

Committee T1 -
Telecommunications

AD-A278 746



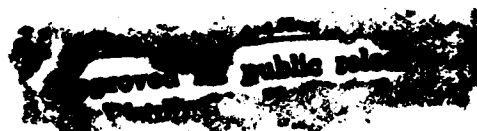
Report No. 24

November 1993

A Technical Report
on
Network Survivability
Performance

SDTIC
ELECTE
APR 07 1994
S E D

Prepared by
T1A1.2
Working Group
on Network Survivability
Performance



94-10693



Committee T1 is sponsored by the Alliance for Telecommunications Industry Solutions
(formerly the Exchange Carriers Standards Association)
Accredited by American National Standards Institute

Technical Report

Copyright © 1993 by Alliance for Telecommunications Industry
Solutions (formerly Exchange Carriers Standards Association).
All rights reserved.

No part of this publication may be reproduced in any form,
in an electronic retrieval system or otherwise, without the
prior written permission of the publisher.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE Nov 1993		3. REPORT TYPE AND DATES COVERED Interim: FYs 1991-1993	
4. TITLE AND SUBTITLE "A Technical Report on Network Survivability Performance, Report No. 24"				5. FUNDING NUMBERS N/A	
6. AUTHOR(S) Committee T1A1.2 Working Group on Network Survivability Performance					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Alliance for Telecommunications Industry Solutions (ATIS) Standards Committee T1 Telecommunications 1200 G Street, N.W. Suite 500 Washington, DC 20005				8. PERFORMING ORGANIZATION REPORT NUMBER Document: T1A1.2/93-001R3	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Information Systems Agency Center for Standards Associate Director, Code TBA Washington, DC 20305-2000				10. SPONSORING/MONITORING AGENCY REPORT NUMBER N/A	
11. SUPPLEMENTARY NOTES This report was prepared by the cooperation of telecommunications network providers vendors, and users, such as the U.S. Government's National Communication System (NCS), Department of Defense (DoD) and Federal Aviation Administration (FAA)					
12a. DISTRIBUTION/AVAILABILITY STATEMENT Public Availability Unlimited				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This Technical Report has been developed to address the survivability of telecommunications networks including services. It responds to the need for a common understanding of, and assessment techniques for network survivability, availability, integrity, and reliability. It provides a basis for designing and operating telecommunications networks to user expectations for network survivability and a foundation for continuing industry activities in the subject area. This report focuses on the survivability of both public and private networks and covers a wide range of users. Two frameworks are established for quantifying and categorizing service outages, and for classifying network survivability techniques and measures. The performance of the network survivability techniques is considered; however, recommended objectives are not established for network survivability performance. Recommendations are made for continuing work on this topic by the T1A1.2. (Continued)					
14. SUBJECT TERMS Telecommunications, network, survivability, reliability, availability, maintainability, quality of service, parameters, measures, techniques, failure signaling, network management, user expectation, restoration,...				15. NUMBER OF PAGES 88	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified		20. LIMITATION OF ABSTRACT	

13. Abstract (Continued)

The intended audience for this Technical Report includes providers, users, and regulators of telecommunications networks and services. These include designers, planners, traffic engineers, and individuals in charge of operations, maintenance, management and administration. They can use this report to quantify and enhance the survivability of their networks. It also provides a service outage framework that can help users communicate their expectations of service and their requirements for network survivability. Equipment vendors can use it to design and build equipment to enhance network reliability. The government may use it for the acquisition, procurement, and the preceding purposes!

A Technical Report on Network Survivability Performance

Abstract

This Technical Report provides information on the network survivability performance of telecommunications networks. Although techniques, parameters and methods needed to study network survivability performance are defined, recommended parameter objectives are not established.

Document T1A1.2/93-001R3

Prepared by T1A1.2

Working Group on Network Survivability Performance

Accession For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

94 4 7 061

Foreword

This report addresses the growing concerns from the telecommunications community about the survivability of telecommunications networks, including the services provided. It is intended to provide a basis for designing and operating telecommunications networks to meet users' expectations regarding network survivability.

The intended audience of this report includes providers, users and regulators of telecommunications networks and services, as well as telecommunications equipment providers.

Suggestions for improvement of this report are welcome. They should be sent to the Exchange Carriers Standards Association, Suite 500, 1200 G Street NW., Washington, DC 20005.

Working Group T1A1.2 (formerly T1Q1.2) on Network Survivability, which developed this report had the following officers and participants:

Chair:	A. Zolfaghari
Vice-Chair and Chief Editor:	F. Kaudel

Editors: J. Cassano	M. Petty
R. Doverspike	J. Yuristy
G. Koerner, D.O.D.	F. Zamora
J. Lord	A. Zolfaghari
S. Makris	

Active Participants:

T. Adcock	M. Kanotz	M. Petty
R. Arsaga	C. Karpewicz	T. Pillai
L. Barker	F. Kaudel	R. Pinto
N. Barlett	G. Koerner	A. Reilly
F. Burns	B. Lollar	L. Sayadian
J. Cassano	J. Lord	T. Soejima
M. Daneshmand	S. Makris	J. Sosnosky
R. Doverspike	H. Mar	E. Turner
F. Ellefson	J. McDonough	R. Waller
W. Grover	M. Mostrel	G. Williams
H. Holton	S. Nadkarni	J. Yuristy
J. James	D. Nak	F. Zamora
R. Jensen	C. Pack	A. Zolfaghari

Table of Contents

0.	Executive Summary	1
1.	Purpose, Scope, Application and Outline	2
1.1	Purpose	2
1.2	Scope	2
1.3	Application	3
1.4	Outline	3
2.	Related Work	3
2.1	Committee T1 Standards Work	3
2.2	International Standards Work	4
2.3	Other Domestic Forums and Committees	4
2.4	U.S. NSTAC Task Force Work	5
3.	Introduction	6
3.1	User Categories	8
4.	Framework for Quantifying and Categorizing Service Outages	9
4.1	Service Outage Parameters	10
4.2	Service Outage Categories	11
5.	Framework for Classifying Network Survivability Techniques and Measures	12
5.1	Physical Layer	14
5.2	System Layer	15
5.3	Logical Layer	15
5.4	Service Layer	17
5.5	Summary	17
6.	Network Survivability Techniques	18
6.1	Physical Layer	20
6.2	System Layer	21
6.2.1	Point-to-Point Systems with Automatic Protection Switching	21
6.2.2	Rings	21
6.2.2.1	Unidirectional Ring	22
6.2.2.2	Bidirectional Ring	22
6.2.2.3	Line Protection Switched Ring	23
6.2.2.4	Path Protection Switched Ring	25
6.3	Logical Layer	26
6.3.1	DCS Reconfiguration Strategies	26
6.3.1.1	Detection and Notification	27
6.3.1.2	Identification	27
6.3.1.3	Path (Route) Selection	27
6.3.1.4	Rerouting	30
6.3.1.5	Normalization	30
6.3.1.6	Summary of DCS Reconfiguration Methods	30
6.4	Service Layer	31
6.4.1	Circuit Switching	31
6.4.1.1	Size Limits	31

