, and the workload may not always demand additional capacity. Adjustments of administrative weights and other parameters associated with routing protocols provide finer grained control, but is difficult to use and imprecise because of the routing interactions that occur across the network. In certain network contexts, more flexible, finer grained approaches which provide more precise control over the mapping of traffic to routes and over the selection and placement of routes may be appropriate and useful.

Control mechanisms can be manual (e.g. administrative configuration), partially automated (e.g. scripts) or fully automated (e.g. policy based management systems). Automated mechanisms are particularly required in large scale networks. Multi-vendor interoperability can be facilitated by developing and deploying standardized management systems (e.g. standard MIBs) and policies (PIBs) to support the control functions required to address traffic engineering objectives such as load distribution and protection/restoration.

Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX4-3]

Internet Draft TE & QoS Methods for IP,ATM,TDM-Based Networks August 2001

MPLS depends on layer 3 mechanisms to determine LSP routes, and also the way that the routes are used. That is, MPLS has no routing of its own built in (it is "between layer 3 and layer 2"). Unlike layer 3 protocols, MPLS lacks addressing and routing components. It has to rely on IP, OSPF/BGP etc. for that. MPLS is not a layer 2 protocol either as it does not have a single format for data transmission, which is a requirement for a layer 2 protocol. How exactly the OSPF/IS-IS extensions get used, how policy-based capabilities get used, etc., to determine MPLS routing, is going to be a matter of vendor implementation. What is emerging are a lot of different capabilities to implement MPLS/TE in many ways, and service providers may provide requirements for a standardized "generic TE method", somewhat like a generic CAC in the ATM context discussed below. These standards requirement would then be used to drive the vendor implementations in the direction of network operator requirements and vendor interoperability.

The following assumptions are made regarding the outcomes of these IP-based routing standardization directions:

a) Call routing in support of connection establishment functions on a per-connection basis to determine the routing address based on a name/number translation, and uses a protocol such as H.323 [H.323] or the session initation protocol (SIP) [RFC2543]. It is assumed that the call routing interworks with the broadband ISDN user part (B-ISUP) [Q.2761] and bearer-independent call control (BICC) protocols [Q.1901] to accommodate setup and release of connection requests.

b) Connection/bandwidth-allocation routing in support of bearer-path selection is assumed to employ OSPF/BGP path selection methods in combination with MPLS. MPLS employs a constraint-based routing label
distribution protocol (CRLDP) [J00] or a resource reservation protocol (RSVP) [RFC2205] to establish constraint-based routing label switched paths (CRLSPs). Bandwidth allocation to CRLSPs is managed in support of QoS resource management, as discussed in ANNEX 3.

c) The MPLS label request message (equivalent to the setup message) carries the explicit route parameter specifying the via nodes (VNs) and destination node (DN) in the selected CRLSP and the depth-of-search (DoS) parameter specifying the allowed bandwidth selection threshold on a link.

d) The MPLS notify (equivalent to the release) message is assumed to carry the crankback/bandwidth-not-available parameter specifying return of control of the connection/bandwidth-allocation request to the originating node (ON), for possible further alternate routing to establish additional CRLSPs.

e) Call control routing is coordinated with connection/bandwidth-allocation for bearer-path establishment.

f) Reachability information is exchanged between all nodes. To provision a new IP address, the node serving that IP address is provisioned. The reachability information is flooded to all nodes in the network using the OSPF LSA flooding mechanism.

g) The ON performs destination name/number translation, service processing, and all steps necessary to determine the routing table for the connection/bandwidth-allocation request across the IP network. The ON makes a connection/bandwidth-allocation request admission if bandwidth is available and places the connection/bandwidth-allocation request on a selected CRLSP.

IP-based networks employ an IP addressing method to identify node endpoints [S94]. A mechanism is needed to translate E.164 AESAs to IP addresses in an efficient manner. Work is in progress [RFC2916, B99] to interwork between IP addressing and E.164 numbering/addressing, in which a translation database is required, based on domain name system (DNS) technology, to convert E.164 addresses to IP addresses. With such a capability, IP nodes could make this translation of E.164 AESAs directly, and thereby provide interworking with TDM- and ATM-based networks which use E.164 numbering and addressing. If this is the case, then E.164 AESAs could become a standard addressing method for interworking across IP-, ATM-, and TDM-based networks.

As stated above, path selection in an IP-based network is assumed to employ OSPF/BGP in combination with the MPLS protocol that functions efficiently in combination with call control establishment of individual connections. In OSPF-based layer 3 routing, as illustrated in Figure 4.1, an ON N1 determines a list of shortest paths by using, for example, Dijkstra’s algorithm.

This path list could be determined based on administrative weights of each link, which are communicated to all nodes within the AS group. These administrative weights may be set, for example, to 1 + epsilon x distance, where epsilon is a factor giving a relatively smaller weight to the distance in comparison to the hop count. The ON selects a path from the list based on, for example, FR, TDR, SDR, or EDR path selection, as described in ANNEX 2. For example, to establish a CRLSP on the first path, the ON N1 sends an MPLS label request message to VN N2, which in turn forwards the MPLS label request message to VN N3, and finally to DN N4. The VNs N2 and N3 and DN N4 are passed in the explicit route (ER) parameter contained in the MPLS label request message. Each node in the path reads the ER information, and passes the MPLS label request message to the next node listed in the ER parameter. If the first-choice path is blocked at any of the links in the path, a MPLS notify message with crankback/bandwidth-not-available parameter is returned to the ON which can then attempt the next path. If FR is used, then this
path is the next path in the shortest path list, for example path N1-N6-N7-N8-N4. If TDR is used, then the next path is the next path in the routing table for the current time period. If SDR is used, OSPF implements a distributed method of flooding link status information, which is triggered either periodically and/or by crossing load state threshold values. As described in the beginning of this Section, this method of distributing link status information can be resource intensive and indeed may not be any more efficient than simpler path selection methods such as EDR. If EDR is used, then the next path is the last successful path, and if that path is unsuccessful another alternate path is searched out according to the EDR path selection method.

Bandwidth-allocation control information is used to seize and modify bandwidth allocation on LSPs, to release bandwidth on LSPs, and for purposes of advancing the LSP choices in the routing table. Existing CRLSP label request (setup) and notify (release) messages, as described in [J00], can be used with additional parameters to control CRLSP bandwidth modification, DoS on a link, or CRLSP crankback/bandwidth-not-available to an ON for further alternate routing to search out additional bandwidth on alternate CRLSPs. Actual selection of a CRLSP is determined from the routing table, and CRLSP control information is used to establish the path choice. Forward information exchange is used in CRLSP setup and bandwidth modification, and includes for example the following parameters:

1. LABEL REQUEST - ER: The explicit route (ER) parameter in MPLS specifies each VN and the DN in the CRLSP, and used by each VN to determine the next node in the path.
2. LABEL REQUEST - DoS: The depth-of-search (DoS) parameter is used by each VN to compare the load state on each CRLSP link to the allowed DoS threshold to determine if the MPLS setup or modification request is admitted or blocked on that link.
3. LABEL REQUEST - MODIFY: The MODIFY parameter is used by each VN/DN to update the traffic parameters (e.g., committed data rate) on an existing CRLSP to determine if the MPLS modification request is admitted or blocked on each link in the CRLSP.

The setup-priority parameter serves as a DoS parameter in the MPLS LABEL REQUEST message to control the bandwidth allocation, queuing priorities, and bandwidth modification on an existing CRLSP [AAFJLLS00].

Backward information exchange is used to release a connection/bandwidth-allocation request on a link such as from a DN to a VN or from a VN to an ON, and includes for example the following parameter:

4. NOTIFY-BNA: The bandwidth-not-available parameter in the notify (release) message sent from the VN to ON or DN to ON, and allows for possible further alternate routing at the ON to search out alternate CRLSPs for additional bandwidth.

A bandwidth-not-available parameter is already planned for the MPLS NOTIFY message to allow the ON to search out additional bandwidth on additional CRLSPs.

In order to achieve automatic update and synchronization of the topology database, which is essential for routing table design, IP-based networks already interpret HELLO protocol mechanisms to identify links in the network. For topology database synchronization the OSPF LSA exchange is used to automatically provision nodes, links, and reachable addresses in the topology database. This information is exchanged between one node and
another node, and in the case of OSPF a flooding mechanism of LSA information is used.

5. HELLO: Provides for the identification of links between nodes in the network.

6. LSA: Provides for the automatic updating of nodes, links, and reachable addresses in the topology database.

In summary, IP-based networks already incorporate standard signaling for routing table management functions, which includes the ER, HELLO, and LSA capabilities. Additional requirements needed to support QoS resource management include the DoS parameter and MODIFY parameter in the MPLS LABEL REQUEST message, the crankback/bandwidth-not-available parameter in the MPLS notify message, as proposed in [FIA00, AALJ99], and the support for QUERY, STATUS, and RECOM routing table design information exchange, as required in Section 4.5. Call control with the H.323 [H.323] and session initiation protocol [RFC2543] protocols needs to be coordinated with MPLS CRLSP connection/bandwidth-allocation control.

4.3 Routing Table Management for ATM-Based Networks

PNNI is a standardized signaling and dynamic routing strategy for ATM networks adopted by the ATM Forum [ATM960055]. PNNI provides interoperability among different vendor equipment and scaling to very large networks. Scaling is provided by a hierarchical peer group structure that allows the details of topology of a peer group to be flexibly hidden or revealed at various levels within the hierarchical structure. Peer group leaders represent the nodes within a peer group for purposes of routing protocol exchanges at the next higher level. Border nodes handle inter-level interactions at call setup. PNNI routing involves two components: a) a topology distribution protocol, and b) the path selection and crankback procedures. The topology distribution protocol floods information within a peer group. The peer group leader abstracts the information from within the peer group and floods the abstracted topology information to the next higher level in the hierarchy, including aggregated reachable address information. As the peer group leader learns information at the next higher level, it floods it to the lower level in the hierarchy, as appropriate. In this fashion, all nodes learn of network-wide reachability and topology.

PNNI path selection is source-based in which the ON determines the high-level path through the network. The ON performs number translation, screening, service processing, and all steps necessary to determine the routing table for the connection/bandwidth-allocation request across the ATM network. The node places the selected path in the designated transit list (DTL) and passes the DTL to the next node in the SETUP message. The next node does not need to perform number translation on the called party number but just follows the path specified in the DTL. When a connection/bandwidth-allocation request is blocked due to network congestion, a PNNI crankback/bandwidth-not-available is sent to the first ATM node in the peer group. The first ATM node may then use the PNNI alternate routing after crankback/bandwidth-not-available capability to select another path for the connection/bandwidth-allocation request. If the network is flat, that is, all nodes have the same peer group level, the ON controls the edge-to-edge path. If the network has more than one level of hierarchy, as the call progresses from one peer group into another, the border node at the new peer group selects a path through that peer group to the next peer group downstream, as determined by the ON. This occurs recursively through the levels of hierarchy. If at any point the call is...
blocked, for example when the selected path bandwidth is not available, then
the call is cranked back to the border node or ON for that level of the
hierarchy and an alternate path is selected. The path selection algorithm
is not stipulated in the PNNI specification, and each ON implementation can
make its own path selection decision unilaterally. Since path selection is
done at an ON, each ON makes path selection decisions based on its local
topology database and specific algorithm. This means that different path
selection algorithms from different vendors can interwork with each other.

In the routing example illustrated in Figure 4.1 now used to illustrate
PNNI, an ON N1 determines a list of shortest paths by using, for example,
Dijkstra’s algorithm. This path list could be determined based on
administrative weights of each link which are communicated to all nodes
within the peer group through the PTSE flooding mechanism. These
administrative weights may be set, for example, to 1 + epsilon x distance,
where epsilon is a factor giving a relatively smaller weight to the distance
in comparison to the hop count. The ON then selects a path from the list
based on any of the methods described in ANNEX 2, that is FR, TDR, SDR, and
EDR. For example, in using the first choice path, the ON N1 sends a PNNI
setup message to VN N2, which in turn forwards the PNNI setup message to VN
N3, and finally to DN N4. The VNs N2 and N3 and DN N4 are passed in the DTL
parameter contained in the PNNI setup message. Each node in the path reads
the DTL information, and passes the PNNI setup message to the next node
listed in the DTL.

If the first path is blocked at any of the links in the path, or overflows
or is excessively delayed at any of the queues in the path, a
crankback/bandwidth-not-available message is returned to the ON which can
then attempt the next path. If FR is used, then this path is the next path
in the shortest path list, for example path N1-N6-N7-N8-N4. If TDR is used,
then the next path is the next path in the routing table for the current
Ash <draft-ietf-tewg-qos-routing-02.txt> [Page ANNEX4-7]

Connection/bandwidth-allocation control information is used in
collection/bandwidth-allocation set up to seize bandwidth in links, to
release bandwidth in links, and to advance path choices in the routing
table. Existing connection/bandwidth-allocation setup and release messages
[ATM960055] can be used with additional parameters to control SVC bandwidth
modification, DoS on a link, or SVC bandwidth-not-available to an ON for
further alternate routing. Actual selection of a path is determined from
the routing table, and connection/bandwidth-allocation control information
is used to establish the path choice. Forward information exchange is used
in connection/bandwidth-allocation set up, and includes for example the
following parameters:

1. SETUP-DTL/ER: The designated-transit-list/explicit-route (DTL/ER)
   parameter in PNNI specifies each VN and the DN in the path, and used by each
   VN to determine the next node in the path.
2. SETUP-DoS: The DoS parameter used by each VN to compare the load
   state on the link to the allowed DoS to determine if the SVC
   connection/bandwidth-allocation request is admitted or blocked on that link.
3. MODIFY REQUEST - DoS: The DoS parameter used by each VN to compare the load state on the link to the allowed DoS to determine if the SVP modification request is admitted or blocked on that link.

It is required that the DoS parameter be carried in the SVP MODIFY REQUEST and SVC SETUP messages, to control the bandwidth allocation and queuing priorities.

Backward information exchange is used to release a connection/bandwidth-allocation request on a link such as from a DN to a VN or from a VN to an ON, and includes for example the following parameter:

4. RELEASE-CB: The crankback/bandwidth-not-available parameter in the release message is sent from the VN to ON or DN to ON, and allows for possible further alternate routing at the ON.

5. MODIFY REJECT-BNA: The bandwidth-not-available parameter in the modify reject message is sent from the VN to ON or DN to ON, and allows for possible further alternate routing at the ON to search out additional bandwidth on alternate SVPs.

SVC crankback/bandwidth-not-available is already defined for PNNI-based signaling. We propose a bandwidth-not-available parameter in the SVP MODIFY REJECT message to allow the ON to search out additional bandwidth on additional SVPs.

In order to achieve automatic update and synchronization of the topology database, which is essential for routing table design, ATM-based networks already interpret HELLO protocol mechanisms to identify links in the network. For topology database synchronization the PTSE exchange is used to automatically provision nodes, links, and reachable addresses in the topology database. This information is exchanged between one node and another node, and in the case of PNNI a flooding mechanism of PTSE information is used.

6. HELLO: Provides for the identification of links between nodes in the network.

7. PTSE: Provides for the automatic updating of nodes, links, and reachable addresses in the topology database.

In summary, ATM-based networks already incorporate standard signaling and messaging directly applicable to routing implementation, which includes the DTL, crankback/bandwidth-not-available, HELLO, and PTSE capabilities. ATM protocol capabilities are being progressed [ATM000102, ATM000102, AM99] to support QoS resource management, which include the DoS parameter in the SVC SETUP and SVP MODIFY REQUEST messages, the bandwidth-not-available parameter in the SVP MODIFY REJECT message, and the QUERY, STATUS, and RECOM routing table design information exchange, as required in Section 4.5.

4.4 Routing Table Management for TDM-Based Networks

TDM-based voice/ISDN networks have evolved several dynamic routing methods, which are widely deployed and include TDR, SDR, and EDR implementations [A98]. TDR includes dynamic nonhierarchical routing (DNHR), deployed in the US Government FTS-2000 network. SDR includes dynamically controlled routing (DCR), deployed in the Stentor Canada, Bell Canada, MCI, and Sprint networks, and real-time network routing (RTNR), deployed in the AT&T network. EDR includes dynamic alternate routing (DAR), deployed in the British Telecom network, and STT, deployed in the AT&T network.

Internet Draft    TE & QoS Methods for IP,ATM,TDM-Based Networks  August 2001

Ash              <draft-ietf-tewg-qos-routing-02.txt>      [Page ANNEX4-8]
TDM-based network call routing protocols are described for example in [Q.1901] for BICC, and in [Q.2761] for the B-ISUP signaling protocol. We summarize here the information exchange required between network elements to implement the TDM-based path selection methods, which include connection control information required for connection set up, routing table design information required for routing table generation, and topology update information required for the automatic update and synchronization of topology databases.

Routing table management information is used for purposes of applying the routing table design rules for determining path choices in the routing table. This information is exchanged between one node and another node, such as between the ON and DN, for example, or between a node and a network element such as a BBP. This information is used to generate the routing table, and then the routing table is used to determine the path choices used in the selection of a path. The following messages can be considered for this function:

1. QUERY: Provides for an ON to DN or ON to BBP link and/or node status request.
2. STATUS: Provides ON/VN/DN to BBP or DN to ON link and/or node status information.
3. RECOM: Provides for an BBP to ON/VN/DN routing recommendation.

These information exchange messages are already deployed in non-standard TDM-based implementations, and need to be extended to standard TDM-based network environments.

In order to achieve automatic update and synchronization of the topology database, which is essential for routing table design, TDM-based networks need to interpret at the gateway nodes the HELLO protocol mechanisms of ATM-Ash <draft-ietf-teq-routing-02.txt> [Page ANNEX4-9] and IP-based networks to identify links in the network, as discussed above for ATM-based networks. Also needed for topology database synchronization is a mechanism analogous to the PTSE exchange, as discussed above, which automatically provisions nodes, links, and reachable addresses in the topology database.

Path-selection and QoS-resource management control information is used in connection/bandwidth-allocation set up to seize bandwidth in links, to release bandwidth in links, and for purposes of advancing path choices in the routing table. Existing connection/bandwidth-allocation setup and release messages, as described in Recommendations Q.71 and Q.2761, can be used with additional parameters to control path selection, DoS on a link, or crankback/bandwidth-not-available to an ON for further alternate routing. Actual selection of a path is determined from the routing table, and connection/bandwidth-allocation control information is used to establish the path choice.