6.4.1.2	Dynamic Routing	31
6.4.1.3	Reconfiguration	32
6.4.1.4	Network Management	32
6.4.1.5	Multi-Serving	32
6.4.2	Packet Switching	32
6.4.3	Common Channel Signaling	33
6.4.3.1	Service Layer Architecture for MTP	33
6.4.3.2	Service Layer Architecture for SCCP	37
6.4.3.3	Manual Traffic Management Controls	38
6.4.3.4	Relationship of Software Diversity to CCSN Reliability ...	38
6.5	Integrated Techniques	42
7.	Network Survivability Performance Analysis	44
7.1	Network Survivability Characterization	44
7.2	Network Survivability Analysis Model	44
7.2.1	Given Occurrence of Failure Survivability Model	44
7.2.2	Random Occurrence of Failure Survivability Model	45
7.2.3	Application of GOF and ROF Models	45
7.3	GOF Network Survivability Measures	46
7.3.1	Service Layer Examples	47
7.3.2	Logical and System Layer Examples	48
7.4	ROF Network Survivability Measures	50
7.4.1	Service Layer Examples	50
7.4.1.1	Circuit Switching	50
7.4.1.2	Packet Switching	51
7.4.1.3	Common Channel Signaling	51
7.4.2	Logical and System Layer Examples	55
7.5	Qualitative Assessment of Network Survivability Techniques	57
8.	Suggestions to General Industry	58
9.	Recommendations to Standards Organizations	59
9.1	Recommendations for Committee T1	59
9.2	Recommendations for Future T1A1.2 Work	59
10.	Summary	59
11.	Bibliography	60
12.	Definitions	64
13.	Abbreviations and Acronyms	66
	Appendix A. Telecommunications Service Priority	67
	Appendix B. Example Network Survivability Analyses and Assessments	70
	Appendix B.1 Bidirectional Ring Survivability Example	70
	Appendix B.2 Network Survivability Assessment Example	71
	Appendix C. User Expectations	73
	Appendix C.1 U.S. Government User Expectations	73
	Appendix D. Tolerance Categories for Restoration Times	76
	Index	81

0. Executive Summary

As a result of growing concerns from the telecommunications community, this Technical Report has been developed to address the survivability of telecommunications networks, including the services provided.

The report is needed to provide a common understanding and common assessment techniques for network survivability. It provides a basis for designing and operating telecommunications networks to meet users' expectations regarding network survivability. The intended audience of this report includes providers, users and regulators of telecommunications networks and services, as well as telecommunications equipment suppliers. The report also provides a foundation for continuing industry activities in this area.

Terminology to characterize network survivability is provided. In particular, network survivability is defined to be: (i) the ability of a network to maintain or restore an acceptable level of performance during network failures by applying various restoration techniques, and (ii) the mitigation or prevention of service outages from network failures by applying preventative techniques. Network survivability includes other industry terms, such as "network integrity" and "network reliability," and is related to "network availability."

A framework for quantifying service outages is developed. The parameters for this framework are the unservability of services affected by the network failure, the duration of the outage, and the extent of the failure (e.g., geographical area, population, or network). Categories of service outage are outlined. The categories depend on type of user, network and service. Types of users include carriers, residential customers, government agencies, educational and medical institutions, as well as business and financial customers.

A four-layer framework is described for classification of network survivability techniques in telecommunication networks. These layers are physical, system, logical and service. In addition to providing a common basis for describing and comparing techniques, the framework identifies layer(s) responsible for reacting to the various types of failures and their interaction.

Techniques available to network providers to enhance the survivability of their telecommunications networks at each layer are described. Two basic approaches to compare survivability techniques and evaluate network survivability performance are given. The first approach (Given Occurrence of Failure, or GOF) uses a conditional approach and defines survivability measures for a network assuming that given failures have occurred. The second approach (Random Occurrence of Failure, or ROF) uses probability of network failure and, possibly,

Technical Report No. 24

rates of repair and/or restoration, to calculate various measures of network unservability or loss.

Suggestions are given for the general industry. Key suggestions are outlined here:

- quantify service outages with the framework described herein,
- use the terminology defined herein for describing network survivability, including network reliability and network integrity,
- use the layered network survivability framework described herein for clarifying failure survivability analyses, objectives and methods,
- plan survivability jointly (e.g., interexchange carrier and exchange carrier interworking), and
- use the performance measures defined herein to compare survivability techniques and to evaluate network survivability performance.

Recommendations are also given for future standards work. Key recommendations are outlined here:

- better quantification of the qualifying regions for service outage categories,
- validation of traffic characteristics,
- analysis of user expectations of network survivability performance, planning, engineering and operations guidelines for network survivability performance, and
- standardization of network survivability performance measures.

1. Purpose, Scope, Application and Outline

1.1 Purpose As a result of growing concerns from the telecommunications community, this Technical Report has been developed to address the survivability of telecommunications networks, including the services provided. The report is a response to the need for a common understanding of, and assessment techniques for, network survivability, including availability, integrity, and reliability. It also provides a basis for designing and operating telecommunications networks to meet users expectations for network survivability. Further this report provides a foundation for continuing industry activities in the subject area.

1.2 Scope This Technical Report focuses on the survivability of both public and private networks and covers a wide range of potential users. Two frameworks are established: the first for quantifying and categorizing service outages, and the second for classifying network survivability techniques and measures. The performance of network survivability techniques is considered; however,

Technical Report No. 24

recommended objectives for network survivability performance are not established. Recommendations are made for continuing work on this topic.

1.3 Application The intended audience of this Technical Report includes providers, users, and regulators of telecommunications networks, and services, as well as suppliers of telecommunications equipment. Network provider personnel — including designers, planners, traffic engineers, and individuals in charge of operations, maintenance, management and administration — can use this Technical Report to quantify and enhance the survivability of their networks. This Technical Report also provides a service outage framework that can help users communicate to the providers their expectations of service and a framework that allows network providers to specify their requirements for network survivability. Telecommunications equipment suppliers can use this Technical Report to guide the design and building of equipment to improve network survivability.

1.4 Outline Section 2 and Appendix A review related work. Several aspects of Operations, Administration and Maintenance (OA&M) functions are covered in other Committee T1 documents (see Section 2.1). Section 3 is the introduction to this report. Section 4 provides a framework for quantifying and categorizing service outages. Section 5 introduces a framework for classifying network survivability techniques and measures. Network survivability techniques that mitigate the impact of failures or service degradation are addressed in Section 6. Definitions and examples of parameters and measures to evaluate the survivability performance of any network are presented in Section 7 and Appendix B. Recommended objectives for these parameters are not established. Section 8 gives general suggestions and Section 9 gives recommendations to Committee T1 and Working Group T1A1.2. Section 10 summarizes this Technical Report. Appendix A describes the Telecommunications Service Priority System. Appendix B gives example network survivability analyses and assessments. Some user expectations for service outage duration are given in Appendix C, and tolerance categories for restoration times are introduced in Appendix D.

2. Related Work

2.1 Committee T1 Standards Work In addition to T1A1, the following four other T1 technical subcommittees are involved in work related to network survivability performance: T1E1, T1M1, T1S1 and T1X1.

T1E1 is considering environmental standards. These standards also impact network survivability performance.

Technical Report No. 24

T1M1 Working Group 2 (SONET¹ OA&M) is considering the OA&M aspects of failed networks.

T1S1 Working Group 3 (Signaling System Number 7, or SS7, protocol) is studying congestion controls for Common Channel Signaling Networks (CCSNs) and has produced a report to the Signaling Network Systems Focus Team of the Network Reliability Council (see Section 2.3).

T1X1 is charged with developing standards for ring architectures, protection switching and SONET compatibility. This covers a range of survivable network architectures.

2.2 International Standards Work The following Telecommunication Standardization Sector (ITU-TS, formerly CCITT) study groups are involved in work related to network survivability performance:

Study Group 2 addresses the overall operation of telecommunication networks. This work addresses various design alternatives for survivable networks.

Study Group 4 is studying the restoration of failed international exchanges, and transmission systems.

Study Group 5 is studying survivability issues for outside plant.

Study Group 11 is studying CCSNs and switching systems.

Study Group 13 is looking at digital network architectures. This work addresses rings and other survivable network architectures.

Study Group 15 is addressing transmission system requirements, including the restoration requirements for cross-connect systems.

2.3 Other Domestic Forums and Committees

The Industry Carriers Compatibility Forum (ICCF) consists of representatives from the industry carriers and major vendors and deals with compatibility problems (including SS7) between carriers. It is sponsored by the Carrier Liaison Committee (CLC) of the Exchange Carriers Standards Association (ECSA).

The Network Operations Forum (NOF) consists primarily of representatives from the industry carriers Operations departments who meet periodically to deal with

¹ Synchronous Optical NETWORK.

Technical Report No. 24

operational problems of mutual interest. It also falls under the auspices of the CLC.

The Network Reliability Council (also known as the NRC) was formed by the Federal Communications Commission (FCC) in the fourth quarter of 1991 in response to major telecommunications outages. The Network Reliability Council is expected to provide the FCC and the Telecommunications Industry with recommendations for prevention of public telecommunications network outages and minimization of the impact of such outages. The Network Reliability Council consists of the following seven focus teams: 1) Signaling Network Systems, 2) Fiber Cables (with focus on cable digups), 3) Digital Cross-connect Systems, 4) Power Systems, 5) Switching (with focus on software), 6) Fire Prevention, and 7) Enhanced 911 Systems.

2.4 U.S. NSTAC Task Force Work The following task forces established by the National Security Telecommunications Advisory Committee (NSTAC) studied Network Survivability issues and forwarded their recommendations to the U.S. Government:

Telecommunications Systems Survivability (TSS), June 1988

This task force studied the work of five previous National Security Telecommunications Advisory Committee (NSTAC) task forces:

1. **Commercial Satellite Survivability (CSS), June 1983 and August 1989**
In response to the twelve recommendations of this task force, the government developed the Commercial SATCOM Interconnectivity (CSI) program where commercial satellites could be used to route traffic around damaged areas of public switched networks.
2. **Commercial Network Survivability (CNS), August 1985**
To improve the survivability of public switched networks when Section 706 of the Communications Act of 1934 is invoked, this task force recommended interconnecting the commercial telecommunications networks.
3. **Electromagnetic Pulse (EMP), July 1985**
As a result of this task force's recommendations, the government and industry is addressing EMP in T1E1.
4. **National Coordinating Mechanism (NCM), December 1983**
This task force called for the establishment of the National Coordination Center (NCC): a joint industry-government operation to initiate, coordinate, restore, and reconstitute public switched network services under all conditions of crisis or emergency.

Technical Report No. 24

5. Automated Information Processing (AIP), October 1985

This task force addressed the survivability and security of automated information processes within public switched networks.

Telecommunications Industry Mobilization (TIM), April 1989

This task force examined the telecommunications industry's ability to rapidly and effectively "marshal their telecom resources needed to transition from a normal state to a state of readiness for war or other national emergency."

Telecommunication Service Priority (TSP), September 1990

This task force developed a system for priority provisioning and restoration of critical federal, state, local, and foreign government communications services. It replaces the existing Restoration Priority (RP) System and revokes the public network precedence system. A detailed discussion of the TSP system is in Appendix A.

National Research Council Report Task Force, March 1990

This task force reviewed the National Research Council's 1989 report "Growing Vulnerability of the Public Switched Networks: Implications for National Security Emergency Preparedness" and concluded that public switched networks of the 1990s are growing more survivable because of: (i) the diversity provided by the major interexchange carriers, (ii) the demand by network users for very high availability of network services, and (iii) the development of increasingly robust network architectures that incorporate major advances in transmission, switching, and signaling technologies.

Intelligent Network (IN), November 1990

This task force investigated the evolving capabilities of the IN to provide customized software-controlled network services and examined the vulnerability and interoperability issues associated with this emerging technology.

Energy Task Force, February 1990

This task force recommended the government establish a priority for electric power restoration and fuel distribution to public switched network telecommunications companies during emergency conditions.

3. Introduction

Historically, the designers and operators of telecommunications networks have placed a high emphasis on overall network performance quality and continuity of the services they provide. This has been termed in the industry as network availability.

Technical Report No. 24

Due to the nature of business today, a significant number of end users have an absolute dependency on telecommunications. Advancing technology (in both hardware and software), increased centralized control and services evolution have resulted in increased complexity, and capacity concentration. Users of these networks and services have also raised their requirements and expectations. The demand for high-quality performance, assured service continuity, and transparency to failures has never been greater. Users who require enhanced levels of network survivability performance include not only government, military, and emergency organizations, but the general public as well. This is termed by the users to be network reliability. The term network integrity is also sometimes used in this context. The use of the term "reliability" in this network context should not be confused with the more technical use of the term in the context of the ability of a specific component or equipment to perform its intended function over a specified period of time.

Along with the high concentration of telecommunications facilities and services brought about by technology, there have been continuing occurrences of service outages caused by metallic and fiber cable cuts, fires, natural disasters, human errors, software faults, sabotage, and malicious damage. The potential impact of a single network failure is higher in magnitude, geographic scope, and public perception than ever before. This potential impact heightens the need for network survivability; network survivability has two components: (i) the ability of a network to maintain or restore an acceptable level of performance during network failures by applying various restoration techniques, and (ii) the mitigation or prevention of service outages from network failures by applying preventative techniques. Preventative techniques can reduce the need for restoration techniques.

Depending on the type and scope of the failures, which can range from affecting very large geographic areas to a single customer, service restoration times can range from many days to milliseconds.

The type and scope of service outages, occurring from both natural and human-caused network failures, are categorized in Section 4 as being one of catastrophic, major or minor. These categories are based on the impact of the failure on users. For minor failures, network restoration can be automatic and virtually transparent to the user. However, for catastrophic and major events affecting large populations, some service may be restored automatically, but most service will be restored manually and as quickly as possible, based upon availability of personnel and replacement facilities.

A framework for quantifying and categorizing service outages in specific terms has been developed in Section 4. The parameters of this framework are: (i) the Unservability (U) of some or all of the provided service affected by the failure; (ii) the Duration (D) during which the service outage exists, and (iii) the Extent (E) in

terms of the geographical area, population and traffic volumes affected by the network failure. Throughout this report, this framework will be referred to as the (U, D, E) triple.

To lessen the occurrence and the degree of such service outages, survivability techniques have included redundancy of networks and systems, provisioning of highest reliability systems, good preventative maintenance, monitoring and surveillance, manual response, and optimized restoration and repair techniques.

Survivability techniques are being developed, based on intelligent and efficient technology, that provide the potential and capability to develop new and preemptive techniques for network survivability. Innovative developments in all layers and elements of networks are making it technically and practically possible to cope with many types of network failures. When this is coupled with well-planned reactive techniques such as readiness and contingency planning (Disaster Recovery), the structure is in place for the capability to provide a full spectrum of network survivability. The information contained within this Technical Report can help to provide a focus and impetus for ensuring the survivability performance of networks and services in general, as well as a base for enhanced network survivability techniques for special applications needs.

3.1 User Categories This section lists a range of telecommunications network users, and groups them into categories based on broad functional or service related characteristics.

User concerns have been heightened by telecommunications failures that have dramatized the vulnerability of telecommunications networks to fires, viruses, software errors, optical fiber/metallic cable cuts, power failures and other natural or human-caused accidents.