Forward information exchange is used in connection/bandwidth-allocation set up, and includes for example the following parameters:

4. SETUP-DTL/ER: The designated-transit-list/explicit-route (DTL/ER) parameter specifies each VN and the DN in the path, and used by each VN to determine the next node in the path.
5. SETUP-DoS: The DoS parameter is used by each VN to compare the load state on the link to the allowed DoS to determine if the connection/bandwidth-allocation request is admitted or blocked on that link.

In B-ISUP these parameters could be carried in the initial address message
Backward information exchange is used to release a connection/bandwidth-allocation on a link such as from a DN to a VN or from a VN to an ON, and includes for example the following parameter:

6. RELEASE-CB: The crankback/bandwidth-not-available parameter in the release message is sent from the VN to ON or DN to ON, and allows for possible further alternate routing at the ON.

In B-ISUP signaling this parameter could be carried in the RELEASE message.

### 4.5 Signaling and Information Exchange Requirements

Table 4.1 summarizes the required signaling and information exchange methods supported within each routing technology which are required to be supported across network types. Table 4.1 identifies

a) the required information-exchange parameters, shown in non-bold type, to support the routing methods, and
b) the required standards, shown in bold type, to support the information-exchange parameters.

<table>
<thead>
<tr>
<th>Table 4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required Signaling and Information-Exchange Parameters to Support Routing Methods (Required Standards in Bold)</td>
</tr>
<tr>
<td>These information-exchange parameters and methods are required for use within each network type and for interworking across network types.</td>
</tr>
</tbody>
</table>

Therefore it is required that all information-exchange parameters identified in Table 4.1 be supported by the standards identified in the table, for each of the five network technologies. That is, it is required that standards be developed for all information-exchange parameters not currently supported, which are identified in Table 4.1 as references to Sections of this ANNEX. This will ensure information-exchange compatibility when interworking between the TDM-, ATM-, and IP-based network types, as denoted in the left three network technology columns. To support this information-exchange interworking across network types, it is further required that the information exchange at the interface be compatible across network types. Standardizing the required information routing methods and information-exchange parameters also supports the network technology cases in the right two columns of Table 4.1, in which PSTNs incorporate ATM- or IP-based technology.

We first discuss the routing methods identified by the rows of Table 4.1, and we then discuss the harmonization of PSTN/ATM-Based and PSTN/IP-Based information exchange, as identified by columns 4 and 5 of Table 4.1. In Sections 4.5.1 to 4.5.4, we describe, respectively the call routing (number translation to routing address), connection routing, QoS resource management, and routing table management information-exchange parameters required in Table 4.1. In Section 4.5.5, we discuss the harmonization of routing methods standards for the two technology cases in the right two columns of Table 4.1 in which PSTNs incorporate ATM- or IP-based technology.

### 4.5.1 Call Routing (Number Translation to Routing Address)

Information-Exchange Parameters
As stated before, in the draft we assume the separation of call-control signaling for call establishment from connection/bandwidth-allocation-control signaling for bearer-channel establishment. Call-control signaling protocols are described for example in [Q.2761] for the B-ISUP signaling protocol, [Q.1901] for BICC, [H.323] for the H.323 protocol, [RFC2805, GR99] for the media gateway control (MEGACO) protocol, and in [RFC2543] for SIP. Connection control protocols include for example [Q.2761] for B-ISUP signaling, [ATM960055] for PNNI signaling, [ATM960061] for UNI signaling, [ATM000148, DN99] for SVP signaling, and [J00, ABGLSS00] for MPLS signaling.

As discussed in ANNEX 2, number/name translation should result in the E.164 AESA addresses, INRAs, and/or IP addresses. It is required that provision be made for carrying E.164-AESA addresses, INRAs, and IP addresses in the connection-setup IE. In addition, it is required that a call identification code (CIC) be carried in the call-control and bearer-control connection-setup IEs in order to correlate the call-control setup with the bearer-control setup, [ATM000146]. Carrying these additional parameters in the Signaling System 7 (SS7) ISDN User Part (ISUP) connection-setup IEs is specified in the BICC protocol [Q.1901].

As shown in Table 4.1, it is required that provision be made for carrying E.164-AESA addresses, INRAs, and IP addresses in the connection-setup IE. In particular, it is required that E.164-AESA-address, INRA, and IP-address elements be developed within IP-based and PSTN/IP-based networks. It is required that number translation/routing methods supported by these parameters be developed for IP-based and PSTN/IP-based networks. When this is the case, then E.164-AESA addresses, INRAs, and IP addresses will become the standard addressing method for interworking across TDM-, ATM-, and IP-based networks.

4.5.2 Connection Routing Information-Exchange Parameters

Connection/bandwidth-allocation control information is used to seize bandwidth on links in a path, to release bandwidth on links in a path, and for purposes of advancing path choices in the routing table. Existing connection/bandwidth-allocation setup and connection-release IEs, as described in [Q.2761, ATM960055, ATM960061, ATM000148, J00], can be used with additional parameters to control SVC/SVP/CRLSP path routing, DoS bandwidth-allocation thresholds, and crankback/bandwidth-not-available to allow further alternate routing. Actual selection of a path is determined from the routing table, and connection/bandwidth-allocation control information is used to establish the path choice.

Source routing can be implemented through the use of connection/bandwidth-allocation control signaling methods employing the DTL or ER parameter in the connection-setup (IAM, SETUP, MODIFY REQUEST, and LABEL REQUEST) IE and the crankback (CBK)/bandwidth-not-available (BNA) parameter in the connection-release (RELEASE, MODIFY REJECT, and NOTIFY) IE. The DTL or ER parameter specifies all VNs and DN in a path, as determined by the ON, and the crankback/bandwidth-not-available parameter allows a VN to return control of the connection request to the ON for further alternate routing.

Forward information exchange is used in connection/bandwidth-allocation setup, and includes for example the following parameters:

1. Setup with designated-transit list/explicit-route (DTL/ER) parameter: The DTL parameter in PNNI or the ER parameter in MPLS specifies
each VN and the DN in the path, and is used by each VN to determine the next node in the path.

Backward information exchange is used to release a connection/bandwidth-allocation request on a link such as from a DN to a VN or from a VN to an ON, and the following parameters are required:

2. Release with crankback/bandwidth-not-available (CBK/BNA) parameter: The CBK/BNA parameter in the connection-release IE is sent from the VN to ON or DN to ON, and allows for possible further alternate routing at the ON.

It is required that the CBK/BNA parameter be included (as appropriate) in the RELEASE IE for TDM-based networks, the SVC RELEASE and SVP MODIFY REJECT IE for ATM-based networks, and MPLS NOTIFY IE for IP-based networks. This parameter is used to allow the ON to search out additional bandwidth on additional SVC/SVP/CRLSPs.

As shown in Table 4.1, it is required that the DTL/ER and CBK/BNA elements be developed within TDM-based networks, which will be compatible with the DTL element in ATM-based networks and the ER element in IP-based networks. It is required [E.350, E.351] that path-selection methods be developed supported by these parameters for TDM-based networks. Furthermore it is required that TDR and EDR path-selection methods be developed supported by these parameters for ATM-based, IP-based, PSTN/ATM-based, and PSTN/IP-based networks. When this is the case, then the DTL/ER and CBK/BNA parameters will become the standard path-selection method for interworking across TDM-*, ATM-*, and IP-based networks.

4.5.3 QoS Resource Management Information-Exchange Parameters

QoS resource management information is used to provide differentiated service priority in seizing bandwidth on links in a path and also in providing queuing resource priority. These parameters are required:

3. Setup with QoS parameters (QoS-PAR): The QoS-PAR include QoS thresholds such as transfer delay, delay variation, and packet loss. The QoS-PAR parameters are used by each VN to compare the link QoS performance to the requested QoS threshold to determine if the connection/bandwidth-allocation request is admitted or blocked on that link.

4. Setup with traffic parameters (TRAF-PAR): The TRAF-PAR include traffic parameters such as average bit rate, maximum bit rate, and minimum bit rate. The TRAF-PAR parameters are used by each VN to compare the link traffic characteristics to the requested TRAF-PAR thresholds to determine if the connection/bandwidth-allocation request is admitted or blocked on that link.

5. Setup with depth-of-search (DoS) parameter: The DoS parameter is used by each VN to compare the load state on the link to the allowed DoS to determine if the connection/bandwidth-allocation request is admitted or blocked on that link.

6. Setup with modify (MOD) parameter: The MOD parameter is used by each VN to compare the requested modified traffic parameters on an existing SVP/CRLSP to determine if the modification request is admitted or blocked on that link.

7. Differentiated services (DIFFSERV) parameter: The DIFFSERV parameter is used in ATM-based and IP-based networks to support priority queuing. The DIFFSERV parameter is used at the queues associated with each link to designate the relative priority and management policy for each queue.

It is required that the QoS-PAR, TRAF-PAR, DTL/ER, DoS, MOD, and DIFFSERV
parameters be included (as appropriate) in the initial address message (IAM) for TDM-based networks, the SVC/SVP SETUP IE and SVP MODIFY REQUEST IE for ATM-based networks, and MPLS LABEL REQUEST IE for IP-based networks. These parameters are used to control the routing, bandwidth allocation, and routing/queuing priorities.

As shown in Table 4.1, it is required that the QoS-PAR and TRAF-PAR elements be developed within TDM-based networks to support bandwidth allocation and protection, which will be compatible with the QoS-PAR and TRAF-PAR elements in ATM-based and IP-based networks. In addition, it is required that the DoS element be developed within TDM-based networks, which will be compatible with the DoS element in ATM-based and IP-based networks. Finally, it is required that the DIFFSERV element should be developed in ATM-based and IP-based networks to support priority queuing. It is required that QoS-resource-management methods be developed supported by these parameters for TDM-based networks. When this is the case, then the QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters will become the standard QoS-resource-management methods for interworking across TDM-, ATM-, and IP-based networks.

4.5.4 Routing Table Management Information-Exchange Parameters

Routing table management information is used for purposes of applying the routing table design rules for determining path choices in the routing table. This information is exchanged between one node and another node, such as between the ON and DN, for example, or between a node and a network element such as a BBP. This information is used to generate the routing table, and then the routing table is used to determine the path choices used in the selection of a path.

In order to achieve automatic update and synchronization of the topology database, which is essential for routing table design, ATM- and IP-based based networks already interpret HELLO protocol mechanisms to identify links in the network. For topology database synchronization the PTSE exchange is used in ATM-based networks and LSA is used in IP-based networks to automatically provision nodes, links, and reachable addresses in the topology database. Hence these parameters are required for this function:

8. HELLO parameter: Provides for the identification of links between nodes in the network.

These information exchange parameters are already deployed in ATM- and IP-based network implementations, and are required to be extended to TDM-based network environments.

The following parameters are required for the status query and routing recommendation function:

10. Routing-query-element (RQE) parameter: Provides for an ON to DN or ON to BBP link and/or node status request.
11. Routing-status-element (RSE) parameter: Provides for a node to BBP or DN to ON link and/or node status information.

These information exchange parameters are being standardized with Recommendation [E.350, E.351], and are required to be extended to ATM- and IP-based network environments.
As shown in Table 4.1, it is required that a TSE parameter be developed within TDM-based PSTN networks. It is required that topology update routing methods supported by these parameters be developed for PSTN/TDM-based networks. When this is the case, then the HELLO and TSE/PTSE/LSA parameters will become the standard topology update method for interworking across TDM-, ATM-, and IP-based networks.

As shown in Table 4.1, it is required that a RSE parameter be developed within TDM-based networks, which will be compatible with the PTSE parameter in ATM-based networks and the LSA parameter in IP-based networks. It is required [E.350, E.351] that status update routing methods supported by these parameters be developed for TDM-based networks. When this is the case, then the RSE/PTSE/LSA parameters will become the standard status update method for interworking across TDM-, ATM-, and IP-based networks.

As shown in Table 4.1, it is required that a RQE parameter be developed within ATM-based, IP-based, PSTN/ATM-based, and PSTN/IP-based networks. It is required that query-for-status routing methods supported by these parameters be developed for ATM-based, IP-based, PSTN/ATM-based, and PSTN/IP-based networks. When this is the case, then the RQE parameters will become the standard query for status method for interworking across TDM-, ATM-, and IP-based networks.

4.5.6 Open Routing Application Programming Interface (API)

Application programming interfaces (APIs) are being developed to allow control of network elements through open interfaces available to individual applications. APIs allow applications to access and control network
functions including routing policy, as necessary, according to the specific application functions. The API parameters under application control, such as those specified for example in [PARLAY], are independent of the individual protocols supported within the network, and therefore can provide a common language and framework across various network technologies, such as TDM-, ATM-, and IP-based technologies.

The signaling/information-exchange connectivity management parameters specified in this Section which need to be controlled through an applications interface include QoS-PAR, TRAF-PAR, DTL/ER, DoS, MOD, DIFFSERV, E.164-AESA, INRA, CIC, and perhaps others. The signaling/information-exchange routing policy parameters specified in this Section which need to be controlled through an applications interface include TSE, RQE, RRE, and perhaps others. These parameters are required to be specified within the open API interface for routing functionality, and in this way applications will be able to access and control routing functionality within the network independent of the particular routing protocol(s) used in the network.

4.6 Examples of Internetwork Routing

A network consisting of various subnetworks using different routing protocols is considered in this Section. As illustrated in Figure 4.2, consider a network with four subnetworks denoted as networks A, B, C, and D, where each network uses a different routing protocol. In this example, network A is an ATM-based network which uses PNNI EDR path selection, network B is a TDM-based network which uses centralized periodic SDR path selection, network C is an IP-based network which uses MPLS EDR path selection, and network D is a TDM-based network which uses TDR path selection. Internetwork E is defined by the shaded nodes in Figure 4.2 and is a virtual network where the interworking between networks A, B, C, and D is actually taking place.

Figure 4.2. Example of an Internetwork Routing Scenario. BBPb denotes a bandwidth broker processor in network B for a centralized periodic SDR method. The set of shaded nodes is internetwork E for routing of connection/bandwidth-allocation requests between networks A, B, C, and D.

4.6.1 Internetwork E Uses a Mixed Path Selection Method

Internetwork E can use various path selection methods in delivering connection/bandwidth-allocation requests between the subnetworks A, B, C, and D. For example, internetwork E can implement a mixed path selection method in which each node in internetwork E uses the path selection method used in its home subnetwork. Consider a connection/bandwidth-allocation request from node a1 in network A to node b4 in network B. Node a1 first paths the connection/bandwidth-allocation request to either node a3 or a4 in network A and in doing so uses EDR path selection. In that regard node a1 first tries to route the connection/bandwidth-allocation request on the direct link a1-a4, and assuming that link a1-a4 bandwidth is unavailable then selects the current successful path a1-a3-a4 and routes the connection/bandwidth-allocation request to node a4 via node a3. In so doing node a1 and node a3 put the DTL/ER parameter (identifying ON a1, VN a3, and DN a4) and QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection/bandwidth-allocation request connection-setup IE.

Node a4 now proceeds to route the connection/bandwidth-allocation request to node b1 in subnetwork B using EDR path selection. In that regard node a4 first tries to route the connection/bandwidth-allocation request on the...
direct link a4-b1, and assuming that link a4-b1 bandwidth is unavailable then selects the current successful path a4-c2-b1 and routes the connection/bandwidth-allocation request to node b1 via node c2. In so doing node a4 and node c2 put the DTL/ER parameter (identifying ON a4, VN c2, and DN b1) and QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection/bandwidth-allocation request connection-setup IE.

If node c2 finds that link c2-b1 does not have sufficient available bandwidth, it returns control of the connection/bandwidth-allocation request to node a4 through use of a CBK/BNA parameter in the connection-release IE. If now node a4 finds that link d4-b1 has sufficient idle bandwidth capacity based on the RSE parameter in the status response IE from node b1, then node a4 could next try path a4-d3-d4-b1 to node b1. In that case node a4 routes the connection/bandwidth-allocation request to node d3 on link a4-d3, and node d3 is sent the DTL/ER parameter (identifying ON a4, VN d3, VN d4, and DN b1) and the DoS parameter in the connection-setup IE. In that case node d3 tries to seize idle bandwidth on link d3-d4, and assuming that there is sufficient idle bandwidth routes the connection/bandwidth-allocation request to node d4 with the DTL/ER parameter (identifying ON a4, VN d3, VN d4, and DN b1) and the QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection-setup IE. Node d4 then routes the connection/bandwidth-allocation request on link d4-b1 to node b1, which has already been determined to have sufficient idle bandwidth capacity. If on the other hand there is insufficient idle d4-b1 bandwidth available, then node d3 returns control of the call to node a4 through use of a CRK/BNA parameter in the connection-release IE. At that point node a4 may try another multilink path, such as a4-a3-b3-b1, using the same procedure as for the a4-d3-d4-b1 path.

Node b1 now proceeds to route the connection/bandwidth-allocation request to node b4 in network B using centralized periodic SDR path selection. In that regard node b1 first tries to route the connection/bandwidth-allocation request on the direct link b1-b4, and assuming that link b1-b4 bandwidth is unavailable then selects a two-link path b1-b2-b4 which is the currently recommended alternate path identified in the RRE parameter from the BBPb for network B. BBPb bases its alternate routing recommendations on periodic (say every 10 seconds) link and traffic status information in the RSE parameters received from each node in network B. Based on the status information, BBPb then selects the two-link path b1-b2-b4 and sends this alternate path recommendation in the RRE parameter to node b1 on a periodic basis (say every 10 seconds). Node b1 then routes the connection/bandwidth-allocation request on link b1-a4 to node a4. In so doing node b1 and node b2 put the DTL/ER parameter (identifying ON b1, VN b2, and DN b4) and QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection/bandwidth-allocation request connection-setup IE.

A connection/bandwidth-allocation request from node b4 in network B to node a1 in network A would mostly be the same as the connection/bandwidth-allocation request from a1 to b4, except with all the above steps in reverse order. The difference would be in routing the connection/bandwidth-allocation request from node b1 in network B to node a4 in network A. In this case, based on the mixed path selection assumption in virtual network E, the b1 to a4 connection/bandwidth-allocation request would use centralized periodic SDR path selection, since node b1 is in network B, which uses centralized periodic SDR. In that regard node b1 first tries to route the connection/bandwidth-allocation request on the direct link b1-a4, and assuming that link b1-a4 bandwidth is unavailable then selects a two-link path b1-c2-a4 which is the currently recommended alternate path identified in the RRE parameter from the BBPb for virtual
network E. BBPb bases its alternate routing recommendations on periodic (say every 10 seconds) link and traffic status information in the RSE parameters received from each node in virtual subnetwork E. Based on the status information, BBPb then selects the two-link path b1-c2-a4 and sends this alternate path recommendation in the RRE parameter to node b1 on a periodic basis (say every 10 seconds). Node b1 then routes the connection/bandwidth-allocation request to node a4 via VN c2. In so doing node b1 and node c2 put the DTL/ER parameter (identifying ON b1, VN c2, and DN a4) and QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection/bandwidth-allocation request connection-setup IE.