Users have unique requirements and expectations for uninterrupted service, depending on the user type, service value and subscription cost. To meet these user expectations, service providers may make use of: specific network configurations, dynamic routing, restoration techniques or procedures, network/system hardening, prevention of unauthorized access, and emerging technologies. Specific user groups with similar or unique expectations include:

- Interexchange Carriers (IC),
- Exchange Carriers (EC),
- residential customers (including Plain Old Telephone Service, or POTS),
- government agencies (federal, state and municipal),
- educational institutions,
- business customers, (excluding some financial institutions)
- cellular carriers,

- information service providers,
- other access providers,
- financial institutions, and
- emergency organizations.

Note that many service providers are also considered to be network users. Users need to identify their survivability expectations. An example of the U.S. Government user expectations is discussed in Appendix C.

As an example of a user network, a public switched network illustrated in Figure 3.1 is considered to be divided into IC and EC portions². In this example, an IC is a user of two ECs' access services.

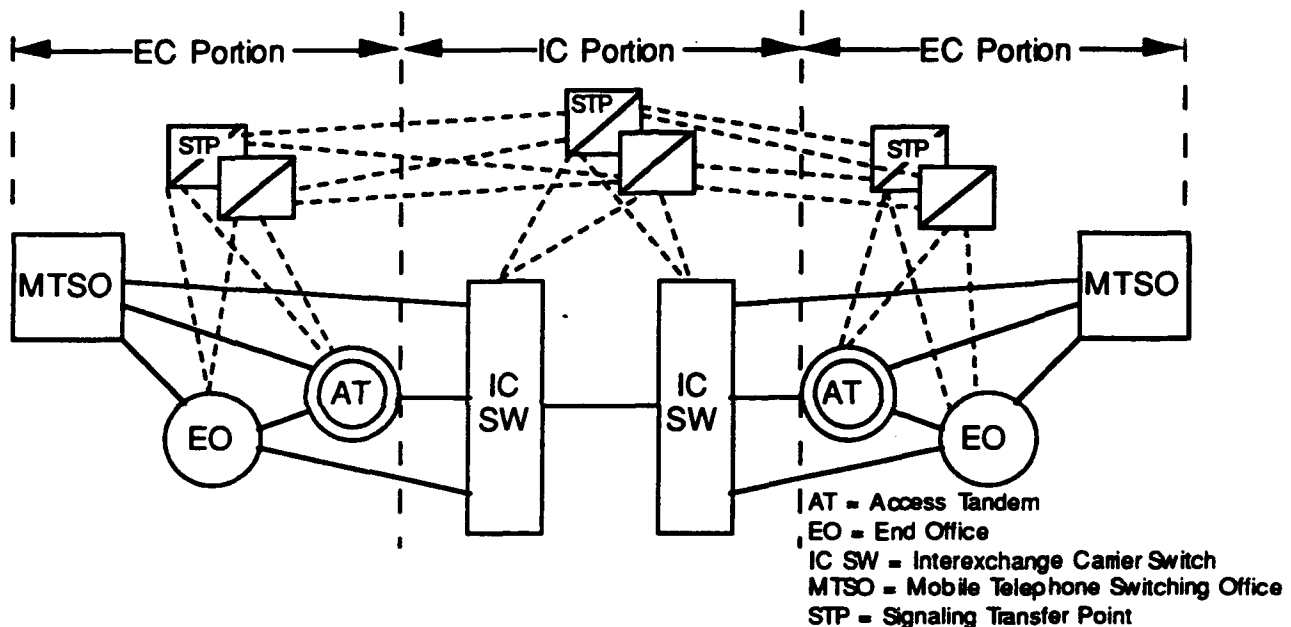


Figure 3.1: IC and EC Portions of a Connection in a Public Switched Network

4. Framework for Quantifying and Categorizing Service Outages

This section introduces a general framework for quantifying service outages from a users' perspective, applying it to three examples: circuit switched, packet switched and leased line networks. Before describing this framework, the term service outage is defined as follows:

A service outage is the state of a service when network failure(s) impair the initiation of new requests for service and/or the continued use of the service and

² Cellular carriers are considered to be both IC and EC carriers.

where the service outage parameters defined below do not fall into the "no outage" qualifying region.

4.1 Service Outage Parameters The three parameters of the framework are introduced below:

Unservability (U) A major impact of a service outage is service disruption to all or a portion of users. Unservability is defined in terms of a unit of usage. For example:

- In a circuit switched network, the most common function of the network is the ability to establish or maintain a connection from a source to a destination within the limits of engineered blocking and transmission performance. The unit of usage is a call attempt or call. In this instance, unservability is the percentage of units that fail.
- In a packet network, the unit of usage is a packet. The unservability is defined as the percentage of packets that were not delivered within the design delay.
- In a leased line network, unservability is defined as the percentage of service units failed (e.g., DS-0, DS-1 or DS-3 service units).

Duration (D) This parameter defines the time during which the service is unservable. It is measured by determining the beginning and the end points of a network failure, based on the performance exceeding the unservability threshold.

Extent (E) This parameter includes the context of geographic area, population affected, traffic volumes, and customer traffic patterns, in which the unservability exceeds a given threshold.

One can categorize service outages by different sets of values for which the (U, D, E) triple qualifies for the particular category of outage. If one were to graph the (U, D, E) triple in 3-space, then the values for a particular outage category would form a region, which we define as the *qualifying region*. For example, Figure 4.1 illustrates three simple qualifying regions. The (U, D, E) triple is the basis for quantifying the impact of service outages on users. Service outages may be classified as catastrophic, major, or minor, depending on whether the values of the (U, D, E) triple fall in the appropriate qualifying region. The different sets of values for the (U, D, E) triple that define each of these categories are under study.

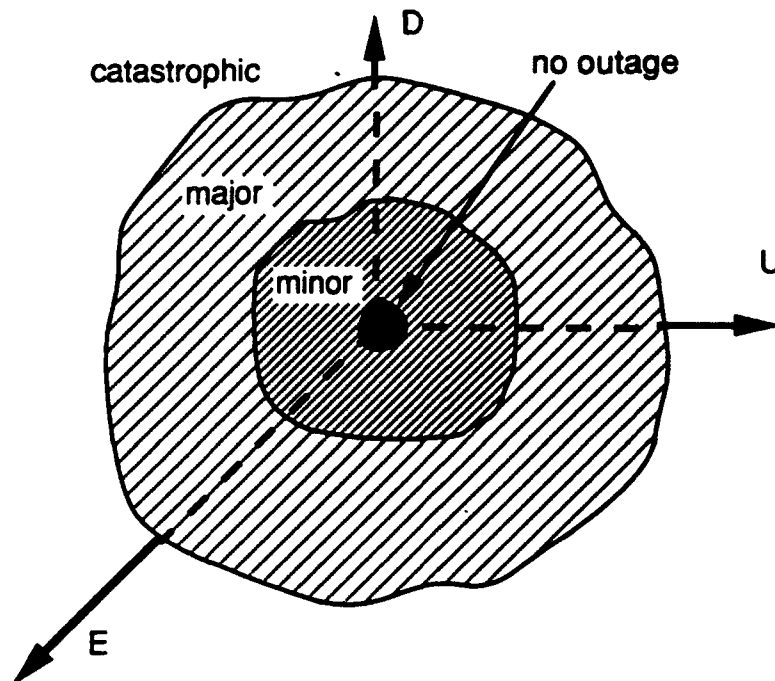


Figure 4.1: Example Qualifying Regions for (U, D, E)

Further investigation is needed to determine the impacts, if any, on the infrastructure of network elements, operations support systems and interfaces for measuring, collecting, and analyzing information to support the U, D and E parameters. If the network infrastructure is not initially able to provide the information, investigation into the identification of interim procedures may be needed.

4.2 Service Outage Categories This section generically defines and describes the service outage categories based upon the above framework.

Depending on the combination of a failure's U, D and E, the following three distinct service outage categories can be defined for both natural and human-caused failures (in order of decreasing severity):

- | | |
|----------------------------|---|
| <i>Catastrophic</i> | Possible causes: severe network node failure(s) ³ , caused by earthquake, flood, hurricane, tornado, power company grid failure, global hardware/software defect, act of war, etc. |
| <i>Major</i> | Possible causes: toll center, common channel signaling, network node failure caused by fire, terrorism, failure of cable backbone |

³ Examples include a critical Central Office (CO) failure, one or more STP pair failures, multiple EO, AT, or MTSO failures.

Technical Report No. 24

route, Digital Cross-connect System (DCS)⁴ failure, access tandem switch failure, CO power/battery failure, software defects, fiber cable failure, etc.

Minor Possible causes: single fiber failure, equipment failure (including shelf or unit).

The goal of a service provider is to restore interrupted service as quickly as possible. For catastrophic outages, affecting large populations, some service may be restored automatically, but most service will be restored manually. For major outages, a larger percentage of service may be restored automatically and the remainder of the service would be restored manually. For minor outages, service restoration can be automatic.

It also should be recognized that the users' expectations of uninterrupted service is tempered by the category and type of service outage. For example, in a catastrophic outage caused by a hurricane and affecting a very large number of users in one area, most users may recognize that service will be restored as quickly as is humanly possible. The time for total restoration of service for every user may take days to weeks. Future standards activity should specify quantifications of these service outage qualifying regions for particular services, networks, and users.

Considering the above framework for measuring service outages, Appendix C discusses users' expectations.

5. Framework for Classifying Network Survivability Techniques and Measures

This section discusses a four-layer framework (see Table 5.1) that allows for the classification of network survivability techniques in telecommunication networks⁵. Each layer can be characterized by a collection of networks (or subnetworks). Each of the networks consists of nodes, links, traffic or demand⁶, and control mechanisms and control networks. The links provide capacity to carry the traffic and the nodes provide points of access, switching, or routing for the traffic. The control mechanisms for each of these networks can

⁴ The term EDSX (Electronic Digital Signal Cross-connect) is often used for a DCS whose interface rate equals its cross-connect rate, e.g., a DCS-3/3 that cross-connects DS-3s. This distinction is not used here.

⁵ The network layers defined here should not be confused with layers defined in other more restricted contexts, such as the OSI layers for packet networks (see Section 6.4.2).

⁶ Throughout the rest of this document the terms "traffic" and "demand" are used interchangeably.

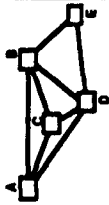
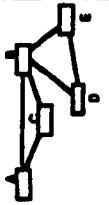
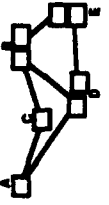
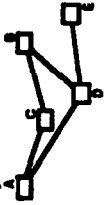
Layer	Function, Components	Type of Traffic	Survivability Role	Survivability Techniques	Control Mechanisms & Networks
Service 	provide public and private data and voice services nodes: switches (circuit, packet), signaling points, AIN node equipment links: trunks, data links, signaling linksets	telephone calls, data packets, cells (ATM), signaling units	mitigate the impact of service, logical, system and physical layer failures	size limits, routing (e.g., dynamic routing), reconfiguration (e.g., dual-homing, quad linkset arrangements)	Common Channel Signaling (CCS) networks, ATM congestion controls, packet protocols, packet congestion controls, protocol-based controls (e.g., SS7)
Logical 	meet the needs of each service using the physical capacity nodes: transmission equipment (e.g., MUXs, DCSs) links: channels (e.g., DS-0, DS-1, DS-3)	DS-0, DS-1, DS-3, SONET STS-n signals	reconfigure around logical, system and physical layer failures	restoration switching	network controllers, operations support systems, manual interfaces and distributed DCS methods
System 	provide bandwidth and diversity nodes: full-rate transmission equipment (terminals) links: full-rate channels (optical fiber, metallic, radio systems including satellite)	DS-3, SONET STS-n signals	reconfigure around system and physical layer failures	protection switching (e.g., self-healing rings, 1+1 diverse routing)	subnetwork controllers, operations support systems and manual interfaces
Physical 	provide geographical and media assets nodes: buildings, underground vaults links: rights-of-way, media	-	provide physical and electromagnetic protection	geographical diversity, building size limits, protection against physical hazards	automated power, security and fire control systems and manual intervention

Table 5.1: Network Survivability Layers

provide the functions of signaling, traffic routing control, traffic admissibility control, and data communications. The control mechanisms also use a control data communications network; for communication among control logic points and nodes. The control data communications network may either be separate (out-of-band) or coincident (in-band) with the traffic bearing network. The survivability of the control data communications network itself also should be studied and evaluated.

In this representation, the link capacity of the service layer in turn forms the traffic for the logical layer, whose link capacity in turn forms traffic for the system layer, etc. Nodes and links of the system and logical layers represent transmission equipment and provide the bandwidth required for the service layer links. Service layer nodes represent the switching or processing equipment that uses this bandwidth to provide user service networks.⁷

Network survivability techniques can be deployed at various layers of this framework to ensure that failures interrupt end-users as little as possible. Failures within a layer can be guarded against by techniques either in that layer or at a higher layer⁸. Beginning with the physical layer, the function of each layer is described below. Sample network survivability techniques representing network based solutions for link and node failures for each layer are described in Section 6.

5.1 Physical Layer The physical layer can be viewed as having two sublayers: the geographical sublayer and the media sublayer. The geographical sublayer consists of the assets that house the media and other equipment associated with the upper layers, e.g., buildings, conduits, manholes and rights-of-way. The media sublayer contains the transmission media used by the transmission systems, e.g., fiber cable, copper cable and radio spectra.

Figure 5.1 illustrates an example physical layer subnetwork. The links of this example network represent fiber cables (media) that in turn route over conduit sections (not shown) between the buildings or possible repeater site locations.

⁷ The service, logical, system and physical layers defined here correspond roughly to the circuit, path, section and physical media layers, respectively, used in CCITT G.803.

⁸ For example, a cable cut may be repaired or restored in the physical, system or logical layers. Otherwise, it can be restored by the service layer. In contrast, a switch failure at the service layer can only be restored in the service layer.

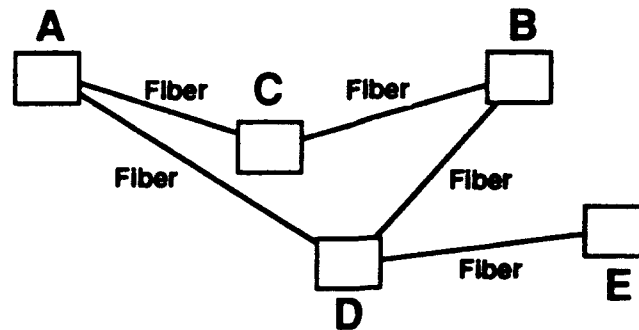


Figure 5.1: Example Physical Layer Subnetwork

5.2 System Layer The system layer represents the network transmission systems, and the system terminating and full-rate interface equipment. Network transmission systems include both cable-based (fiber and metallic) and wireless technologies (e.g., cellular, Personal Communications Network (PCN) and radio). Network survivability techniques at the system layer reconfigure the network components to protect the logical layer, and thus the service layer, against physical or system layer failures. These techniques combine system diversity with geographical diversity and reserve capacity to allow bandwidth lost by failures to be restored.