If node c2 finds that link c2-a4 does not have sufficient available bandwidth, it returns control of the connection/bandwidth-allocation request to node b1 through use of a CRK/BNA parameter in the connection-release IE. If now node b1 finds that path b1-d4-d3-a4 has sufficient idle bandwidth capacity based on the RSE parameters in the status IEs to BBPb, then node b1 could next try path b1-d4-d3-a4 to node a4. In that case node b1 routes the connection/bandwidth-allocation request to node d4 on link b1-d4, and node d4 is sent the DTL/ER parameter (identifying ON b1, VN d4, VN d3, and DN a4) and the QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection-setup IE. In that case node d4 tries to seize idle bandwidth on link d4-d3, and assuming that there is sufficient idle bandwidth routes the connection/bandwidth-allocation request to node d3 with the DTL/ER parameter (identifying ON b1, VN d4, VN d3, and DN a4) and the QoS-PAR, TRAF-PAR, DoS, and DIFFSERV parameters in the connection-setup IE. Node d3 then routes the connection/bandwidth-allocation request on link d3-a4 to node a4, which is expected based on status information in the RSE parameters to have sufficient idle bandwidth capacity. If on the other hand there is insufficient idle d3-a4 bandwidth available, then node d3 returns control of the call to node b1 through use of a CRK/BNA parameter in the connection-release IE. At that point node b1 may try another multilink path, such as b1-b3-a3-a4, using the same procedure as for the b1-d4-d3-a4 path.

Allocation of end-to-end performance parameters across networks is addressed in Recommendation I.356, Section 9. An example is the allocation of the maximum transfer delay to individual network components of an end-to-end connection, such as national network portions, international portions, etc.

4.6.2 Internetwork E Uses a Single Path Selection Method

Internetwork E may also use a single path selection method in delivering connection/bandwidth-allocation requests between the networks A, B, C, and D. For example, internetwork E can implement a path selection method in which each node in internetwork E uses EDR. In this case the example connection/bandwidth-allocation request from node a1 in network A to node b4 in network B would be the same as described above. A connection/bandwidth-allocation request from node b4 in network B to node a1 in network A would be the same as the connection/bandwidth-allocation request from a1 to b4, except with all the above steps in reverse order. In this case the routing of the connection/bandwidth-allocation request from node b1 in network B to node a4 in network A would also use EDR in a similar manner to the a1 to b4 connection/bandwidth-allocation request described above.

4.7 Modeling of Traffic Engineering Methods

In this Section, we again use the full-scale national network model developed in ANNEX 2 to study various TE scenarios and tradeoffs. The
135-node national model is illustrated in Figure 2.9, the multiservice traffic demand model is summarized in Table 2.1, and the cost model is summarized in Table 2.2.

As we have seen, routing table management entails many different alternatives and tradeoffs, such as:

* centralized routing table control versus distributed control
* pre-planned routing table control versus on-line routing table control
* per-flow traffic management versus per-virtual-network traffic management
* sparse logical topology versus meshed logical topology
* FR versus TDR versus SDR versus EDR path selection
* multilink path selection versus two-link path selection
* path selection using local status information versus global status information
* global status dissemination alternatives including status flooding, distributed query for status, and centralized status in a bandwidth-broker processor

Here we evaluate the tradeoffs in terms of the number of information elements and parameters exchanged, by type, under various TE scenarios. This approach gives some indication of the processor and information exchange load required to support routing table management under various alternatives. In particular, we examine the following cases:

Tables 4.2 and 4.3 summarize the comparative results for these cases, for the case of SDR path selection and STT path selection, respectively. The 135-node multiservice model was used for a simulation under a 30% general network overload in the network busy hour.

Table 4.2
Signaling and Information-Element Parameters Exchanged for Various TE Methods with SDR Per-Flow Bandwidth Allocation
Number of IE Parameters Exchanged under 30% General Overload in Network Busy Hour
(135-Node Multiservice Network Model)

Table 4.3
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Bandwidth Allocation
Number of IE Parameters Exchanged under 30% General Overload in Network Busy Hour
(135-Node Multiservice Network Model)

Tables 4.4 and 4.5 summarize the comparative results for the case of SDR path selection and STT path selection, respectively, in which the 135-node multiservice model was used for a simulation under a 6-times focused overload on the OKBK node in the network busy hour.

Table 4.4
Signaling and Information-Element Parameters Exchanged for Various TE Methods with SDR Per-Flow Bandwidth Allocation
Number of IE Parameters Exchanged under 6X Focused Overload on OKBK in Network Busy Hour
(135-Node Multiservice Network Model)

Table 4.5
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Bandwidth Allocation
Number of IE Parameters Exchanged under 6X Focused Overload on OKBK in Network Busy Hour
(135-Node Multiservice Network Model)

Tables 4.6 and 4.7 summarize the comparative results for the case of SDR path selection and STT path selection, respectively, in which the 135-node multiservice model was used for a simulation under a facility failure on the CHCG-NYCM link in the network busy hour.

Table 4.6
Signaling and Information-Element Parameters Exchanged for Various TE Methods with SDR Per-Flow Bandwidth Allocation
Number of IE Parameters Exchanged under Failure of CHCG-NYCM Link in Network Busy Hour
(135-Node Multiservice Network Model)

Table 4.7
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Bandwidth Allocation
Number of IE Parameters Exchanged under Failure of CHCG-NYCM Link in Network Busy Hour
(135-Node Multiservice Network Model)

Tables 4.8 - 4.10 summarize the comparative results for the case of STT path selection, in the hierarchical network model shown in Figure 2.10, for the 30% general overload, the 6-times focused overload, and the link failure scenarios, respectively. Both the per-flow bandwidth allocation and per-virtual network bandwidth allocation cases are given in these tables.

Table 4.8
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Per-Virtual-Network Bandwidth Allocation
Number of IE Parameters Exchanged under 30% General Overload in Network Busy Hour
(135-Edge-Node & 21-Backbone-Node Hierarchical Multiservice Network Model)

Table 4.9
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Per-Virtual-Network Bandwidth Allocation
Number of IE Parameters Exchanged under 6X Focused Overload on OKBK in Network Busy Hour;
(135-Edge-Node & 21-Backbone-Node Hierarchical Multiservice Network Model)

Table 4.10
Signaling and Information-Element Parameters Exchanged for Various TE Methods with STT-EDR Per-Virtual-Network Bandwidth Allocation
Number of IE Parameters Exchanged under Failure of CHCG-NYCM Link in Network Busy Hour;
(135-Edge-Node & 21-Backbone-Node Hierarchical Multiservice Network Model)
Tables 4.2 - 4.10 illustrate the potential benefits of EDR methods in reducing the routing table management overhead. In ANNEX 3 we discussed EDR methods applied to QoS resource management, in which the connection bandwidth-allocation admission control for each link in the path is performed based on the local status of the link. That is, the ON selects any path for which the first link is allowed according to QoS resource management criteria. Each VN then checks the local link status of the links specified in the ER parameter against the DoS parameter. If a subsequent link is not allowed, then a release with crankback/bandwidth-not-available is used to return to the ON which may then select an alternate path. This use of this EDR path selection method, then, which entails the use of the release with crankback/bandwidth-not-available mechanism to search for an available path, is an alternative to the SDR path selection alternatives, which may entail flooding of frequently changing link state parameters such as available-cell-rate.

A "least-loaded routing" strategy based on available-bit-rate on each link in a path, is used in the SDR dynamic routing methods illustrated in the above tables, and is a well-known, successful way to implement dynamic routing. Such SDR methods have been used in several large-scale network applications in which efficient methods are used to disseminate the available-link-bandwidth status information, such as the query for status method using the RQE and RRE parameters. However, there is a high overhead cost to obtain the available-link-bandwidth information when using flooding techniques, such as those which use the TSE parameter for link-state flooding. This is clearly evident in Tables 4.2 - 4.10. As a possible way around this, the EDR routing methods illustrated above do not require the dynamic flooding of available-bit-rate information. When EDR path selection with crankback is used in lieu of SDR path selection with link-state flooding, the reduction in the frequency of such link-state parameter flooding allows for larger peer group sizes. This is because link-state flooding can consume substantial processor and link resources, in terms of message processing by the processors and link bandwidth consumed on the links. Crankback/bandwidth-not-available is then an alternative to the use of link-state-flooding algorithm for the ON to be able to determine which subsequent links in the path will be allowed.

4.8 Conclusions/Recommendations

The conclusions/recommendations reached in this ANNEX are as follows:

* Per-VNET bandwidth allocation is recommended and is preferred to per-flow allocation because of the much lower routing table management overhead requirements. Per-VNET bandwidth allocation is essentially equivalent to per-flow bandwidth allocation in network performance and efficiency, as discussed in ANNEX 3.

* EDR TE methods are recommended and can lead to a large reduction in ALB flooding overhead without loss of network throughput performance. While SDR TE methods typically use ALB flooding for TE path selection, EDR TE methods do not require ALB flooding. Rather, EDR TE methods typically search out capacity by learning models, as in the STT method. ALB flooding can be very resource intensive, since it requires link bandwidth to carry LSAs, processor capacity to process LSAs, and the overhead can limit area/autonomous system (AS) size.

* EDR TE methods are recommended and lead to possible larger administrative areas as compared to SDR-based TE methods because of lower routing table management overhead requirements. This can help achieve
single-area flat topologies which, as discussed in ANNEX 3, exhibit better network performance and, as discussed in ANNEX 6, greater design efficiencies in comparison with multi-area hierarchical topologies.

Figures:

Figure 4.1 IP/MPLS Routing Example
Figure 4.2 Example of an Internetwork Routing Scenario

ANNEX 5
Dynamic Transport Routing Methods

5.1 Introduction

This ANNEX describes and analyzes transport network architectures in light of evolving technology for integrated broadband networks. Dynamic transport routing offers advantages of simplicity of design and robustness to load variations and network failures. Dynamic transport routing can combine with dynamic traffic routing to shift transport bandwidth among node pairs and services through use of flexible transport switching technology. Dynamic transport routing can provide automatic link provisioning, diverse link routing, and rapid link restoration for improved transport capacity utilization and performance under stress.

We present reliable transport routing models to achieve reliable network design, so as to provide service for predefined restoration objectives for any transport link or node failure in the network and continue to provide connections to customers with essentially no perceived interruption of service. We show that robust routing techniques such as dynamic traffic routing, multiple ingress/egress routing, and logical link diversity routing improve response to node or transport failures.

Cross-connect devices, such as optical cross-connects (OXC)s, are able to node transport channels, for example OC48 channels, onto different higher-capacity transport links such as an individual WDM channel on a fiberoptic cable. Transport paths can be rearranged at high speed using OXC,s, typically within tens of milliseconds switching times. These OXC,s can reconfigure logical transport capacity on demand, such as for peak day traffic, weekly redesign of link capacity, or emergency restoration of capacity under node or transport failure. Re-arrangement of logical link capacity involves reallocating both transport bandwidth and node terminations to different links. OXC technology is amenable to centralized traffic management.

There is recent work in extending MPLS control capabilities to the setup of layer 2 logical links through OXC,s, this effort dubbed multiprotocol lambda switching, after the switching of wavelengths in dense wavelength division multiplexer (DWDM) technology [ARDC99].

5.2 Dynamic Transport Routing Principles

An important element of network architecture is the relationship between the transport network and the traffic network. An illustration of a transport network is shown in Figure 5.1, and Figure 5.2 illustrates the mapping of layer-2 logical links in the traffic network onto the layer-1 physical transport network of Figure 5.1. Some logical links overlay two or more
fiber-backbone links. For example, in Figure 5.1, logical link AD traverses fiber-backbone links AB, BC, and CD.

Figure 5.2 further illustrates the difference between the physical transport network (layer 1) and the logical transport network (layer 2). Logical links are individual logical connections between network nodes, which make up the logical link connections and are routed on the physical transport network. Logical links can be provisioned at given rates, such as T1, OC3, OC12, OC48, OC192, etc., and is dependent on the level of traffic demand between nodes.

Figure 5.2 indicates that a direct logical link is obtained by cross-connecting through a transport switching location. Thus, the traffic network is a logical network overlaid on a sparse physical one. A cross-connect device is traversed at each network node on a given logical link path, as illustrated in Figure 5.2. This is particularly promising when such a device has low cost.

It is clear from Figures 5.1 and 5.2 that in a highly interconnected traffic network, or logical transport network, many node pairs may have a "direct" logical link connection where none exists in the physical transport network. In this case a direct logical link is obtained by cross-connecting through a transport switching location, such as an OXC. This is distinct from the traffic routing situation, in which a bearer connection is actually switched at an intermediate location. This distinction between cross-connecting and switching is a bit subtle, but it is fundamental to traffic routing of calls and transport routing of logical links. Referring to Figure 5.2, we illustrate one of the logical inconsistencies we encounter when we design the traffic network to be essentially separate from the transport network. On the alternative traffic path from node B to node D through A, the physical path is, in fact, up and back from B to A (a phenomenon known as "backhauling") and then across from B to D. The sharing of capacity by various traffic loads in this way actually increases the efficiency of the network because the backhauled capacity to and from B and A is only used when no direct A-to-B or A-to-D traffic wants to use it. It is conceivable that under certain conditions, capacity could be put to more efficient use, and this is studied in this ANNEX.

Hence a logical link connection is obtained by cross-connecting through transport switching devices, such as OXCs, and this is distinct from per-flow routing, which switches a call on the logical links at each node in the call path. In this way, the logical transport network is overlaid on a sparser physical transport network. In ANNEX 2 we discussed a wide variety of dynamic traffic routing methods. Dynamic transport routing methods incorporate dynamic path selection which seeks out and uses idle network capacity by using frequent, perhaps call-by-call, traffic and transport routing table update decisions. The trend in both traffic and transport routing architecture is toward greater flexibility in resource allocation, which includes transport and switching resource allocation. A fixed transport routing architecture may have dynamic traffic routing but fixed transport routing of logical link capacity. In a dynamic transport routing architecture, however, the logical link capacities can be rapidly rearranged ---that is, they are not fixed.

With dynamic transport routing, the logical transport bandwidth is shifted rapidly at layer 2 among node pairs and services through the use of dynamic cross-connect devices. In this case, the layer-1 physical fiber-link bandwidth is allocated among the layer-2 logical links. Bandwidth allocation at layer 3 also creates the equivalent of direct links, and we
refer to these links as traffic trunks, which in turn comprise virtual networks (VNETs) as described in ANNEX 3. Traffic trunks can be implemented, for example, by using MPLS label switched paths (LSPs). Bandwidth is allocated to traffic trunks in accordance with traffic demands, and normally not all logical link bandwidth is assigned; thus, there is a pool of unassigned bandwidth. In cases of traffic overload for a given node pair, the node first sets up calls on the traffic trunk that connects the node pair. If that is not possible the node then sets up calls on the available pool of bandwidth. If there is available bandwidth, then the bandwidth is allocated to the traffic trunk and used to set up the call. If bandwidth is not available, then the layer-2 logical link bandwidth might be dynamically increased by the bandwidth broker, and then allocated to the traffic trunk and finally the call. In a similar manner, in the event that bandwidth is underutilized in a traffic trunk, excess bandwidth is released to the available pool of bandwidth and then becomes available for assignment to other node pairs. If logical link bandwidth is sufficiently underutilized, the bandwidth might be returned to the available pool of layer-1 fiber-link bandwidth. The bandwidth broker reassigns network resources on a dynamic basis, through analysis of traffic data collected from the individual nodes.

In the dynamic transport architecture, we allow logical link between the various nodes to be rearranged rapidly, such as by hour of the day, or perhaps in real time. Dynamic transport routing capability enables rearrangement of the logical link capacities on demand. This capability appears most desirable for use in relatively slow rearrangement of capacity, such as for busy-hour traffic, weekend traffic, peak-day traffic, weekly redesign of logical link capacities, or for emergency restoration of capacity under node or transport failure. At various times the demands for node and transport capacity by the various node pairs and services that ride on the same optical fibers will differ. In this network, if a given demand for logical link capacity between a certain node pair decreases and a second goes up, we allow the logical link capacity to be reassigned to the second node pair. The ability to rearrange logical link capacity dynamically and automatically results in cost savings. Large segments of bandwidth can be provided on fiber routes, and then the transport capacity can be allocated at will with the rearrangement mechanism. This ability for simplified capacity management is discussed further in ANNEX 6.

Figure 5.3 illustrates the concept of dynamic traffic (layer 3) and transport routing (layer 2) from a generalized switching node point of view. Figure 5.3 illustrates the relationship of the call-level and transport-level dynamic routing methods used in the dynamic transport routing network. Dynamic connection routing, such as discussed in ANNEX 2, is used to route calls comprising the underlying traffic demand. Traffic trunk capacity allocations are made for each VNET on the transport link capacity. For each call the originating node analyzes the called number and determines the terminating node, class-of-service, and virtual network. The originating node tries to set up the call on the traffic trunk to the terminating node and, if unavailable, dynamic routing is used to rearrange the traffic trunk capacity as required to match the traffic demands and to achieve inter-node diversity, access diversity, and traffic trunk restoration following node, OXC, or fiber transport failures. The traffic trunk capacities are allocated by the traffic router to the logical link bandwidth, and the logical link bandwidth allocated by the bandwidth broker to the fiber-link bandwidth, such that the bandwidth is efficiently used according to the level of traffic demand between the nodes.

At the traffic demand level in the transmission hierarchy, flow requests are
switched using dynamic traffic routing on the logical link network by node routing logic. At the OC3 and higher demand levels in the transmission hierarchy, logical link demands are switched using OXC systems, which allow dynamic transport routing to route transport demands in accordance with traffic levels. Real-time logical link and real-time response to traffic congestion can be provided by OXC dynamic transport routing to improve network performance.