Figure 5.2 illustrates a subnetwork at the system layer. It consists of a SONET [44,38] OC-48 (2.448 gigabits/second) self-healing ring [40,42] and two SONET point-to-point OC-12 (622 megabits/second) fiber systems. The capacity over the links of this network becomes demand (traffic) for the physical layer example network shown in Figure 5.1. For example, the B-E OC-12 fiber system shown in Figure 5.2 routes over two fiber cable links (node pairs B-D and D-E) in the physical layer example of Figure 5.1.

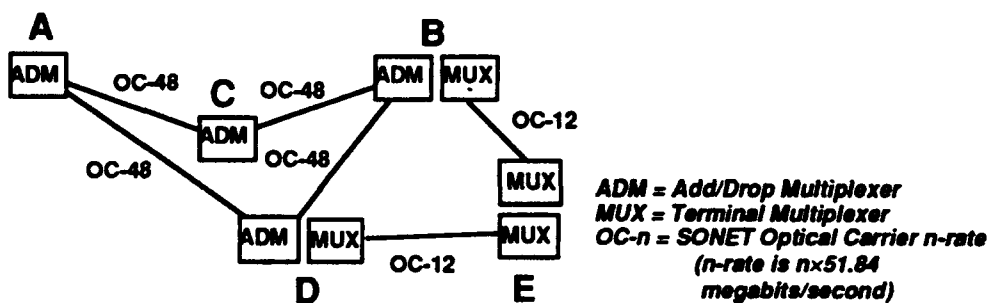


Figure 5.2: Example System Layer Subnetwork

5.3 Logical Layer The next layer of the planning framework is the logical layer, in which lower-rate transmission channels such as DS-0s, DS-1s or DS-3s and their interface equipment appear. Using the full-rate transmission channels of the system layer, the logical layer provides channels for use by the service

networks in the service layer. After a failure is detected by the logical layer, network survivability techniques act to preserve the logical network that the service layer uses.

There are many possible distinctions between the traffic characteristics of a network at the logical layer and a network at the service layer. Traffic holding times are one such distinction. For example, traffic for a wideband DCS network (a *wideband* DCS interfaces digital signals at the DS-3, or STS-n rate, and cross-connects⁹ the constituent channels at the DS-1 or SONET Virtual Tributary (VT) group rate; a *broadband* DCS interfaces digital signals at the DS-3 or STS-n rate and cross-connects the constituent channels at the DS-3 or STS-1 rate) at the logical layer are DS-1s that represent private line demand or multiplexed trunks of a circuit switched network. Here, traffic holding time represents the time between connection and disconnection of a DS-1, which could range from days to years. Also, this traffic does not have easy re-connect ("redial") capability as in a circuit switched network. In contrast, the traffic in a circuit switched network at the service layer is calls, whose holding times are typically short – in the range of minutes.

The example in Figure 5.3 shows a logical layer subnetwork of wideband DCSs. Here, the links represent families of DS-3 or SONET STS-1 capacity between the DCSs. These capacities become demand (traffic) for the example system layer subnetwork in Section 5.2. For example, the DS-3s between nodes A and B route over nodes A-C-B of the SONET self-healing ring in the system layer example subnetwork. Note, although not illustrated in this simple example, there can be further layering of subnetworks within the logical layer (e.g., the link capacities of a subnetwork of wideband DCSs may route over a subnetwork of broadband DCSs).

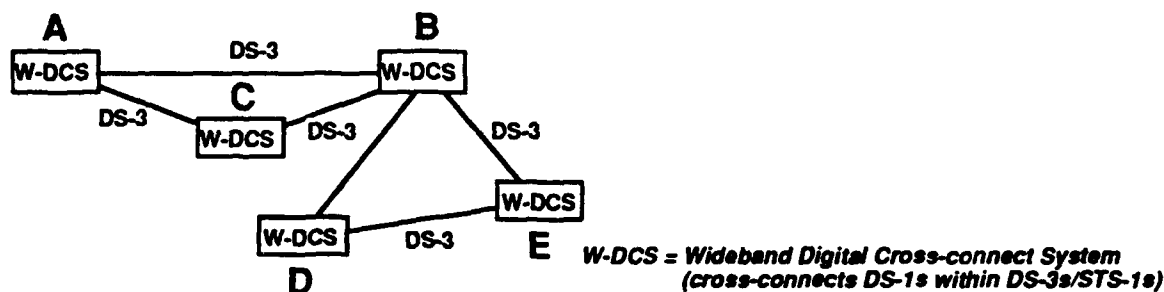


Figure 5.3: Example Logical Layer Subnetwork

⁹ Cross-connecting is the process of connecting between terminal blocks on the two sides of a distribution frame, or between terminals on a terminal block. The same connection is also made through software implementations of this process by a DCS.

5.4 Service Layer The top layer of the planning framework is the service layer. This layer contains user service networks such as public and private data, voice and CCSNs. Survivability techniques offered in this layer can include:

- size limits,
- dynamic routing,
- message retransmission,
- multi-serving, and
- reconfiguration (reconnection and redirection).

Service layer rerouting and reconfiguration techniques would normally be activated for capacity that is not restored by techniques invoked at the system or logical layers. Service layer techniques form the last safeguards before failures become apparent to the user.

Figure 5.4 illustrates an example service layer subnetwork. This represents a portion of a circuit-switched network. Here, the links represent trunk groups (with integrated DS-1 interfaces) between circuit switching systems. These capacities become demand (traffic) for the example logical layer subnetwork in Figure 5.3. For example, the direct trunks between switches A and D form DS-1s that route over nodes A-B-D in the logical layer. The demand (traffic) transported by this example service layer subnetwork are normal "telephone" calls, often measured as "traffic load," (in units of Erlangs or hundred call seconds).

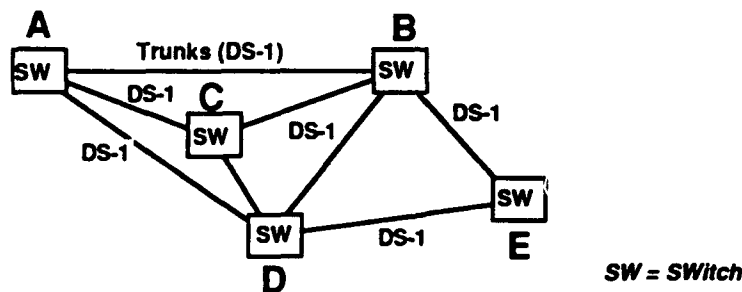


Figure 5.4: Example Service Layer Subnetwork

5.5 Summary The increased impact of failures has led to a greater awareness of the need to design survivable networks. There is a need for a common framework and terminology for discussing and comparing the wide variety of network survivability techniques that are available. The four-layer model described above provides such a framework. Table 5.1 summarizes this framework and lists some of the techniques that can be applied at each layer to minimize service interruptions to users.

This model provides a method of categorizing the various network survivability techniques in terms of the layers at which a network responds to failure. In

addition to providing a common basis for describing and comparing techniques, this framework identifies the layer(s) responsible for reacting to the various types of failures and their interaction. Failures within a layer can be guarded against by techniques either in that layer or at a higher layer. To protect the service users, network planners need to select effective combinations of network survivability techniques across the layers for incremental deployment.