As illustrated in Figure 5.4, the dynamic transport routing network concept includes backbone routers (BRs), access routers (ARs), and OXCs. Access routers could route traffic from local offices, access tandems, customer premises equipment, and overseas international switching centers. Here a logical link transmission channel could consist, for example, of OC3-, OC12-, OC48-, or OCx-level bandwidth allocation. An OXC can cross-connect (or "switch") a logical link transmission channel within one terminating fiber wavelength channel in a dense wavelength division multiplex (DWDM) system to a like-channel within another fiber DWDM system. In the example illustrated, access routers connect to the OXC by means of transport links such as link AX1, and BRs connect to OXCs by means of transport links such as BX1. A number of backbone fiber/DWDM transport links interconnect the OXC network elements, such as links XX1 and XX2. Backbone logical links are terminated at each end by OXCs and are routed over fiber/DWDM spans on the physical transport network on the shortest physical paths. Inter-BR logical links are formed by cross-connecting the bandwidth channels through OXCs between a pair of BRs.

For example, the backbone logical link B2 from BR1 to BR3 is formed by connecting between BR1 and BR3 through fiber/DWDM links BX1, XX1, XX2, and BX3 by making appropriate cross-connects through OXC1, OXC2, and OXC3. Logical links have variable bandwidth capacity controlled by the bandwidth broker implementing the dynamic transport routing network. Access logical links are formed by cross-connecting between ARs and BRs---for example, access router AR1 connected on fiber/DWDM links AX1 and BX1 through OXC1 to BR1 or, alternatively, access router AR1 connected on fiber/DWDM links AX1, XX1, and BX2 cross-connected through OXC1 and OXC2 to BR2. For additional network reliability, backbone routers and access routers may be dual-homed to two OXCs, possibly in different building locations.

5.3 Dynamic Transport Routing Examples

There are significant network design opportunities with dynamic transport routing, and in this Section we give examples of dynamic transport routing over different time scales. These examples illustrate the network efficiency and performance improvements possible with seasonal, weekly, daily, and real-time transport rearrangement.

An illustration of dynamic transport routing for varying seasonal traffic demands is given in Figure 5.5. As seasonal demands shift, the dynamic transport network is better able to match demands to routed transport capacity, thus gaining efficiencies in transport requirements. Figure 5.5 illustrates how dynamic transport routing achieves network capacity reductions, and shows how transport demand is routed according to varying seasonal requirements. As seasonal demands shift, the dynamic transport network is better able to match demands to routed transport capacity, thus gaining efficiencies in transport requirements. The figure illustrates the variation of winter and summer capacity demands. With fixed transport routing, the maximum termination capacity and transport capacity are provided across the seasonal variations, because in a manual environment without dynamic transport rearrangement it is not possible to disconnect and reconnect capacity on such short cycle times. When transport rearrangement
is automated with dynamic transport routing, however, the termination and transport design can be changed on a weekly, daily, or, with high-speed packet switching, real-time basis to exactly match the termination and transport design with the actual network demands. Notice that in the fixed transport network there is unused termination and transport capacity that cannot be used by any demands; sometimes this is called "trapped capacity," because it is available but cannot be accessed by any actual demand. The dynamic transport network, in contrast, follows the capacity demand with flexible transport routing, and together with transport network design it reduces the trapped capacity.

Therefore, the variation of demands leads to capacity-sharing efficiencies, which in the example of Figure 5.5 reduce termination capacity requirements by 50 node terminations, or approximately 10 percent compared with the fixed transport network, and by 50 transport capacity requirements, or approximately 14 percent. Therefore, with dynamic transport routing capacity utilization can be made more efficient in comparison with fixed transport routing, because with dynamic transport network design the link sizes can be matched to the network load.

With dynamic traffic routing and dynamic transport routing design models, reserve capacity can be reduced in comparison with fixed transport routing. In-place capacity that exceeds the capacity required to exactly meet the design loads with the objective performance is called reserve capacity. Reserve capacity comes about because load uncertainties, such as forecast errors, tend to cause capacity buildup in excess of the network design that exactly matches the forecast loads. Reluctance to disconnect and rearrange traffic trunk and transport capacity contributes to this reserve capacity buildup. Typical ranges for reserve capacity are from 15 to 25 percent or more of network cost. Models show that dynamic traffic routing compared with fixed traffic routing provides a potential 5 percent reduction in reserve capacity while retaining a low level of short-term capacity design [A98].

With dynamic transport network design the link sizes can be matched to the network load. With dynamic transport routing, the link capacity disconnect policy becomes, in effect, one in which link capacity is always disconnected when not needed for the current traffic loads. Models given in [FHH79] predict reserve capacity reductions of 10~percent or more under this policy, and the results presented in Section 5.4 based on weekly dynamic transport design substantiate this conclusion.

Weekly design and rearrangement of logical link capacity can approach zero reserve capacity designs. Figures 5.6 and 5.7 illustrate the changing of routed transport capacity on a weekly basis between node pairs A--B, C--D, and B--E, as demands between these node pairs change on a weekly basis.

These transport routing and capacity changes are made automatically in the dynamic transport network, in which diverse transport routing of logical links A--B and C--D is maintained by the dynamic transport routing network. Logical link diversity achieves additional network reliability.

Daily design and rearrangement of transport link capacity can achieve performance improvements for similar reasons, due to noncoincidence of transport capacity demands that can change daily. An example is given in Figures 5.8 and 5.9 for traffic noncoincidence experienced on peak days such as Christmas Day. In Figure 5.8, we illustrate the normal business-day routing of access demands and inter-BR demands. On Christmas Day, however, there are many busy nodes and many idle nodes. For example, node BR2 may be relatively idle on Christmas Day (for example, if it were a downtown
This kind of access demand and inter-BR demand rearrangement, based on noncoincident traffic shifts, more traffic to and from BR1 can be completed because inter-BR logical link capacity is increased, now using freed-up transport capacity from the reduction in the transport capacity needed by BR2. On a peak day such as Christmas Day, the busy nodes are often limited by inter-BR logical link capacity; this rearrangement reduces or eliminates this bottleneck, as is illustrated in the Christmas Day dynamic transport network design example in Section 5.4.

The balancing of access and inter-BR capacity throughout the network can lead to robustness to unexpected load surges. This load-balancing design is illustrated in Section 5.4 with an example based on a Hurricane-caused focused overload in the northeastern United States. Capacity addition rearrangements based on instantaneous reaction to unforeseen events such as earthquakes could be made in the dynamic transport network.

Dynamic transport routing can provide dynamic restoration of failed capacity, such as that due to fiber cuts, onto spare or backup transport capacity. Dynamic transport routing provides a self-healing network capability to ensure a networkwide path selection and immediate adaptation to failure.

FASTAR [CED91], for example, implements central automatic control of transport switching devices to quickly restore service following a transport failure. As illustrated in Figure 5.10, a fiber cut can disrupt large traffic trunk capacities, and dynamic transport restoration can quickly restore transport capacity. Dynamic transport routing provides a self-healing network capability to ensure a networkwide path selection and immediate adaptation to failure. As illustrated in Figure 5.10, a fiber cut near the Nashville node severed 8.576 Gbps of traffic trunk capacity of switched-network traffic (there was also private-line traffic), and after dynamic transport restoration a total of 3.84 Gbps of traffic trunk capacity was still out of service in the switched network. In the example dynamic transport restoration is implemented by centralized automatic control of transport cross-connect devices to quickly restore service following a transport failure, such as caused by a cable cut. Over the duration of this event, more than 12,000 calls were blocked in the switched network, almost all of them originating or terminating at the Nashville node, and it is noteworthy that the blocking in the network returned to zero after the 4.736 Gbps of traffic trunk capacity was restored in the first 11 minutes, even though there was still 3.84 Gbps of traffic trunk capacity still out of service.

Dynamic traffic routing was able to find paths on which to complete traffic even though there was far less logical link capacity than normal even after the dynamic transport restoration. Hence dynamic traffic routing in combination with dynamic transport restoration provides a self-healing network capability, and even though the cable was repaired two hours after the cable cut, degradation of service was minimal. In this example, dynamic traffic routing also provided priority routing for selected customers and services, as described in ANNEX 3, which permits priority calls to be routed in preference to other calls, and blocking of the priority services is
essentially zero throughout the whole event.

Over the duration of an event, calls are blocked until sufficient capacity is restored for the network to return to zero blocking. That is, both dynamic transport routing and dynamic traffic routing are able to find available paths on which to restore the failed traffic. Hence, this example clearly illustrates how real-time dynamic traffic routing in combination with real-time dynamic transport routing can provide a self-healing network capability, and even if the cable is repaired two hours after the cut, degradation of service is minimal. This improved network performance provides additional service revenues as formerly blocked calls are completed, and it improves service quality to the customer.

These examples illustrate that implementation of dynamic transport routing provides better network performance at reduced cost. These benefits are similar to those achieved by dynamic traffic routing, and, as shown, the combination of dynamic traffic and transport routing provides synergistic reinforcement to achieve these network improvements.

The implementation of a dynamic transport routing network allows significant reductions in capital costs and network management and design expense with rearrangeable transport capacity design methods. Automated logical link provisioning and rearrangement lead to annual operations expense savings. Other network management and design impacts, leading to additional reduction in operations expense, are to simplify logical link provisioning systems; automate preservice logical-link testing and simplify maintenance systems; integrate logical-link capacity forecasting, administration, and bandwidth allocation into capacity planning and delivery; simplify node and transport planning; and automate inventory tracking.

5.4 Reliable Transport Network Design

In the event of link, node, or other network failure, the network design needs to provide sufficient surviving capacity to meet the required performance levels. For example, if a major fiber link fails, it could have a catastrophic effect on the network because traffic for many node pairs could not use the failed link. Similarly, if one of the nodes fails, it could isolate a whole geographic area until the node is restored to service.

With these two kinds of major failures in mind, we present here reliable transport routing models to achieve reliable network design, so as to provide service for predefined restoration objectives for any transport link or node failure in the network and continue to provide connections to customers with essentially no perceived interruption of service. This approach tries to integrate capabilities in both the traffic and transport networks to make the network robust or insensitive to failure. The basic aims of these models are to provide link diversity and protective capacity augmentation where needed so that specific "network robustness" objectives, such as traffic restoration level objectives, are met under failure events. This means that the network is designed so that it carries at least the fraction of traffic known as the traffic restoration level (TRL) under the failure event. For example, a traffic restoration level objective of 70-percent means that under any single transport link failure in the transport network, at least 70-percent of the original traffic for any affected node pair is still carried after the failure; for the unaffected node pairs, the traffic is carried at the normal blocking probability grade-of-service objective. These design models provide designs that address the network response immediately after a network event. It is also desirable to have transport restoration respond after the occurrence of the network
event to bring service back to normal. Transport restoration is also addressed in this ANNEX.

Reliable network performance objectives may require, for example, the network to carry 50-percent of its busy-hour load on each link within five minutes after a major network failure, in order to eliminate isolations among node pairs. Such performance may be provided through traffic restoration techniques, which include link diversity, traffic restoration capacity, and dynamic traffic routing. Reliable network performance objectives might also require a further reduction of connection setup blocking level to less than 5-percent within, say, 30 minutes to limit the duration of degraded service. This is possible through transport restoration methods that utilize transport nodes along with centralized transport restoration control. A further objective may be to restore at least 50-percent of severed trunks in affected links within this time period.

The transport restoration process restores capacity for switched, as well as dedicated ("private-line"), services in the event of link failures. In one implementation, transport restoration is conducted via a centralized system that restores the affected transport capacity until all available restoration capacity is exhausted. Optimization of the total cost of transport restoration capacity is possible through a design that increases sharing opportunities of the restoration capacity among different failure scenarios. Real-time transport restoration may also require the use of dedicated restoration capacity for each link and, thus, a lesser opportunity for sharing the restoration capacity. For the purpose of this analysis, we assume that all network transport may be protected with an objective level of restoration capacity. Transport restoration level (TPRL) is the term used to specify the minimal percentage of capacity on each transport link that is restorable. A transport restoration level is implemented in the model by restoring each affected link in a failure to a specified level. Here, we further distinguish between transport restoration level for switched circuits and dedicated circuits and designate them by TPRLs and TPRLp, respectively.

We now describe logical transport routing design models for survivable networks. Before we describe the models, we discuss the distinction between the traffic and transport networks and the concept of link diversity.

To distinguish between the traffic and transport networks, consider the example of a three-node network in Figure 5.2. The physical transport network is depicted at the bottom of the figure, and the corresponding logical transport (traffic) network at the top. For example, the direct logical link for connecting nodes A and B may ride the path A--C--D--B in the physical transport network. There is not a logical link between nodes B and C, which means there is no direct traffic trunk capacity from node B to C. A single physical transport link failure may affect more than one logical link. For example, in Figure 5.2, the failure of the physical transport link C--D affects logical links A--D, C--D, and A--B. Logical link diversity refers to a logical link design in which direct capacity for a logical link is split on two or more different physical transport paths. For example, in Figure 5.1, the direct logical link capacity for the link A--D may be split on two physical transport paths A--B--C--D and a physically diverse path A--E--F--D. A link diversity policy, say, of 70/30 corresponds to the fact that no more than 70-percent of the direct logical link (traffic trunk) capacity is routed on a single physical transport link for the different transport paths for that logical link. The advantage of logical link diversity is that if a physical transport link fails, the traffic for a particular node pair can still use the direct logical link.
capacity that survived on the physical transport path not on the failed link.

We now present models for transport routing design for physical transport link failure and node failure.

5.4.1 Transport Link Design Models

We assume that we have two distinct transport paths for direct logical link capacity for each node pair. In the model traffic demand is converted to virtual trunk demand such as based on direct and overflow logical transport capacity, such as illustrated in ANNEX 6 (see Figure 6.1). Let v be the virtual trunk requirement for the traffic demand for a particular node pair. Let d be the virtual trunk capacity to be put on the primary physical transport path and s be the virtual trunk capacity to be put on the alternate physical transport paths for the direct logical link of the given node pair. Let b be the number of trunks for this traffic link that are designed by the network design model. Let t be the traffic restoration level (TRL) objective under a link failure scenario. Let delta-b be the link capacity augmentation that may be needed for this logical link.

What we would like in a failure event is to carry a portion tv of the total virtual trunk demand for the affected node pairs. Thus, if tv delta-b/2, we set delta-b = 0 (no augmentation) with d = b - tv and s = tv. In this way, if either transport path fails, we can carry at least tv of the virtual trunk demand. On the other hand, if tv > b/2, then we want

\[(b + \text{delta-b})/2 = tv\]

which implies

\[\text{delta-b} = 2tv - b\]

In this case, we set d = s = (b + delta-b)/2. The above procedure is repeated for every demand pair in the network. The incremental cost of the network is the cost of trunk augmentation and routing the direct logical link capacity for each node pair, if any, on two transport paths.

The above procedure can be extended to the general case of k distinct physical transport paths. So, for k distinct physical transport paths, if

\[tv <= (k - 1)b/k\]

then delta-b = 0 with d = b - tv on the first physical transport path, and tv/(k - 1) on each of the other (k - 1) transport paths. If

\[tv > (k - 1)b/k\]

then

\[\text{delta-b} = ktv - (k - 1)/b\]

with each of the k transport paths having (b + delta-b)/k trunks.

A transport link model is illustrated in Figure 5.11, where each transport cross section is assumed on the average to contain certain fractions of switched (N) and dedicated (M) circuits. A portion of the switched and dedicated circuits is further presumed to be restorable in real time, such as with ring or dual-feed transmission arrangements (lowercase n and m...
Circuits that are not restored in real time are restored with transport restoration to a specified transport restoration level (TPRL) value. The lower part of Figure 5.11 demonstrates the interaction between the switched and dedicated capacity in the restoration process. The restoration process is assumed to restore the first unit of transport capacity (e.g., OC3) after x seconds, with y seconds to restore each additional unit of transport capacity. The restoration times are illustrative and are not critical to the reliable network design principles being discussed. SONET ring restoration can occur in about 50–200 milliseconds, and such real-time restoration is included in the model. A prioritization method is assumed, whereby transport links that carry higher-priority dedicated services are restored first. Because switched and dedicated capacity is often mixed at the logical link level, some switched capacity is also restored.

Different levels of transport restoration may also be assigned for the dedicated (TPRLp) and switched (TPRLs) networks. Each type of circuit demand is then restored to the corresponding level of restoration. Figure 14.3 also shows how the restoration level for switched circuits varies as a function of time. Some level of traffic circuits is restored in real time (n). After x seconds, transport restoration is initiated with one unit of transport capacity being restored in each y seconds, and with a smaller fraction of each transport capacity unit being switched traffic. The switched portion in each transport capacity unit subsequently increases to a larger fraction after dedicated traffic is restored to its designated level TPRLp. Transport restoration stops after both the TPRLp and TPRLs objectives are met.

5.4.2 Node Design Models

Node failure restoration design incorporates the concept of dual homing, as discussed in ANNEX 2, along with multiple ingress/egress routing. With single homing, the traffic from a particular geographical area normally goes to the single node nearest to it in order to carry the ingress and egress traffic. For example, in Figure 5.12 the traffic from the served area A1 enters the network through node B, and, similarly, the served area A2 traffic enters the network through node C. Now, if node B fails, then area A1 gets isolated. To protect against such an event, areas A1 and A2 are homed to more than one node—-to nodes A and B in this case (Figure 5.12, bottom). This is the concept of dual homing, in which we address the issue of designing a reliable network when one of the nodes may fail.

For every node to be protected, we assign a dual-homed node. Before failure, we assume that any load from a node to the protected node and its dual-homed node is equally divided; that is, if the original load between area A1 and node A is a1, and between A and the dual-homed node, B, is a2, then we assume that under normal network conditions, the load between nodes A and B is (a1 + a2)/2, and the same for the load between nodes A and C. We refer to this concept as balanced load. Then, under a failure event such as a node B failure, we carry load equal to (a1 + a2)/2 between nodes A and C. (See Figure 5.12, bottom.) We call this design objective a 50-percent traffic restoration level objective in a manner very analogous to the link failure event. As we can see from the bottom of Figure 14.2, this restoration level of traffic from or to area A1 is then still carried.

In the restoration model the node pairs that are going to be considered for the node failure design scenarios are determined. Next, the dual-homing nodes are determined for each node and the balanced-load traffic routed
accordingly. Then the vk virtual trunks are computed for the balanced loads. For each node pair, one of the nodes is assumed to fail (say, node B). Then this node cannot have any incoming or outgoing traffic and, also, cannot be a via node for any two-link traffic between other node pairs. Using these constraints, we solve a linear programming model that minimizes the incremental augmentation cost. Then, we reverse the roles of the nodes for this pair and solve the linear programming model again with the above-mentioned constraints. This design process is repeated for every pair of candidate nodes for each node failure scenario.