6. Network Survivability Techniques

This section contains short summaries of various techniques that provide for survivability of telecommunications networks. These techniques are generic and are based on the four network layers described in the framework of Section 5.

Basic Steps for Restoration

As illustrated in Figure 6.1, network restoration can be accomplished through six basic steps in response to a network failure within the four network layers:

1. *Detection*: Detection of the network failure by the appropriate Operations Support System (OSS), controller, or network element.
2. *Notification*: Notification of the network failure through the control architecture. Under a centralized scheme this means notifying the central operations support system or controller of the failure, either from the network element directly or via another operations support system for network surveillance. Under distributed control; this means notifying the other network elements via the links connecting network elements or a separate data communications network.

After the notification step, two parallel processes begin: (i) repair of the failed components, and (ii) reconfiguration of affected traffic. These steps can be applied to implementing most network survivability techniques, although some of the steps may not be necessary. Also note that some combinations of the steps may be iterated. For example, the path selection and rerouting steps may be iterated for each demand that is rerouted.

- 3a. *Fault Isolation*: Under the repair process, (see Figure 6.1), fault isolation is the identification or location of the network failure.
- 3b. *Identification*: Under the reconfiguration process, the affected demands are identified for rerouting.
- 4a. *Repair*: The repair process is the effort to repair the network failure to allow normalization of the network.
- 4b. *Path Selection*: Path selection is the selecting of an alternate path for each demand to be rerouted, either through data look-up or algorithm.

5. *Rerouting*: Rerouting is the moving of the affected demands to an alternate path.
6. *Normalization*: Normalization is detection of the repair of the network failure and the return to the normal state of the network. This normalization could include rerouting of affected demands to new or original routes.

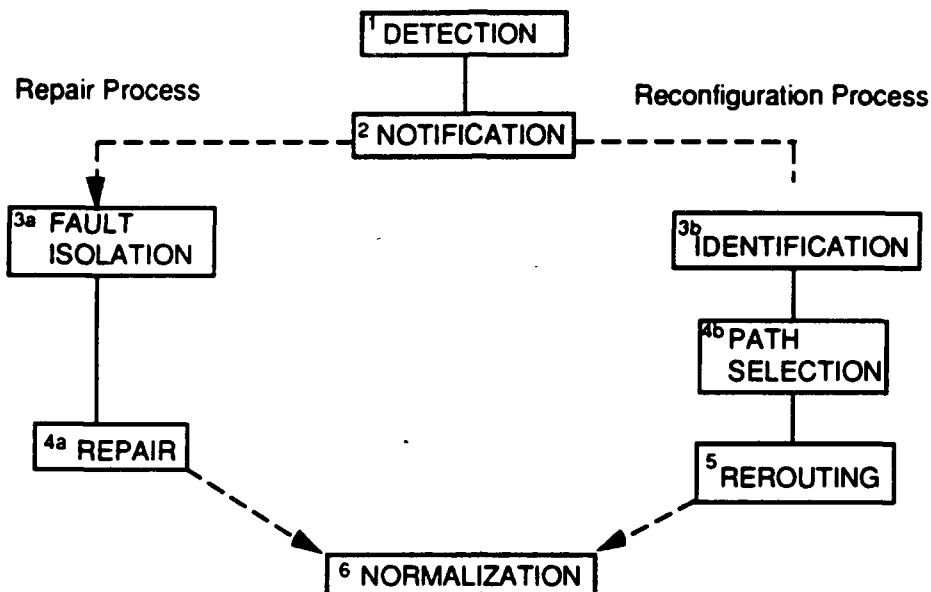


Figure 6.1: Basic Steps for Restoration

Determination of alternate paths in the path selection step can be of two types: precalculated or determined dynamically based on path status. A precalculated method of alternate path selection specifies that for each demand that must be rerouted, a preplanned, reserved alternate path with spare capacity is identified. For example, self-healing rings (Section 6.2) use precalculated methods. Status based methods respond using current network status knowledge to select alternate paths. For example, DCS reconfiguration methods (Section 6.3) often use status-based methods.

The two general approaches for status based path selection methods are preplanned lists of paths stored as local routing tables and dynamic path generation, where the paths are created at the time of failure.

The following network survivability techniques for each layer may utilize paths established via either wireline (cable) or wireless technologies. Wireline technologies consist of copper cable (coaxial or twisted pair) and fiber optics. Wireless technologies include terrestrial radio, satellite, cellular, and the evolving PCNs. Each of these technologies brings to the network survivability planning process its own limitations. Recognizing that a limited capacity system may not be

adequate by itself to fully restore disrupted services, when used in conjunction with other surviving technologies, each could play an important role in minimizing disruptions caused by network failure.

6.1 Physical Layer Physical layer survivability falls into three categories: (i) geographical diversity and flexibility, (ii) security to human-caused intrusion, and (iii) tolerance, i.e., the ability of components to resist human-caused or natural disasters.

Geographical diversity is a technique whereby pairs of buildings are connected via multiple paths that do not share the same locations or rights-of-way and where critical network systems are spread among different locations. To enhance network survivability, network and service providers need to ensure sufficient redundant and diverse facilities of technologically similar media are in place to transport traffic and maintain uninterrupted services should the primary media become damaged. Where technologically similar media are not available, the restoration links need to be equipped to provide the same level of transport as the primary.

Given the higher bit rate and increased capacity of optical fiber cables, restoration of traffic transported via this medium must be moved to other facilities with the same capacity. This suggests that the fiber diversity route must be technologically similar and of equal capacity to restore the affected demands while not overtaxing the unaffected media. It also suggests that the restoration routes could be a combination of technologically similar media which, when combined, has sufficient capacity to restore the affected demands.

Security of the physical layer should be a major point of consideration. Increasing and maintaining a high level of physical security is necessary to ensure protection from damage caused by persons intent on disrupting telecommunications services.

Physical layer techniques which enhance building and telecommunication systems design availability do so by increasing their ability to tolerate external and environmental effects. For example, good fire detection and suppression systems are well-known methods to limit the degree of damage suffered by network assets.

Standards which address some of these areas have been developed or are under development in T1E1:

- earthquake protection and testing (project approved),
- central office telecommunications equipment ambient temperature and humidity (ANSI T1.304-1989),
- DC power systems environment protection (ANSI T1.311-1991),

- central office equipment electrostatic discharge (ANSI T1.308-1991),
- electrical protection for central offices and facilities (ANSI T1.313-1991), and
- telecommunications equipment ignitability requirements (ANSI T1.307-1990).

6.2 System Layer System layer survivability is enhanced by considering techniques which support reliable connectivity over existing surviving links. These include automatic cutover systems, link diversity, spare or excess capacity, coding, spread spectrum transmission, and error correction in addition to the following.

6.2.1 Point-to-Point Systems with Automatic Protection Switching

Automatic Protection Switching (APS) is a transport network survivability technique where both ends of a working channel's termination equipment switch to a full bandwidth protection channel in the event of link failure (see Figure 6.2). APS can either have one protection channel¹⁰ for each working channel (1+1 or 1:1), or have one protection channel for one or more (n) working channels (1: n). The working and protection channels may make use of diverse routing for added network survivability. A point-to-point system with diverse routing does not provide individual channel add-drop capability at intermediate locations and hence another protection architecture must be used to provide this type of application.[27].