5.5 Modeling of Traffic Engineering Methods

In this Section we give modeling results for dynamic transport routing capacity design, performance under network failure, and performance under various network overload scenarios.

5.5.1 Dynamic Transport Routing Capacity Design

Design for traffic loads with week-to-week traffic variation. Dynamic transport routing network design allows more efficient use of node capacity and transport capacity and can lead to a reduction of network reserve trunk capacity by about 10 percent, while improving network performance. Table 5.1 illustrates a comparative forecast of a national intercity network’s normalized logical-link capacity requirements for the base case without dynamic transport routing and the network requirements with dynamic transport routing network design. When week-to-week traffic variations, which reflect seasonal variations, are taken into account, as in this analysis, the dynamic transport routing design can provide a reduction in network reserve capacity. As shown in the table, the traffic trunk savings always exceed 10 percent, which translates into a significant reduction in capital expenditures.

Table 5.1
Dynamic Transport Routing Capacity Savings with Week-to-Week Seasonal Traffic Variations (normalized capacity)

Dynamic transport routing network design for transport capacity achieves higher fiber link fill rates, which further reduces transport costs. The dynamic transport routing network implements automated inter-BR and access logical-link diversity, logical-link restoration, and node backup restoration to enhance the network survivability over a wide range of network failure conditions. We now illustrate dynamic transport routing network performance under design for normal traffic loads, fiber transport failure events, unpredictable traffic load patterns, and peak-day traffic load patterns.

5.5.2 Performance for Network Failures

Simulations are performed for the fixed transport and dynamic transport network performance for a fiber cut in Newark, New Jersey, in which approximately 8.96 Gbps of traffic trunk capacity was lost. The results are shown in Table 5.2. Here, a threshold of 50 percent or more node-pair blocking is used to identify node pairs that are essentially isolated; hence, the rearrangeable transport network design eliminates all isolations during this network failure event.

Table 5.2
Network Performance for Fiber Cut in Newark, NJ
An analysis is also performed for the network performance after transport restoration, in which the fixed and dynamic transport network designs are simulated after 29 percent of the lost trunks are restored. The results are shown in Table 5.3. Again, the dynamic transport network design eliminates all network isolations, some of which still exist in the base network after traffic trunk restoration.

Table 5.3
Network Performance for Fiber Cut in Newark, NJ (after Logical-link Restoration)

>From this analysis we conclude that the combination of dynamic traffic routing, logical-link diversity design, and transport restoration provides synergistic network survivability benefits. Dynamic transport network design automates and maintains logical-link diversity, as well as access network diversity in an efficient manner, and provides automatic transport restoration after failure.

A network reliability example is given for dual-homing transport demands on various OXC transport nodes. In one example, an OXC failure at the Littleton, MA node, in the model illustrated in Figure 5.1 is analyzed, and results given in Table 5.4. Because transport demands are diversely routed between nodes and dual-homed between access nodes and OXC devices, this provides additional network robustness and resilience to traffic node and transport node failures. When the network is designed for load balancing between access and internode demands, and traffic trunk restoration is performed, the performance of the dynamic transport routing network is further improved.

Table 5.4
Dynamic Transport Network Performance under OXC Failure

Figure 5.13 illustrates a typical instance of network design with traffic restoration level objectives and transport restoration level objectives, as compared with the base network with no TRL or TPRL design objectives. In the example, a fiber link failure occurs in the model network f seconds after the beginning of the simulation, severing a large amount of transport capacity in the network and cutting off thousands of existing connections. Therefore, in the simulation results shown in Figure 5.13, we see a large jump in the blocking at the instant of the cut. A transient flurry of reattempts follows as cut-off customers redial and reestablish their calls. This call restoration process is aided by the traffic restoration level design, which provides link diversity and protective transport capacity to meet the TRL objectives immediately following a failure. This TRL design, together with the ability of dynamic traffic routing to find surviving capacity wherever it exists, quickly reduces the transient blocking level, which then remains roughly constant for about x seconds until the transport restoration process begins. At x seconds after the link failure, the transport restoration process begins to restore capacity that was lost due to the failure. Blocking then continues to drop during that period when transport restoration takes place until it reaches essentially a level of zero blocking. Figure 5.13 illustrates the comparison between network performances with and without the traffic and transport restoration design techniques presented in this ANNEX.

This traffic restoration level design allows for varying levels of diversity on different links to ensure a minimum level of performance. Robust routing techniques such as dynamic traffic routing, multiple ingress/egress routing, and logical link diversity routing further improve response to node or
transport failures. Transport restoration is necessary to reduce network blocking to low levels. Given, for example, a 50 percent traffic restoration level design, it is observed that this combined with transport restoration of 50 percent of the failed transport capacity in affected links is sufficient to restore the traffic to low blocking levels. Therefore, the combination of traffic restoration level design and transport restoration level design is seen both to be cost-effective and to provide fast and reliable performance. The traffic restoration level design eliminates isolations between node pairs, and transport restoration level is used to reduce the duration of poor service in the network. Traffic restoration techniques combined with transport restoration techniques provide the network with independent means to achieve reliability against multiple failures and other unexpected events and are perceived to be a valuable part of a reliable network design.

5.5.3 Performance for General Traffic Overloads

The national network model is designed for dynamic transport routing with normal engineered traffic loads using the discrete event flow optimization (DEFO) model described in ANNEX 6, and it results in a 15 percent savings in reserve trunk capacity over the fixed transport routing model. In addition to this large savings in network capacity, the network performance under a 10 percent overload results in the performance comparison illustrated in Table 5.5. Hence, dynamic transport routing network designs achieve significant capital savings while also achieving superior network performance.

Table 5.5
Network Performance for 10% Traffic Overload

5.5.4 Performance for Unexpected Overloads

Dynamic transport routing network design provides load balancing of node traffic load and logical-link capacity so that sufficient reserve capacity is provided throughout the network to meet unexpected demands on the network. The advantage of such design is illustrated in Table 5.6, which compares the simulated network blocking for the fixed transport routing network design and dynamic transport routing network design during an hurricane-caused focused traffic overload in the northeastern United States. Such unexpected focused overloads are not unusual in a switched network, and the additional robustness provided by dynamic transport routing network design to the unexpected traffic overload patterns is clear from these results.

Table 5.6
Network Performance for Unexpected Traffic Overload
(focused overload in Northeastern US caused by hurricane)

Another illustration of the benefits of load balancing is given in Figure 5.14, in which a 25-percent traffic overload is focused on a node in Jackson, Mississippi.

Because the dynamic transport network is load balanced between access demands and inter-BR demands, this provides additional network robustness and resilience to unexpected traffic overloads, even though the dynamic transport routing network in this model has more than 15 percent less capacity than the fixed transport routing network. In this example, blocking-triggered rearrangement is allowed in the dynamic transport network. That is, as soon as node-pair blocking is detected, additional
logical-link capacity is added to the affected links by cross-connecting spare node-termination capacity and spare logical-link capacity, which has been freed up as a result of the more efficient dynamic transport network design. As can be seen from the figure, this greatly improves the network response to the overload.

5.5.5 Performance for Peak-Day Traffic Loads

A dynamic transport network design is performed for the Christmas traffic loads, and simulations performed for the base network and rearrangeable transport network design for the Christmas Day traffic. Results for the inter-BR blocking are summarized in Table 5.7. Clearly, the rearrangeable transport network design eliminates the inter-BR network blocking, although the access node to BR blocking may still exist but is not quantified in the model. In addition to increased revenue, customer perception of network quality is also improved for these peak-day situations.

Table 5.7
Network Performance for Christmas Day Traffic Overload

5.6 Conclusions/Recommendations

In this ANNEX, we present and analyze dynamic transport network architectures. Dynamic transport routing is a routing and bandwidth allocation method, which combines dynamic traffic routing with dynamic transport routing and for which we provide associated network design methods. We find that networks benefit more in efficiency and performance as the ability to reassign transport bandwidth is increased, and can simplify network management and design. We present results of a number of analysis, design, and simulation studies related to dynamic transport network architectures.

Models are used to measure the performance of the network for dynamic transport routing network design in comparison with the fixed transport network design, under a variety of network conditions including normal daily load patterns, unpredictable traffic load patterns such as caused by a hurricane, known traffic overload patterns such as occur on Christmas Day, and a network failure conditions such as a large fiber cut.

The conclusions/recommendations reached in this ANNEX are as follows:

* Dynamic transport routing is recommended and provides greater network throughput and, consequently, enhanced revenue, and at the same time capital savings should result, as discussed in ANNEX 6.
  a. Dynamic transport routing network design enhances network performance under failure, which arises from automatic inter-backbone-router and access logical-link diversity in combination with the dynamic traffic routing and transport restoration of logical links.
  b. Dynamic transport routing network design is recommended and improves network performance in comparison with fixed transport routing for all network conditions simulated, which include abnormal and unpredictable traffic load patterns.

* Traffic and transport restoration level design is recommended and allows for link diversity to ensure a minimum level of performance under failure.
Robust routing techniques are recommended, which include dynamic traffic routing, multiple ingress/egress routing, and logical link diversity routing; these methods improve response to node or transport failures.

Figures:

Figure 5.1 Transport Network Model
Figure 5.2 Logical (Layer 2) & Physical (Layer 1) Transport Networks
Figure 5.3 Dynamic Transport (Layer 2) Routing & Dynamic Connection (Layer 3) Routing
Figure 5.4 Dynamic Transport Routing Network
Figure 5.5 Dynamic Transport Routing vs. Fixed Transport Routing
Figure 5.6 Dynamic Transport Routing Network Weekly Arrangement (Week 1 Load Pattern)
Figure 5.7 Dynamic Transport Routing Network Weekly Arrangement (Week 2 Load Pattern)
Figure 5.8 Dynamic Transport Routing Peak Day Design
Figure 5.9 Dynamic Transport Routing Peak Day Design
Figure 5.10 Fiber Cut Example with Dynamic Traffic Routing & Dynamic Transport Routing
Figure 5.11 Transport Restoration Model
Figure 5.12 Illustration of Dual Homing
Figure 5.13 Network Performance for Link Failure with Traffic & Transport Restoration Design
Figure 5.14 Dynamic Transport Routing Performance for 25% Overload on Jackson, Mississippi, Node

ANNEX 6
Capacity Management Methods

Traffic Engineering & QoS Methods for IP-, ATM-, & TDM-Based Multiservice Networks

6.1 Introduction

In this ANNEX we discuss capacity management principles, as follows:
a) Link Capacity Design Models. These models find the optimum tradeoff between traffic carried on a shortest network path (perhaps a direct link) versus traffic carried on alternate network paths.
b) Shortest Path Selection Models. These models enable the determination of shortest paths in order to provide a more efficient and flexible routing plan.
c) Multihour Network Design Models. Three models are described including i) discrete event flow optimization (DEFO) models, ii) traffic load flow optimization (TLFO) models, and iii) virtual trunking flow optimization (VTFO) models.
d) Day-to-day Load Variation Design Models. These models describe techniques for handling day-to-day variations in capacity design.
e) Forecast Uncertainty/Reserve Capacity Design Models. These models describe the means for accounting for errors in projecting design traffic loads in the capacity design of the network.
6.2 Link Capacity Design Models

As illustrated in Figure 6.1, link capacity design requires a tradeoff of the traffic load carried on the link and traffic that must route on alternate paths.

High link occupancy implies more efficient capacity utilization, however high occupancy leads to link congestion and the resulting need for some traffic not to be routed on the direct link but on alternate paths. Alternate paths may entail longer, less efficient paths. A good balance can be struck between link capacity design and alternate path utilization. For example, consider Figure 6.1, which illustrates a network where traffic is offered on link A-B connecting node A and node B.

Some of the traffic can be carried on link A-B, however when the capacity of link A-B is exceeded, some of the traffic must be carried on alternate paths or be lost. The objective is to determine the direct A-B link capacity and alternate routing path flow such that all the traffic is carried at minimum cost. A simple optimization procedure is used to determine the best proportion of traffic to carry on the direct A-B link and how much traffic to alternate route to other paths in the network. As the direct link capacity is increased, the direct link cost increases while the alternate path cost decreases as more direct capacity is added, because the overflow load decreases and therefore the cost of carrying the overflow load decreases. An optimum, or minimum, cost condition is achieved when the direct A-B link capacity is increased to the point where the cost per incremental unit of bandwidth capacity to carry traffic on the direct link is just equal to the cost per unit of bandwidth capacity to carry traffic on the alternate network. This is a design principle used in many design models, be they sparse or meshed networks, fixed hierarchical routing networks or dynamic nonhierarchical routing networks.

6.3 Shortest Path Selection Models

Some routing methods such as hierarchical routing, limits path choices and provides inefficient design. This limits flexibility and reduces efficiency. If we choose paths based on cost and relax constraints such as a hierarchical network structure, a more efficient network results. Additional benefits can be provided in network design by allowing a more flexible routing plan that is not restricted to hierarchical routes but allows the selection of the shortest nonhierarchical paths. Dijkstra’s method \[\text{Dij}59\], for example, is often used for shortest path selection. Figure 6.2 illustrates the selection of shortest paths between two network nodes, SNDG and BRHM.

Longer paths, such as SNDG-SNBO-ATLN-BRHM, which might arise through hierarchical path selection, are less efficient than shortest path selection, such as SNDG-PHNX-BRHM, SNDR-TCSN-BRHM, or SNDG-MTGM-BRHM. There are really two components to the shortest path selection savings. One component results from eliminating link splintering. Splintering occurs, for example, when more than one node is required to satisfy a traffic load within a given area, such as a metropolitan area. Multiple links to a distant node could result, thus dividing the load among links which are less efficient than a single large link. A second component of shortest path selection savings arises from path cost. Routing on the least costly, most direct, or shortest paths is often more efficient than routing over longer hierarchical paths.

6.4 Multihour Network Design Models
Dynamic routing design improves network utilization relative to fixed routing design because fixed routing cannot respond as efficiently to traffic load variations that arise from business/residential phone use, time zones, seasonal variations, and other causes. Dynamic routing design increases network utilization efficiency by varying routing tables in accordance with traffic patterns and designing capacity accordingly. A simple illustration of this principle is shown in Figure 6.3, where there is afternoon peak load demand between nodes A and B but a morning peak load demand between nodes A and C and nodes C and B.

Here a simple dynamic route design is to provide capacity only between nodes A and C and nodes C and B but no capacity between nodes A and B. Then the A--C and C--B morning peak loads route directly over this capacity in the morning, and the A--B afternoon peak load uses this same capacity by routing this traffic on the A--C--B path in the afternoon. A fixed routing network design provides capacity for the peak period for each node pair and thus provides capacity between nodes A and B, as well as between nodes A and C and nodes C and B.

The effect of multihour network design is illustrated by a national intercity network design model illustrated in Figure 6.4. Here it is shown that about 20 percent of the network's first cost can be attributed to designing for time-varying loads.

As illustrated in the figure, the 17 hourly networks are obtained by using each hourly load, and ignoring the other hourly loads, to size a network that perfectly matches that hour's load. Each hourly network represents the hourly traffic load capacity cost referred to in Table 1.1 in ANNEX 1. The 17 hourly networks show that three network busy periods are visible, where we see morning, afternoon, and evening busy periods, and the noon-hour drop in load and the early-evening drop as the business day ends and residential calling begins in the evening. The hourly network curve separates the capacity provided in the multihour network design into two components: Below the curve is the capacity needed in each hour to meet the load; above the curve is the capacity that is available but is not needed in that hour. This additional capacity exceeds 20 percent of the total network capacity through all hours of the day, which represents the multihour capacity cost referred to in Table 1.1. This gap represents the capacity of the network to meet noncoincident loads.

We now discuss the three types of multihour network design models—discrete event flow optimization models, virtual trunking flow optimization models, and traffic flow optimization models—and illustrate how they are applied to various fixed and dynamic network designs. For each model we discuss steps that include initialization, routing design, capacity design, and parameter update.

### 6.4.1 Discrete Event Flow Optimization (DEFO) Models

Discrete event flow optimization (DEFO) models are used for fixed and dynamic traffic network design. These models optimize the routing of discrete event flows, as measured in units of individual connection requests, and the associated link capacities. Figure 6.5 illustrates steps of the DEFO model.

The event generator converts traffic demands to discrete connection-request events. The discrete event model provides routing logic according to the particular routing method and routes the connection-request events according
to the routing table logic. DEFO models use simulation models for path selection and routing table management to route discrete-event demands on the link capacities, and the link capacities are then optimized to meet the required flow. We generate initial link capacity requirements based on the traffic load matrix input to the model. Based on design experience with the model, an initial node-termination capacity is estimated based on a maximum design occupancy in the node busy hour of 0.93, and the total network occupancy (total traffic demand/total link capacity) in the network busy hour is adjusted to fall within the range of 0.84 to 0.89. Network performance is evaluated as an output of the discrete event model, and any needed link capacity adjustments are determined. Capacity is allocated to individual links in accordance with the Kruithof allocation method [Kru37], which distributes link capacity in proportion to the overall demand between nodes.

Kruithof’s technique is used to estimate the node-to-node requirements pij from the originating node i to the terminating node j under the condition that the total node link capacity requirements may be established by adding the entries in the matrix p = [pij]. Assume that a matrix q = [qij], representing the node-to-node link capacity requirements for a previous iteration, is known. Also, the total link capacity requirements bi at each node i and the total link capacity requirements dj at each node j are estimated as follows:

\[ b_i = \frac{a_i}{\gamma} \]
\[ d_j = \frac{a_j}{\gamma} \]

where ai is the total traffic at node i, aj is the total traffic at node j, and gamma is the average traffic-carrying capacity per trunk, or node design occupancy, as given previously. The terms pij can be obtained as follows:

\[ \text{faci} = \frac{b_i}{\text{sum-j qij}} \]
\[ \text{facj} = \frac{d_j}{\text{sum-i qij}} \]
\[ E_{ij} = \frac{\text{faci} + \text{facj}}{2} \]
\[ pij = qij \times E_{ij} \]

After the above equations are solved iteratively, the converged steady state values of pij are obtained.

The DEFO model can generate connection-request events according to a Poisson arrival distribution and exponential holding times, or with more general arrival streams and arbitrary holding time distributions, because such models can readily be implemented in the discrete routing table simulation model. Connection-request events are generated in accordance with the traffic load matrix input to the model. These events are routed on the selected path according to the routing table rules, as modeled by the routing table simulation, which determines the selected path for each call event and flows the event onto the network capacity.

The output from the routing design is the fraction of traffic lost and delayed in each time period. From this traffic performance, the capacity design determines the new link capacity requirements of each node and each link to meet the design performance level. From the estimate of lost and delayed traffic at each node in each time period, an occupancy calculation determines additional node link capacity requirements for an updated link
capacity estimate. Such a link capacity determination is made based on the amount of blocked traffic. The total blocked traffic \( \Delta a \) is estimated at each of the nodes, and an estimated link capacity increase \( \Delta T \) for each node is calculated by the relationship

\[
\Delta T = \frac{\Delta a}{\gamma}
\]

where again \( \gamma \) is the average traffic-carrying capacity per trunk. Thus, the \( T \) for each node is distributed to each link according to the Kruthof estimation method described above. The Kruthof allocation method \([Kru37]\) distributes link capacity in proportion to the overall demand between nodes and in accordance with link cost, so that overall network cost is minimized. Sizing individual links in this way ensures an efficient level of utilization on each link in the network to optimally divide the load between the direct link and the overflow network. Once the links have been resized, the network is re-evaluated to see if the performance objectives are met, and if not, another iteration of the model is performed.

We evaluate in the model the confidence interval of the engineered blocking/delay. For this analysis, we evaluate the binomial distribution for the 90th percentile confidence interval. Suppose that for a traffic load of \( A \) in which calls arrive over the designated time period of stationary traffic behavior, there are on average \( m \) blocked calls out of \( n \) attempts. This means that there is an average observed blocking/delay probability of \( p_1 = \frac{m}{n} \).

\[
p_1 = \frac{m}{n}
\]

where, for example, \( p_1 = .01 \) for a 1 percent average blocking/delay probability. Now, we want to find the value of the 90th percentile blocking/delay probability \( p \) such that

\[
\sum_{r=m}^{n} \binom{n}{r} p^r (1-p)^{n-r} \geq .90
\]

where

\[
\binom{n}{r} = \frac{n!}{(n-r)!r!}
\]

is the binomial coefficient, and

\[
q = 1-p
\]

Then the value \( p \) represents the 90th percentile blocking/delay probability confidence interval. That is, there is a 90 percent chance that the observed blocking/delay will be less than or equal to the value \( p \). Methods given in \([Wei63]\) are used to numerically evaluate the above expressions.

As an example application of the above method to the DEFO model, suppose that network traffic is such that 1 million calls arrive in a single busy-hour period, and we wish to design the network to achieve 1 percent average blocking/delay or less. If the network is designed in the DEFO model to yield at most .00995 probability of blocking/delay---that is, at most 9,950 calls are blocked out of 1 million calls in the DEFO model---then we can be more than 90 percent sure that the network has a maximum blocking/delay probability of .01. For a specific switch pair where 2,000 calls arrive in a single busy-hour period, suppose we wish to design the switch pair to achieve 1 percent average blocking/delay probability or less. If the network capacity is designed in the DEFO model to yield at most .0075 probability of blocking/delay for the switch pair---that is, at most 15 calls are blocked out of 2,000 calls in the DEFO model---then we can be more
than 90 percent sure that the switch pair has a maximum blocking/delay probability of .01. These methods are used to ensure that the blocking/delay probability design objectives are met, taking into consideration the sampling errors of the discrete event model.

The greatest advantage of the DEFO model is its ability to capture very complex routing behavior through the equivalent of a simulation model provided in software in the routing design module. By this means, very complex routing networks have been designed by the model, which include all of the routing methods discussed in ANNEX 2, TDR, SDR, and EDR methods, and the multiservice QoS resource allocation models discussed in ANNEX 3. Complex traffic processes, such as self-similar traffic, can also be modeled with DEFO methods. A flow diagram of the DEFO model, in which DC-SDR logical blocks described in ANNEX 2 are implemented, is illustrated in Figure 6.6. The DEFO model is general enough to include all TE models yet to be determined.

6.4.2 Traffic Load Flow Optimization (TLFO) Models

Traffic load flow optimization (TLFO) models are used for fixed and dynamic traffic network design. These models optimize the routing of traffic flows and the associated link capacities. Such models typically solve mathematical equations that describe the routing of traffic flows analytically and, for dynamic network design, often solve linear programming flow optimization models. Various types of traffic flow optimization models are distinguished as to how flow is assigned to links, paths, and routes. In fixed network design, traffic flow is assigned to direct links and overflow from the direct links is routed to alternate paths through the network, as described above. In dynamic network design, traffic flow models are often path based, in which traffic flow is assigned to individual paths, or route based, in which traffic flow is assigned to routes.

As applied to fixed and dynamic routing networks, TLFO models do network design based on shortest path selection and linear programming traffic flow optimization. An illustrative traffic flow optimization model is illustrated in Figure 6.7.

There are two versions of this model: route-TLFO and path-TLFO models. Shortest least-cost path routing gives connections access to paths in order of cost, such that connections access all direct circuits between nodes prior to attempting more expensive overflow paths. Routes are constructed with specific path selection rules. For example, route-TLFO models construct routes for multilink or two-link path routing by assuming crankback and originating node control capabilities in the routing. The linear programming flow optimization model strives to share link capacity to the greatest extent possible with the variation of loads in the network. This is done by equalizing the loads on links throughout the busy periods on the network, such that each link is used to the maximum extent possible in all time periods. The routing design step finds the shortest paths between nodes in the network, combines them into candidate routes, and uses the linear programming flow optimization model to assign traffic flow to the candidate routes.

The capacity design step takes the routing design and solves a fixed-point traffic flow model to determine the capacity of each link in the network. This model determines the flow on each link and sizes the link to meet the performance level design objectives used in the routing design step. Once the links have been sized, the cost of the network is evaluated and compared to the last iteration. If the network cost is still decreasing, the update module (1) computes the slope of the capacity versus load curve on each
link, which reflects the incremental link cost, and updates the link "length" using this incremental cost as a weighting factor and (2) recomputes a new estimate of the optimal link overflow using the method described above. The new link lengths and overflow are fed to the routing design, which again constructs route choices from the shortest paths, and so on. Minimizing incremental network costs helps convert a nonlinear optimization problem to a linear programming optimization problem. Yaged [Yag71, Yag73] and Knepley [Kne73] take advantage of this approach in their network design models. This favors large efficient links, which carry traffic at higher utilization efficiency than smaller links. Selecting an efficient level of blocking/delay on each link in the network is basic to the route/path-TLFO model. The link overflow optimization model [Tru54] is used in the TLFO model to optimally divide the load between the direct link and the overflow network.

6.4.3 Virtual Trunking Flow Optimization (VTFO) Models

Virtual trunk flow optimization (VTFO) models are used for fixed and dynamic traffic and transport network design. These models optimize the routing of "virtual trunking (VT)" flows, as measured in units of VT bandwidth demands such as 1.5 mbps, OC1, OC12, etc. For application to network design, VTFO models use mathematical equations to convert traffic demands to VT capacity demands, and the VT flow is then routed and optimized. Figure 6.8 illustrates the VTFO steps. The VT model converts traffic demands directly to VT demands. This model typically assumes an underlying traffic routing structure.

A linear programming VT flow optimization model can be used for network design, in which hourly traffic demands are converted to hourly VT demands by using, for example, TLFO network design methods described above for each hourly traffic pattern. The linear programming VT flow optimization is then used to optimally route the hourly node-to-node VT demands on the shortest, least-cost paths and size the links to satisfy all the VT demands. Alternatively, node-to-node traffic demands are converted to node-to-node VT demands by using the approach described above to optimally divide the traffic load between the direct link and the overflow network, but in this application of the model we obtain an equivalent VT demand, by hour, as opposed to an optimum link-overflow objective.

6.5 Day-to-day Load Variation Design Models

In network design we use the forecast traffic loads, which are actually mean loads about which there occurs a day-to-day variation, characterized, for example, by a gamma distribution with one of three levels of variance [Wil58]. Even if the forecast mean loads are correct, the actual realized loads exhibit a random fluctuation from day to day. Studies have established that this source of uncertainty requires the network to be augmented in order to maintain the required performance objectives. Accommodating day-to-day variations in the network design procedure can use an equivalent load technique that models each node pair in the network as an equivalent link designed to meet the performance objectives. On the basis of day-to-day variation design models, such as [Hin76, Wil58], the link bandwidth N required in the equivalent link to meet the required objectives for the forecasted load R with its specified instantaneous-to-mean ratio (IMR) and specified level of day-to-day variation phi is determined. Holding fixed the specified IMR value and the calculated bandwidth capacity N, we calculate what larger equivalent load Re requires bandwidth N to meet the performance objectives if the forecasted load had had no day-to-day variation. The equivalent traffic load Re is then used which in place of R,
since it produces the same equivalent bandwidth when designed for the same IMR-level but in the absence of day-to-day variation.

6.6 Forecast Uncertainty/Reserve Capacity Design Models

Network designs are made based on measured traffic loads and estimated traffic loads that are subject to error. In network design we use the forecast traffic loads because the network capacity must be in place before the loads occur. Errors in the forecast traffic reflect uncertainty about the actual loads that will occur, and as such the design needs to provide sufficient capacity to meet the expected load on the network in light of these expected errors. Studies have established that this source of uncertainty requires the network to be augmented in order to maintain the blocking/delay probability grade-of-service objectives [FHH79].

The capacity management process accommodates the random forecast errors in the procedures. When some realized node-to-node performance levels are not met, additional capacity and/or routing changes are provided to restore the network performance to the objective level. Capacity is often not disconnected in the capacity management process even when load forecast errors are such that this would be possible without performance degradation. Capacity management, then, is based on the forecast traffic loads and the link capacity already in place. Consideration of the in-service link capacity entails a transport routing policy that could consider (1) fixed transport routing, in which transport is not rearranged; and (2) dynamic transport routing, as discussed in ANNEX 5, which allows periodic transport rearrangement including some capacity disconnects.

The capacity disconnect policy may leave capacity in place even though it is not called for by the network design. In-place capacity that is in excess of the capacity required to exactly meet the design loads with the objective performance is called reserve capacity. There are economic and service implications of the capacity management strategy. Insufficient capacity means that occasionally link capacity must be connected on short notice if the network load requires it. This is short-term capacity management. There is a trade-off between reserve capacity and short-term capacity management. Reference [FHH79] analyzes a model that shows the level of reserve capacity to be in the range of 6–25 percent, when forecast error, measurement error, and other effects are present. In fixed transport routing networks, if links are found to be overloaded when actual loads are larger than forecasted values, additional link capacity is provided to restore the objective performance levels, and, as a result, the process leaves the network with reserve capacity even when the forecast error is unbiased. Operational studies in fixed transport routing networks have measured up to 20 percent and more for network reserve capacity. Methods such as the Kalman filter [PaW82], which provides more accurate traffic forecasts and rearrangeable transport routing, can help reduce this level of reserve capacity. On occasion, the planned design underprovides link capacity at some point in the network, again because of forecast errors, and short-term capacity management is required to correct these forecast errors and restore service.

The model illustrated in Figure 6.9 is used to study network design of a network on the basis of forecast loads, in which the network design accounts for both the current network and the forecast loads in capacity management. Capacity management can make short-term capacity additions if network performance for the realized traffic loads becomes unacceptable and cannot be corrected by routing adjustments. Capacity management tries to minimize reserve capacity while maintaining the design performance objectives and an acceptable level of short-term capacity additions. Capacity management uses
the traffic forecast, which is subject to error, and the existing network. The model assumes that the network design is always implemented, and, if necessary, short-term capacity additions are made to restore network performance when design objectives are not met.

With fixed traffic and transport routing, link capacity augments called for by the design model are implemented, and when the network design calls for fewer trunks on a link, a disconnect policy is invoked to decide whether trunks should be disconnected. This disconnect policy reflects a degree of reluctance to disconnect link capacity, so as to ensure that disconnected link capacity is not needed a short time later if traffic loads grow. With dynamic traffic routing and fixed transport routing reduction in reserve capacity is possible while retaining a low level of short-term capacity management. With dynamic traffic routing and dynamic transport routing, additional reduction in reserve capacity is achieved. With dynamic traffic routing and dynamic transport routing design, as illustrated in Figure 6.10, reserve capacity can be reduced in comparison with fixed transport routing, because with dynamic transport network design the link sizes can be matched to the network load.

6.7 Meshed, Sparse, and Dynamic-Transport Design Models

In the meshed network designs we assume an overlay network structure, such as for example MPLS traffic trunk formed by label switched paths (LSPs) or ATM virtual paths (VPs). Such LSPs are formed through use of label switched routers (LSRs) to establish the paths. VPs are formed through use of ATM switches, or perhaps might involve the use of ATM cross-connect device. In the meshed network case, traffic is aggregated to many logical links, and the links therefore need to have a bandwidth granularity below OC3 level. Such an overlay network cross-connecting capability is able to establish a mesh of layer-2 logical links, which are multiplexed onto the higher capacity fiber backbone links. With the highly connected mesh of logical links, 1- and 2-link routing methods, such as 2-link STT-EDR and 2-link DC-SDR, can be employed if VPs or LSPs can be used in tandem.

For the sparse network case, as illustrated in Figure 6.11, logical links are established by use of cross-connect switching, such as with optical cross connects (OXC), as discussed in ANNEX 5. In the sparse network case the traffic is aggregated to a fewer number of logical links, in which case the links have larger bandwidth granularity, OC3, OC12, OC48, and higher. For the dynamic transport network design, the traffic is aggregated to an even smaller number of fiber backbone links, and in that case the bandwidth granularity is larger, OC48, OC192, and larger, corresponding to a single wavelength on a DWDM fiber channel.

For design of the dynamic transport routing network, as described in ANNEX 5, the logical links are controlled dynamically within the OXC network by switching bandwidth on the fiber backbone links to the logical links. As a result, the design procedure for dynamic transport networks can be relatively simple. The traffic demands of the various node pairs are aggregated to the backbone fiber transport links, which overlay the logical links, and then each transport link is sized to carry the total traffic demand from all node pairs that use the backbone transport link for voice, data, and broadband traffic. As illustrated in Figure 6.12, one subtlety of the design procedure is deciding what performance objectives (e.g., blocking objective) to use for sizing the backbone transport links. The difficulty is that many node pairs send traffic over the same backbone transport link, and each of these node pairs has a different number of backbone transport links in its path. This means that for each traffic load,
a different level of performance (e.g., blocking) on a given backbone transport link is needed to ensure, say, a 1-percent level of blocking end to end. With many kinds of traffic present on the link, we are guaranteed an acceptable blocking probability grade-of-service objective if we identify the path through each transport link that requires the largest number of links, n, and size the link to a 1/n blocking objective. In Figure 6.12, link L1 has a largest number n equal to 6, and link L2 has a largest number n equal to 4. If the end-to-end blocking objective is 1-percent, then the link-blocking objectives are determined as given in the figure. We show that the dynamic transport routing network sized in this simple manner still achieves significant efficiencies.

6.8 Modeling of Traffic Engineering Methods

In this Section, we again use the full-scale national network model developed in ANNEX 2 to study various TE scenarios and tradeoffs. The 135-node national model is illustrated in Figure 2.6, the multiservice traffic demand model is summarized in Table 2.1, and the cost model is summarized in Table 2.2.

6.8.1 Per-Virtual-Network vs. Per-Flow Network Design

Here we illustrate the use of the DEFO model to design for a per-flow multiservice network design and a per-virtual-network design, and to provide comparisons of these designs. The per-flow and per-virtual network designs for the flat 135-node model are summarized in Table 6.1.

Table 6.1
Design Comparison of Per-Virtual-Network & Per-Flow Bandwidth Allocation
Multilink STT-EDR Connection Routing
Sparse Single-Area Flat Topology
(135-Node Multiservice Network Model; DEFO Design Model)

We see from the above results that the per-virtual network design compared to the per-flow design yields the following results:

* the per-flow design has 0.996 of the total termination capacity of the per-virtual-network design
* the per-flow design has 0.991 of the total transport capacity of the per-virtual-network design
* the per-flow design has 0.970 of the total network cost of the per-virtual-network design

These results indicate that the per-virtual network design and per-flow design are quite comparable in terms of capacity requirements and design cost. In ANNEX 3 we showed that the performance of these two designs was also quite comparable under a range of network scenarios.

6.8.2 Integrated vs. Separate Voice/ISDN & Data Network Designs

The comparative designs for separate and integrated network designs under multilink, STT-EDR, per-flow routing are given in Table 6.2 for the following cases:

* voice/ISDN-only traffic (VNETs 1-8 in Table 2.1)
* data-only traffic (VNETs 9-11 in Table 2.1)
* integrated voice/ISDN and data design (VNETs 1-11 in Table 2.1)

Table 6.2
Comparison of Voice/ISDN-Only Design (VNETs 1-8), Data-Only Design (VNETs
We see from the above results that the separate voice/ISDN and data designs compared to the integrated design yields the following results:

* the integrated design has 0.937 of the total termination capacity as the separate voice/ISDN & data designs
* the integrated design has 0.963 of the total transport capacity as the separate voice/ISDN & data designs
* the integrated design has 0.947 of the total cost as the separate voice/ISDN & data designs

These results indicate that the integrated design is somewhat more efficient in design owing to the economy-of-scale of the higher-capacity network elements, as reflected in the cost model given in Table 2.2.

The comparative designs for separate and integrated network designs under 2-link STT-EDR connection routing with per-flow QoS resource management are given in Table 6.3 for the following cases:

* voice/ISDN-only traffic (VNETs 1-8 in Table 2.1)
* data-only traffic (VNETs 9-11 in Table 2.1)
* integrated voice/ISDN and data design (VNETs 1-11 in Table 2.1)

We see from the above results that the separate voice/ISDN and data designs compared to the integrated design yields the following results:

* the integrated design has 0.948 of the total termination capacity as the separate voice/ISDN & data designs
* the integrated design has 0.956 of the total transport capacity as the separate voice/ISDN & data designs
* the integrated design has 0.804 of the total cost as the separate voice/ISDN & data designs

These results indicate that the integrated design is somewhat more efficient in design termination and transport capacity. It is about 20 percent more efficient in cost owing to the economy-of-scale of the higher-capacity network elements, as reflected in the cost model given in Table 2.2.

The comparative designs for separate and integrated network designs under 2-link DC-SDR connection routing with per-flow QoS resource management are given in Table 6.4 for the following cases:

* voice/ISDN-only traffic (VNETs 1-8 in Table 2.1)
* data-only traffic (VNETs 9-11 in Table 2.1)
* integrated voice/ISDN and data design (VNETs 1-11 in Table 2.1)
Table 6.4
Comparison of Voice/ISDN-Only Design (VNETs 1-8), Data-Only Design (VNETs 9-11), & Integrated Voice/ISDN & Data Design (VNETs 1-11)
2-Link DC-SDR Connection Routing; Per-Flow Bandwidth Allocation
Sparse Single-Area Flat Topology
(135-Node Multiservice Network Model; DEFO Design Model)

We see from the above results that the separate voice/ISDN and data designs compared to the integrated design yields the following results:

* the integrated design has 0.951 of the total termination capacity as the separate voice/ISDN & data designs
* the integrated design has 0.958 of the total transport capacity as the separate voice/ISDN & data designs
* the integrated design has 0.806 of the total cost as the separate voice/ISDN & data designs

These results indicate that the integrated design is somewhat more efficient in design termination and transport capacity. It is about 20 percent more efficient in cost owing to the economy-of-scale of the higher-capacity network elements, as reflected in the cost model given in Table 2.2.

6.8.3 Multilink vs. 2-Link Network Design

We see from the results in Tables 6.2 and 6.3 that the multilink EDR network design compared to 2-link EDR design yields the following results:

* the voice/ISDN-only multilink-EDR design has 0.735 of the total termination capacity of the 2-link design
* the voice/ISDN-only multilink-EDR design has 0.906 of the total transport capacity of the 2-link design
* the voice/ISDN-only multilink-EDR design has 0.792 of the total cost of the 2-link design

* the data-only multilink-EDR design has 0.631 of the total termination capacity of the 2-link design
* the data-only multilink-EDR design has 0.770 of the total transport capacity of the 2-link design
* the data-only multilink-EDR design has 0.897 of the total cost of the 2-link design

* the integrated multilink-EDR design has 0.640 of the total termination capacity of the 2-link design
* the integrated multilink-EDR design has 0.798 of the total transport capacity of the 2-link design
* the integrated multilink-EDR design has 1.023 of the total cost of the 2-link design

These results show that the multilink designs are generally more efficient than the 2-link designs in transport and termination capacity, and have lower cost for the separate designs and comparable cost for the integrated design.

6.8.4 Single-Area Flat vs. 2-Level Hierarchical Network Design

In Table 6.5 we illustrate the use of the DEFO model to design for a per-flow 2-level hierarchical multiservice network design and a 2-level hierarchical per-virtual-network design, and to provide comparisons of these
designs. Recall that the hierarchical model, illustrated in Figure 3.7, consists of 135-edge-nodes and 21 backbone-nodes. The edge-nodes are homed onto the backbone nodes in a hierarchical relationship. The per-flow and per-virtual network designs for the hierarchical 135-edge-node and 21-backbone-node model are summarized in Table 6.5.

Table 6.5
Design Comparison of Per-Virtual-Network & Per-Flow Bandwidth Allocation
Multilink STT-EDR Connection Routing
135-Edge-Node and 21-Backbone-Node Sparse Multi-Area 2-Level Hierarchical Topology
(Multiservice Network Model; DEFO Design Model)

We see from the above results that the hierarchical per-virtual network design compared to the hierarchical per-flow design yields the following results:

* the hierarchical per-flow design has 0.956 of the total termination capacity of the hierarchical per-virtual-network design
* the hierarchical per-flow design has 0.992 of the total transport capacity of the hierarchical per-virtual-network design
* the hierarchical per-flow design has 0.971 of the total network cost of the hierarchical per-virtual-network design

These results indicate that the hierarchical per-virtual-network design and hierarchical per-flow designs are quite comparable in terms of capacity requirements and design cost. In ANNEX 3 we showed that the performance of these two designs was also quite comparable under a range of network scenarios.

By comparing Tables 6.1 and 6.5, we can find the relative capacity of the single-area flat network design and the multiple-area, 2-level hierarchical network design (per-flow case):

* the single-area flat design has 0.776 of the total termination capacity of the multi-area 2-level hierarchical design
* the single-area flat design has 0.937 of the total transport capacity of the multi-area 2-level hierarchical design
* the single-area flat design has 1.154 of the total network cost of the multi-area 2-level hierarchical design

In this model the single-area flat designs have less total termination and transport capacity as the multi-area hierarchical designs, and are therefore more efficient in engineered capacity. However, the hierarchical designs appear to be less expensive than the flat designs. This is because of the larger percentage of OC48 links in the hierarchical designs, which is also considerably sparser than the flat design and therefore the traffic loads are concentrated onto fewer, larger, links. As discussed in ANNEX 2, there is an economy of scale built into the cost model which affords the higher capacity links (e.g., OC48 as compared to OC3) a considerably lower per-unit-of-bandwidth cost, and therefore a lower overall network cost is achieved as a consequence. However, the performance analysis results discussed in ANNEX 3 show that the flat designs perform better than the hierarchical designs under the overload and failure scenarios that were modeled. This also is a consequence of the sparser hierarchical network and lesser availability of alternate paths for more robust network performance.

6.8.5 EDR vs. SDR Network Design

Next we examine the meshed network designs for the 2-link STT-EDR network
and the 2-link DC-SDR network, which were discussed in Section 6.7. The designs for the 2-link STT-EDR and 2-link DC-SDR connection routing networks, with per-flow QoS resource management, are given in Table 6.6, which again are obtained using the DEFO model on the 135-node model.

Table 6.6
Design Comparison of 2-Link STT-EDR & 2-Link DC-SDR Connection Routing
Per-Flow Bandwidth Allocation
Meshed Single-Area Flat Topology
(135-Edge-Node Multiservice Network Model; DEFO Design Model)

We see from the above results that the EDR network design compared to the SDR design yields the following results:

* the EDR design has 0.999 of the total termination capacity of the SDR design
* the EDR design has 1.000 of the total transport capacity of the SDR design
* the EDR design has 0.999 of the total network cost of the SDR design

We note that the designs are very comparable to each other and have essentially the same total network design costs. This suggests that there is not a significant advantage for employing link-state information in these network designs, and given the high overhead in flooding link-state information, EDR methods are preferred.

6.8.6 Dynamic Transport Routing vs. Fixed Transport Routing Network Design

Finally we examine the design comparisons of dynamic transport routing compared with the fixed transport routing. In the model we assume multilink STT-EDR connection routing with per-flow QoS resource management, and once again use the DEFO design model for the flat 135-node network model. The results are summarized in Table 6.7.

Table 6.7
Design Comparison of Fixed Transport Routing & Dynamic Transport Routing
Multilink STT-EDR Connection Routing; Per-Flow Bandwidth Allocation
Sparse Single-Area Flat Topology
(135-Node Multiservice Network Model; DEFO Design Model)

We see from the above results that the fixed transport network design compared to the dynamic transport design yields the following results:

* the dynamic transport design has 1.097 of the total termination capacity of the fixed-transport-network design
* the dynamic transport design has 1.048 of the total transport capacity of the fixed-transport-network design
* the dynamic transport design has 0.516 of the total network cost of the fixed-transport-network design

These results indicate that the dynamic transport design has more termination capacity and transport capacity than the fixed transport network design, but substantially lower cost. The larger capacity comes about because of the larger fiber backbone link bandwidth granularity compared to the logical link granularity in the fixed transport routing case. The lower cost of the dynamic transport network comes about, however, because of the economies of scale of the higher capacity transport and termination elements, as reflected in Table 2.2. In ANNEX 3 we showed that the performance of these two designs was also quite comparable under a range of...
6.9 Conclusions/Recommendations

The conclusions/recommendations reached in this ANNEX are as follows:

* Discrete event flow optimization (DEFO) design models are recommended and are shown to be able to capture very complex routing behavior through the equivalent of a simulation model provided in software in the routing design module. By this means, very complex routing networks have been designed by the model, which include all of the routing methods discussed in ANNEX 2 (FR, TDR, SDR, and EDR methods) and the multiservice QoS resource allocation models discussed in ANNEX 3.

* Sparse topology options are recommended, such as the multilink STT-EDR/DC-SDR/DP-SDR options, which lead to capital cost advantages, and more importantly to operation simplicity and cost reduction. Capital cost savings are subject to the particular switching and transport cost assumptions. Operational issues are further detailed in ANNEX 7.

* Voice and data integration is recommended and
  a. can provide capital cost advantages, and
  b. more importantly can achieve operational simplicity and cost reduction, and
  c. if IP-telephony takes hold and a significant portion of voice calls use voice compression technology, this could lead to more efficient networks.

* Multilink routing methods are recommended and exhibit greater design efficiencies in comparison with 2-link routing methods. As discussed and modeled in ANNEX 3, multilink topologies exhibit better network performance under overloads in comparison with 2-link routing topologies; however the 2-link topologies do better under failure scenarios.

* Single-area flat topologies are recommended and exhibit greater design efficiencies in termination and transport capacity, but higher cost, and, as discussed and modeled in ANNEX 3, better network performance in comparison with multi-area hierarchical topologies. As illustrated in ANNEX 4, larger administrative areas can be achieved through use of EDR-based TE methods as compared to SDR-based TE methods.

* EDR methods are recommended and exhibit comparable design efficiencies to SDR. This suggests that there is not a significant advantage for employing link-state information in these network designs, especially given the high overhead in flooding link-state information in SDR methods.

* Dynamic transport routing is recommended and achieves capital savings by concentrating capacity on fewer, high-capacity physical fiber links and, as discussed in ANNEX 5, achieves higher network throughput and enhanced revenue by their ability to flexibly allocate bandwidth on the logical links serving the access and inter-node traffic.

Figures:

Figure 6.1 Tradeoff Between Direct Link Capacity and Alternate Path Capacity
7.1 Introduction

As discussed in the draft, Figure 1.1 illustrates a model for network routing and network management and design. The central box represents the network, which can have various configurations, and the traffic routing tables and transport routing tables within the network. Routing tables describe the route choices from an originating node to a terminating node for a connection request for a particular service. Hierarchical, nonhierarchical, fixed, and dynamic routing tables have all been discussed in the draft. Routing tables are used for a multiplicity of services on the telecommunications network, such as an MPLS/TE-based network used for illustration in this ANNEX.

Traffic engineering functions include traffic management, capacity management, and network planning. Figure 1.1 illustrates these functions as interacting feedback loops around the network. The input driving the network is a noisy traffic load, consisting of predictable average demand components added to unknown forecast error and other load variation components. The feedback controls function to regulate the service provided by the network through traffic management controls, capacity adjustments, and routing adjustments. Traffic management provides monitoring of network performance through collection and display of real-time traffic and performance data and allows traffic management controls such as code blocks, connection request gapping, and reroute controls to be inserted when circumstances warrant. Capacity management includes capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment. Forecasting operates over a multiyear forecast interval and drives network capacity expansion. Daily and weekly performance monitoring identify any service problems in the network. If service problems are detected, short-term network adjustment can include routing table updates and, if necessary, short-term capacity additions to alleviate service problems. Updated routing tables are sent to the switching systems either directly or via an automated routing update system. Short-term capacity additions are the exception, and most capacity changes are normally forecasted, planned, scheduled, and managed over a period of months or a year or more. Network design embedded in capacity management includes routing design and capacity design. Network planning includes longer-term node planning and transport network planning, which operates over a horizon of months to years to plan and implement new
In Sections 7.2 to 7.5, we focus on the steps involved in traffic management of the MPLS/TE-based network (Section 7.2), capacity forecasting in the MPLS/TE-based network (Section 7.3), daily and weekly performance monitoring (Section 7.4), and short-term network adjustment in the MPLS/TE-based network (Section 7.5). For each of these three topics, we illustrate the steps involved with examples.

Monitoring of traffic and performance data is a critical issue for traffic management, capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment. This topic is receiving attention in IP-based networks [FGLRR99] where traffic and performance data has been somewhat lacking, in contrast to TDM-based networks where such TE monitoring data has been developed to a sophisticated standard over a period of time [A98]. The discussions in this ANNEX intend to point out the kind and frequency of TE traffic and performance data required to support each function.

7.2 Traffic Management

In this section we concentrate on the surveillance and control of the MPLS/TE-based network. We also discuss the interactions of traffic managers with other work centers responsible for MPLS/TE-based network operation. Traffic management functions should be performed at a centralized work center, and be supported by centralized traffic management operations functions (TMOF), perhaps embedded in a centralized bandwidth-broker processor (here denoted TMOF-BBP). A functional block diagram of TMOF-BBP is illustrated in Figure 7.1.

7.2.1 Real-Time Performance Monitoring

The surveillance of the MPLS/TE-based network should be performed through monitoring the highest bandwidth-overflow/delay-count node-pair, preferably on a geographical display, which is normally monitored at all times. This display should be used in the auto-update mode, which means that every five minutes TMOF-BBP automatically updates the exceptions shown on the map itself and displays the node pairs with the highest bandwidth overflow/delay count. TMOF-BBP also should have displays that show the high bandwidth-overflow/delay-percent pairs within threshold values.

Traffic managers are most concerned with what connection requests can be rerouted and therefore want to know the location of the heaviest concentrations of blocked call routing attempts. For that purpose, overflow/delay percentages can be misleading. From a service revenue standpoint, the difference between 1 percent and 10 percent blocking/delay on a node pair may favor concentration on the 1 percent blocking/delay situation, because there are more connection requests to reroute. TMOF-BBP should also display all the exceptions that there are with the auto threshold display, which displays everything exceeding the present threshold---for example either 1 percent bandwidth-overflow/delay or 1 or more blocked connection requests, in 5 minutes. In the latter case, this display then shows the total blocked connection requests and not just the highest pairs.

For peak-day operation, or operation on a high day (such as a Monday after Thanksgiving), traffic managers should work back and forth between the auto threshold display and the highest blocked-connection-count pair display. They can spend most of their time with the auto threshold display, where
they can see everything that is being blocked. Then, when traffic managers want to concentrate on clearing out some particular problem, they can look at the highest blocked-connection-count pair display, an additional feature of which is that it allows the traffic manager to see the effectiveness of controls.

The traffic manager can recognize certain patterns from the surveillance data. For example, a focused overload on a particular city/node such as caused by a flooding situation discussed further in Sections 7.3, 7.4, and 7.5. The typical traffic pattern under a focused overload is that most locations show heavy overflow/delay into and out of the focus-overload node. Under such circumstances, the display should show the bandwidth overflow/delay percent for any node pair in the MPLS/TE-based network that exceeds 1 percent bandwidth overflow/delay percent.

Ash              <draft-ietf-tewg-qos-routing-02.txt>       [Page ANNEX7-2]

Internet Draft  TE & QoS Methods for IP, ATM, TDM-Based Networks  August 2001

One of the other things traffic managers should be able to see with TMOF-BBP using the highest bandwidth-overflow/delay-count pair display is a node failure. Transport failures should also show on the displays, but the resulting display pattern depends on the failure itself.

7.2.2 Network Control

The MPLS/TE-based network needs automatic controls built into the node processing and also has automatic and manual controls that can be activated from TMOF-BBP. We first describe the required controls and what they do, and then we discuss how the MPLS/TE-based traffic managers work with these controls. Two protective automatic traffic management controls are required in the MPLS/TE-based network: dynamic overload control (DOC), which responds to node congestion, and dynamic bandwidth reservation (DBR), which responds to link congestion. DOC and DBR should be selective in the sense that they control traffic destined for hard-to-reach points more stringently than other traffic.

The complexity of MPLS/TE networks makes it necessary to place more emphasis on fully automatic controls that are reliable and robust and do not depend on manual administration. DOC and DBR should respond automatically within the node software program. For DBR, the automatic response can be coupled, for example, with two bandwidth reservation threshold levels, represented by the amount of idle bandwidth on an MPLS/TE-based link. DBR bandwidth reservation levels should be automatic functions of the link size.

DOC and DBR are not strictly link-dependent but should also depend on the node pair to which a controlled connection request belongs. A connection request offered to an overloaded via node should either be canceled at the originating node or advanced to an alternate via node, depending on the destination of the call. DBR should differentiate between primary (shortest) path and alternate path connection requests.

DOC and DBR should also use a simplified method of obtaining hard-to-reach control selectivity. In the MPLS/TE-based network, hard-to-reach codes can be detected by the terminating node, which then communicates them to the originating nodes and via nodes. Because the terminating node is the only exit point from the MPLS/TE-based network, the originating node should treat a hard-to-reach code detected by a terminating node as hard to reach on all MPLS/TE-based links.

DOC should normally be permanently enabled on all links. DBR should automatically be enabled by an originating node on all links when that originating node senses general network congestion. DBR is particularly important in the MPLS/TE-based network because it minimizes the use of less
efficient alternate path connections and maximizes useful network throughput during overloads. The automatic enabling mechanism for DBR ensures its proper activation without manual intervention. DOC and DBR should automatically determine whether to subject a controlled connection request to a cancel or skip control. In the cancel mode, affected connection requests are blocked from the network, whereas in the skip mode such connection requests skip over the controlled link to an alternate link. DOC and DBR should be completely automatic controls. Capabilities such as automatic enabling of DBR, the automatic skip/cancel mechanism, and the DBR one-link/two-link traffic differentiation adapt these controls to the MPLS/TE-based network and make them robust and powerful automatic controls.

Code-blocking controls block connection requests to a particular destination code. These controls are particularly useful in the case of focused overloads, especially if the connection requests are blocked at or near their origination. Code blocking controls need not block all calls, unless the destination node is completely disabled through natural disaster or equipment failure. Nodes equipped with code-blocking controls can typically control a percentage of the connection requests to a particular code. The controlled E.164 name (dialed number code), for example, may be NPA, NXX, NPA-NXX, or NPA-NXX-XXXX, when in the latter case one specific customer is the target of a focused overload.

A call-gapping control, illustrated in Figure 7.2, is typically used by network managers in a focused connection request overload, such as sometimes occurs with radio call-in give-away contests.

Call gapping allows one connection request for a controlled code or set of codes to be accepted into the network, by each node, once every x seconds, and connection requests arriving after the accepted connection request are rejected for the next x seconds. In this way, call gapping throttles the connection requests and prevents the overload of the network to a particular focal point.

An expansive control is also required. Reroute controls should be able to modify routes by inserting additional paths at the beginning, middle, or end of a path sequence. Such reroutes should be inserted manually or automatically through TMOF-BBP. When a reroute is active on a node pair, DBR should be prevented on that node pair from going into the cancel mode, even if the overflow/delay is heavy enough on a particular node pair to trigger the DBR cancel mode. Hence, if a reroute is active, connection requests should have a chance to use the reroute paths and not be blocked prematurely by the DBR cancel mode.

In the MPLS/TE-based network, a display should be used to graphically represent the controls in effect. Depending on the control in place, either a certain shape or a certain color should tell traffic managers which control is implemented. Traffic managers should be able to tell if a particular control at a node is the only control on that node. Different symbols should be used for the node depending on the controls that are in effect.

7.2.3 Work Center Functions

7.2.3.1 Automatic Controls

The MPLS/TE-based network requires automatic controls, as described above, and if there is spare capacity, traffic managers can decide to reroute. In the example focus-overload situation, the links are occupied sufficiently,
and there is often no network capacity available for reroutes. The DBR control is normally active at the time. In order to get connection requests out of focus-overload-node, traffic managers sometimes must manually disable the DBR control at the focus-overload-node. This gave preference to connection requests going out of the focus-overload-node. Thereby, the focus-overload-node gets much better completion of outgoing connection requests than will the other nodes at completing calls into the focus-overload node. This control results in using the link capacity more efficiently.

Traffic managers should be able to manually enable or inhibit DBR and also inhibit the skip/cancel mechanism for both DBR and DOC. Traffic managers should monitor DOC controls very closely because they indicate switching congestion or failure. Therefore, DOC activations should be investigated much more thoroughly and more quickly than DBR activations, which are frequently triggered by normal heavy traffic.

7.2.3.2 Code Controls

Code controls are used to cancel connection requests for very hard-to-reach codes. Code control is used when connection requests cannot complete to a point in the network or there is isolation. For example, traffic managers should use code controls for a focus overload situation, such as caused by an earthquake, in which there can be isolation. Normal hard-to-reach traffic caused by heavy calling volumes will be blocked by the DBR control, as described above.

Traffic managers should use data on hard-to-reach codes in certain situations for problem analysis. For example, if there is a problem in a particular area, one of the early things traffic managers should look at is the hard-to-reach data to see if they can identify one code or many codes that are hard to reach and if they are from one location or several locations.

7.2.3.3 Reroute Controls

Traffic managers should sometimes use manual reroute even when an automatic reroute capability is there. Reroutes are used primarily for transport failures or heavy traffic surges, such as traffic on heavier than normal days, where the surge is above the normal capabilities of the network to handle the load. Those are the two prime reasons for rerouting. Traffic managers do not usually reroute into a disaster area.

7.2.3.4 Peak-Day Control

Peak-day routing in the MPLS/TE-based network should involve using the primary (shortest) path (CRLSP) as the only engineered path and then the remaining available paths as alternate paths all subject to DBR controls. The effectiveness of the additional alternate paths and reroute capabilities depends very much on the peak day itself. The greater the peak-day traffic, the less effective the alternate paths are. That is, on the higher peak days, such as Christmas and Mother’s Day, the network is filled with connections mostly on shortest paths. On lower peak days, such as Easter or Father’s Day, the use of alternate paths and rerouting capabilities are more effective. This is because the peaks, although they are high and have an abnormal traffic pattern, are not as high as on Christmas or Mother’s Day. So on these days there is additional capacity to complete connection requests on the alternate paths. Reroute paths are particularly available in the early morning and late evening. Depending on the peak day, at times there is also a lull in the afternoon, and TMOP-BBP should normally be able
to find reroute paths that are available.

7.2.4 Traffic Management on Peak Days

A typical peak-day routing method uses the shortest path between node pairs as the only engineered path, followed by alternate paths protected by DBR. This method is more effective during the lighter peak days such as Thanksgiving, Easter, and Father’s Day. With the lighter loads, when the network is not fully saturated, there is a much better chance of using the alternate paths. However, when we enter the network busy hour or combination of busy hours, with a peak load over most of the network, the routing method at that point drops back to shortest-path routing because of the effect of bandwidth reservation. At other times the alternate paths are very effective in completing calls.

7.2.5 Interfaces to Other Work Centers

The main interaction traffic managers have is with the capacity managers. Traffic managers notify capacity managers of conditions in the network that are affecting the data that they use in making decisions as to whether or not to add capacity. Examples are transport failures and node failures that would distort traffic data. A node congestion signal can trigger DOC; DOC cancels all traffic destined to a node while the node congestion is active. All connection requests to the failed node are reflected as overflow connection requests for the duration of the node congestion condition. This can be a considerable amount of canceled traffic. The capacity manager notifies traffic managers of the new link capacity requirements that they are trying to get installed but that are delayed. Traffic managers can then expect to see congestion on a daily basis or several times a week until the capacity is added. This type of information is passed back and forth on a weekly or perhaps daily basis.

7.3 Capacity Management -- Forecasting

In this section we concentrate on the forecasting of MPLS/TE-based node-to-node loads and the sizing of the MPLS/TE-based network. We also discuss the interactions of network forecasters with other work centers responsible for MPLS/TE-based network operations.

Network forecasting functions should be performed from a capacity administration center and supported by network forecasting operations functions integrated into the BBP (NFOF-BBP). A functional block diagram of NFOF-BBP is illustrated in Figure 7.3. In the following two sections we discuss the steps involved in each functional block.

7.3.1 Load Forecasting

7.3.1.1 Configuration Database Functions

As illustrated in Figure 7.3, the configuration database is used in the forecasting function, and within this database are defined various specific components of the network itself, for example: backbone nodes, access nodes, transport points of presence, buildings, manholes, microwave towers, and other facilities.

Forecasters maintain configuration data for designing and forecasting the MPLS/TE-based network. Included in the data for each backbone node and access node, for example, are the number/name translation capabilities, equipment type, type of signaling, homing arrangement, international routing capabilities, operator service routing capabilities, and billing/recording
capabilities. When a forecast cycle is started, which is normally each month, the first step is to extract the relevant pieces of information from the configuration database that are necessary to drive network forecasting operations functions (NFOF-BBP). One of information items indicates the date of the forecast view; this is when the configuration files were frozen, which then represents the network structure at the time the forecast is generated.

7.3.1.2 Load Aggregation, Basing, and Projection Functions

NFOF-BBP should process data from a centralized message database, which represents a record of all connection requests placed on the network, over four study periods within each year, for example, March, May, August, and November, each a 20-day study period. From the centralized a sampling method can be used, for example a 5 percent sample of recorded connections for 20 days. Forecasters can then equate that 5 percent, 20-day sample to one average business day. The load information then consists of messages and traffic load by study period. In the load aggregation step, NFOF-BBP may apply nonconversation time factors to equate the traffic load obtained from billed traffic load to the actual holding time traffic load.

The next step in load forecasting is to aggregate all of the access-node-to-access-node loads up to the backbone node-pair level. This produces the backbone-node-to-backbone-node traffic item sets. These node-to-node traffic item sets are then routed to the candidate links. NFOF-BBP should then project those aggregated loads into the future, using smoothing techniques to compare the current measured data with the previously projected data and to determine an optimal estimate of the base and projected loads. The result is the initially projected loads that are ready for forecaster adjustments and business/econometric adjustments.

7.3.1.3 Load Adjustment Cycle and View of Business Adjustment Cycle

Once NFOF-BBP smoothes and projects the data, forecasters can then enter a load adjustment cycle. This should be an online process that has the capability to go into the projected load file for all the forecast periods for all the years and apply forecaster-established thresholds to those loads. For example, if the forecaster requests to see any projected load that has deviated more than 15 percent from what it was projected to be in the last forecast cycle, a load analysis module in NFOF-BBP should search through all the node pairs that the forecaster is responsible for, sort out the ones that exceed the thresholds, and print them on a display. The forecaster then has the option to change the projected loads or accept them.

After the adjustment cycle is complete and the forecasters have adjusted the loads to account for missing data, erroneous data, more accurate current data, or specifically planned events that cause a change in load, forecasters should then apply the view of the business adjustments. Up to this point, the projection of loads has been based on projection models and network structure changes, as well as the base study period billing data. The view of the business adjustment is intended to adjust the future network loads to compensate for the effects of competition, rate changes, and econometric factors on the growth rate. This econometric adjustment process tries to encompass those factors in an adjustment that is applied to the traffic growth rates. Growth rate adjustments should be made for each business, residence, and service category, since econometric effects vary according to service category.

7.3.2 Network Design
Given the MPLS/TE-based node-pair loads, adjusted by the forecasters, and also adjusted for econometric projections, the network design model should then be executed by NFOF-BCC based on those traffic loads. The node-to-node loads are estimated for each hourly backbone-node-to-backbone-node traffic load, including the minute-to-minute variability and the day-to-day variation, plus the control parameters. The access-node-to-backbone-node links should also be sized in this step.

A list of all the MPLS/TE-based node pairs should then be sent to the transport planning database, from which is extracted transport information relative to the transport network between the node pairs on that list. Once the information has been processed in the design model, NFOF-BBP should output the MPLS/TE-based forecast report. Once the design model has run for a forecast cycle, the forecast file and routing information should be sent downstream to the provisioning systems, planning systems, and capacity management system, and the capacity manager takes over from there as far as implementing the routing and the link capacity called for in the forecast.

7.3.3 Work Center Functions

Capacity management and forecasting operation should be centralized. Work should be divided on a geographic basis so that the MPLS/TE-based forecaster and capacity manager for a region can work with specific work centers within the region. These work centers include the node planning and implementation organizations and the transport planning and implementation organizations. Their primary interface should be with the system that is responsible for issuing the orders to augment link capacity. Another interface is with the routing organization that processes the routing information coming out of NFOF-BBP.

NFOF-BBP should provide a considerable amount of automation, and as such people can spend their time on more productive activities. By combining the forecasting job and the capacity management job into one centralized operation, additional efficiencies are achieved from a reduction in fragmentation. By centralizing the operations, this avoids duplication from distributing the operation within regional groups. And, with the automation, time need only be spent to clear a problem or analyze data outliers, rather than to check and verify everything.

This operation requires people who are able to understand and deal with a more complex network, and the network complexity will continue to increase as new technology and services are introduced. Other disciplines can usefully centralize their operations, for example, node planning, transport planning, equipment ordering, and inventory control. With centralized equipment-ordering and inventory control, for example, all equipment required for the network can be bulk ordered and distributed. This leads to a much more efficient use of inventory.

7.3.4 Interfaces to Other Work Centers

Network forecasters work cooperatively with node planners, transport planners, traffic managers, and capacity managers. With an MPLS/TE network, forecasting, capacity management, and traffic management must tie together closely. One way to develop those close relationships is by having centralized, compact work centers. The forecasting process essentially drives all the downstream construction and planning processes for an entire network operation.
7.4 Capacity Management - Daily and Weekly Performance Monitoring

In this section we concentrate on the analysis of node-to-node capacity management data and the design of the MPLS/TE-based network. Capacity management becomes mandatory at times, as seen from the node-to-node traffic data, when significant congestion problems are extant in the network or when it is time to implement a new forecast. We discuss the interactions of capacity managers with other work centers responsible for MPLS/TE-based network operation. Capacity management functions should be performed from a capacity administration center and should be supported by the capacity management operations functions embedded, for example, in the BBP (denoted here as the CMOF-BBP). A functional block diagram of the CMOF-BBP is illustrated within the lower three blocks of Figure 7.3. In the following sections we discuss the processes in each functional block.

7.4.1 Daily Congestion Analysis Functions

A daily congestion summary should be used to give a breakdown of the highest to the lowest node-pair congestion that occurred the preceding day. This is an exception-reporting function, in which there should be an ability to change the display threshold. For example, the capacity manager can request to see only node pairs whose congestion level is greater than 10 percent. Capacity managers investigate to find out whether they should exclude these data and, if so, for what reason. One reason for excluding data is to keep them from downstream processing if they are associated with an abnormal network condition. This would prevent designing the network for this type of nonrecurring network condition. In order to find out what the network condition was, capacity managers consult with the traffic managers. If, for example, traffic managers indicate that the data is associated with an abnormal network condition, such as a focused overload due to flooding the night before, then capacity managers may elect to exclude the data.

7.4.2 Study-week Congestion Analysis Functions

The CMOF-BBP functions should also support weekly congestion analysis. This should normally occur after capacity managers form the weekly average using the previous week’s data. The study-week data should then be used in the downstream processing to develop the study-period average. The weekly congestion data are set up basically the same way as the daily congestion data and give the node pairs that had congestion for the week. This study-week congestion analysis function gives another opportunity to review the data to see if there is a need to exclude any weekly data.

7.4.3 Study-period Congestion Analysis Functions

Once each week, the study-period average should be formed using the most current four weeks of data. The study-period congestion summary gives an idea of the congestion during the most current study period, in which node pairs that experienced average business day average blocking/delay greater than 1 percent are identified. If congestion is found for a particular node pair in a particular hour, the design model may be exercised to solve the congestion problem. In order to determine whether they should run the design model for that problem hour, capacity managers should first look at the study-period congestion detail data. For the node pair in question they look at the 24 hours of data to see if there are any other hours for that node pair that should be investigated. Capacity managers should also determine if there is pending capacity addition for the problem node pair.

7.5 Capacity Management - Short-Term Network Adjustment
7.5.1 Network Design Functions

There are several features should be available in the design model. First, capacity managers should be able to select a routing change option. With this option, the design model should make routing table changes to utilize the network capacity that is in place to minimize congestion. The design model should also design the network to the specified grade-of-service objectives. If it cannot meet the objectives with the network capacity in place, it specifies how much capacity to add to which links in order to meet the performance objectives. The routing table update implementation should be automatic from the CMOF-BBP all the way through to the network nodes. An evaluator option of the design model should be available to determine the carried traffic per link, or network efficiency, for every link in the network for the busiest hour.

7.5.2 Work Center Functions

Certain sections of the network should be assigned so that all capacity managers have an equal share of links that they are responsible for. Each capacity manager therefore deals primarily with one region. Capacity managers also need to work with transport planners so that the transport capacity planned for the links under the capacity manager’s responsibility is available to the capacity manager. If, on a short-term basis, capacity has to be added to the network, capacity managers find out from the transport planner whether the transport capacity is available. CMOF-BBP is highly automated, and the time the capacity manager spends working with CMOF-BBP system displays should be small compared with other daily responsibilities. One of the most time-consuming work functions is following up on the capacity orders to determine status: Are they in the field? Does the field have them? Do they have the node equipment working? If capacity orders are delayed, the capacity manager is responsible for making sure that the capacity is added to the network as soon as possible. With the normal amount of network activity going on, that is the most time-consuming part of the work center function.

7.5.3 Interfaces to Other Work Centers

The capacity manager needs to work with the forecasters to learn of network activity that will affect the MPLS/TE-based network. Of concern are new nodes coming into the network capacity management activity that affects the MPLS/TE-based network. Capacity managers should interact quite frequently with traffic managers to learn of network conditions such as cable cuts, floods, or disasters. Capacity managers detect such activities the next day in the data; the network problem stands out immediately. Before they exclude the data, however, capacity managers need to talk with the traffic managers to find out specifically what the problem was in the network. In some cases, capacity managers will share information with them about something going on that they may not be aware of. For example, capacity managers may be able to see failure events in the data, and they can share this type of information with the traffic managers. Other information capacity managers might share with traffic managers relates to peak days. Capacity managers are able the next morning to give the traffic managers the actual reports and information of the load and congestion experienced in the network.

Capacity managers also work with the data collection work center. If they miss collecting data from a particular node for a particular day, capacity managers should discuss this with that work center to get the data into CMOF-BBP. In CMOF-BBP, capacity managers should have some leeway in getting data into the system that may have been missed. So if data are missed one night on a particular node, the node should be available to be repolled to
Capacity managers frequently communicate with the routing work centers because there is so much activity going on with routing. For example, capacity managers work with them to set up the standard numbering/naming plans so that they can access new routing tables when they are entered into the network. Capacity managers also work with the people who are actually doing the capacity order activity on the links. Capacity managers should try to raise the priority on capacity orders if there is a congestion condition, and often a single congestion condition may cause multiple activities in the MPLS/TE network.

Hence an important simplification introduced with the SDR/EDR-based networks is that routing tables need not be calculated by the design model, because these are computed in real time by the node or BBP. This leads to simplifications in that the routing tables computed in TDR-based networks are no longer needed. Hence simplifications are introduced into the administration of network routing. With TDR, routing tables must be periodically reoptimized and downloaded into nodes via the CMOF-BBP process. Reoptimizing and changing the routing tables in the TDR-based network represents an automated yet large administrative effort involving perhaps millions of records. This function is simplified in SDR/EDR-based networks since the routing is generated in real time for each connection request and then discarded. Also, because SDR/EDR-based TE adapts to network conditions, less network churn and short-term capacity additions are required. This is one of the operational advantages of SDR/EDR-based MPLS/TE networks—that is, to automatically adapt TE so as to move the traffic load to where capacity is available in the network.

Conclusions/recommendations reached in this ANNEX include the following:

* Monitoring of traffic and performance data is recommended and is required for traffic management, capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment.

* Traffic management is recommended and is required to provide monitoring of network performance through collection and display of real-time traffic and performance data and allow traffic management controls such as code blocks, connection request gapping, and reroute controls to be inserted when circumstances warrant.

* Capacity management is recommended and is required for capacity forecasting, daily and weekly performance monitoring, and short-term network adjustment.

* Forecasting is recommended and is required to operate over a
multiyear forecast interval and drive network capacity expansion.

* Daily and weekly performance monitoring is recommended and is required to identify any service problems in the network. If service problems are detected, short-term network adjustment can include routing table updates and, if necessary, short-term capacity additions to alleviate service problems. Updated routing tables are sent to the switching systems either directly or via an automated routing update system.

* Short-term capacity additions are recommended and are required as needed, but only as an exception, whereas most capacity changes are normally forecasted, planned, scheduled, and managed over a period of months or a year or more.

---

Ash <draft-ietf-tegw-qos-routing-02.txt> [Page ANNEX7-11]

Internet Draft TE & QoS Methods for IP, ATM, TDM-Based Networks August 2001

* Network design, which includes routing design and capacity design, is recommended and is required within the capacity management function.

* Network planning is recommended and is required for longer-term node planning and transport network planning, and operates over a horizon of months to years to plan and implement new node and transport capacity.

Figures:

Figure 7.1 Traffic Management Operations Functions with Bandwidth-Broker Processor
Figure 7.2 Call Gap Control
Figure 7.3 Capacity Management Functions within Bandwidth-Broker Processor

Ash <draft-ietf-tegw-qos-routing-02.txt> [Page ANNEX7-12]