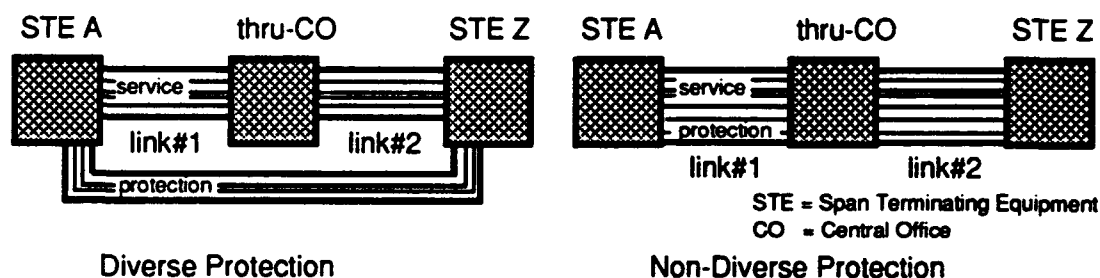


Figure 6.2: Diverse vs. Non-Diverse Protection

6.2.2 Rings A ring is a collection of nodes forming a closed loop whereby each node is connected to two adjacent nodes via a duplex communications facility. A ring provides redundant bandwidth and/or network equipment so disrupted services can be automatically restored following network failures.

¹⁰ 1+1 systems transmit on the protection channel at all times. 1:1 systems transmit on the protection channel only after the failure (the protection channel is normally idle).

Rings can be divided into two general categories, unidirectional and bidirectional, according to the direction of traffic flow under normal conditions. In the SONET application [40,42], unidirectional and bidirectional architectures are defined based on the direction in which incoming and associated returning tributaries travel around the ring. An incoming tributary is the signal added into the ring, and the returning tributary is the associated signal dropped from the ring.

6.2.2.1 Unidirectional Ring In a Unidirectional Ring (UR), working traffic is carried around the ring in one direction only (e.g., clockwise). Refer to Figure 6.3. Traffic from any node A to any other node B is routed along the working communications ring from A to B via D and C, and the return traffic continues around the ring from B back to A, in the same direction of transport using the remaining portion of the working communications ring. Thus, traffic arrives at nodes A and B by different routes. Because the transmission of normal working traffic on the UR is in one direction only, ring capacity is determined by the sum of demands between each pair of nodes. URs are sometimes called “counter-rotating rings” because the second communications ring (for protection only) transmits in the opposing direction as the first. This is also known as 1:1 protection, or 1:1 UR.

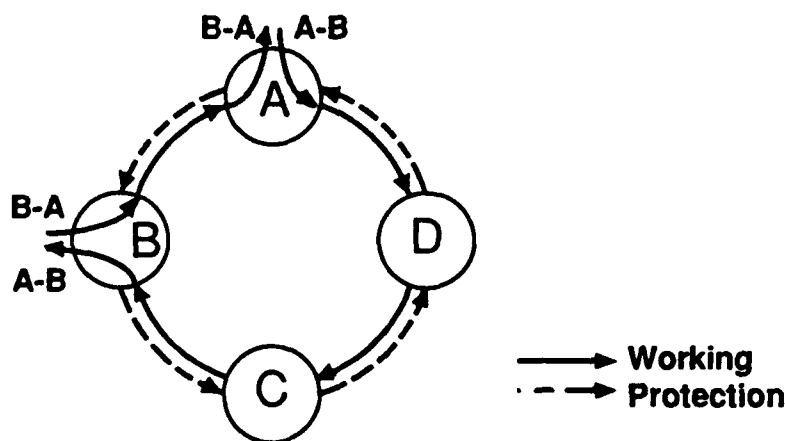


Figure 6.3: Unidirectional Ring

6.2.2.2 Bidirectional Ring In a Bidirectional Ring (BR), working traffic travels in both directions over a single path that uses the two parallel communications paths (operating in opposite directions) between the nodes of the ring (e.g., between A and B), hence the name BR. Since traffic is routed over a single path between nodes, spare capacity around the ring can be shared on a per span basis (i.e., demand carried per span), and not dedicated to the total demand on the ring as for a UR. In a BR, traffic is often routed over the shortest path and is identified during provisioning. This is referred to as a shortest path algorithm or nondemand splitting. (For optimal load balancing algorithms, however, some traffic may be routed over the longer path.) Nondemand splitting basically means that the total demand between any two points on the ring travels the same route

(shortest path), and is not normally split between two different routes (i.e., split between shortest and longest paths). BRs are engineered based on the span carrying the maximum amount of traffic. A BR may use four fibers or two fibers depending upon the spare capacity arrangement, as shown in Figure 6.4.

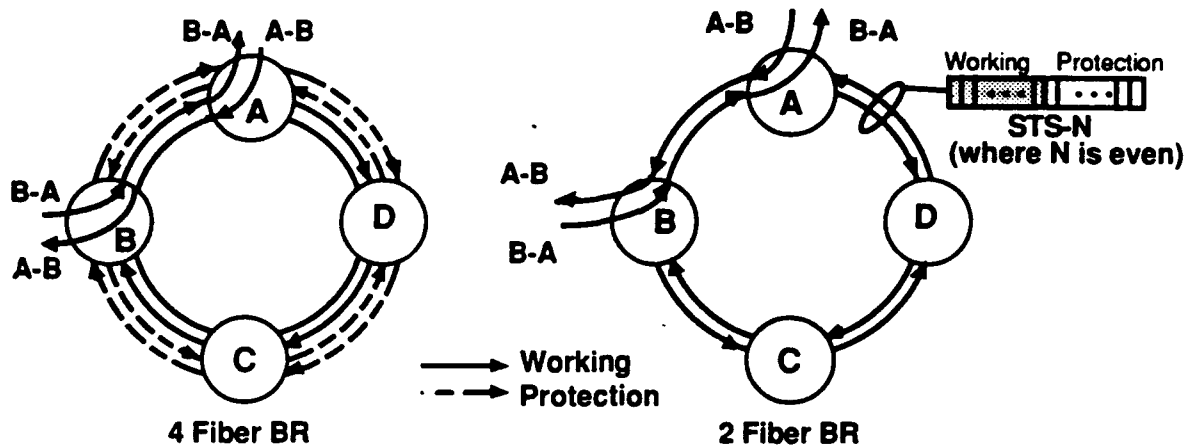


Figure 6.4: Bidirectional Rings

In a four fiber BR (or 1:1 configuration), working and protection channels use separate communications paths (fibers). The 1:1 configuration has evolved from today's point-to-point protected systems and can provide line protection switching using similar SONET automatic protection switching procedures.

In a two fiber BR, working and protection channels use the same two parallel communications paths (i.e., reserve half bandwidth for protection). This ring arrangement can provide line protection switching by using a time slot interchange method for merging working channels with protection channels.

Unidirectional and bidirectional rings can be further categorized into line and path protection switched rings according to the SONET layer used to trigger the protection switch action that enables the ring to automatically recover from failures. Bidirectional path switched rings are not discussed here.

6.2.2.3 Line Protection Switched Ring A line protection switched ring architecture uses SONET line layer indications to trigger the protection switching action and may be unidirectional or bidirectional. Switching action is performed at only the line level to recover from failures, and does not involve path level indications. Line level indications include line level failure conditions and signaling messages that are sent between nodes to effect a coordinated line protection switch.

An example of a unidirectional, line protection switched ring (UR/L) is a "folded," UR implementation, that folds (or loops) the disrupted line signal (i.e., the optical

line signal or n line SPEs¹¹) onto a separate protection fiber ring (1:1 concept). This example is shown in Figure 6.5. The nodes adjacent to the break (nodes A and B in Figure 6.5) perform the fold or looping function. Because of the loopback capabilities at the nodes adjacent to the break, the folded ring remains in a ring topology.

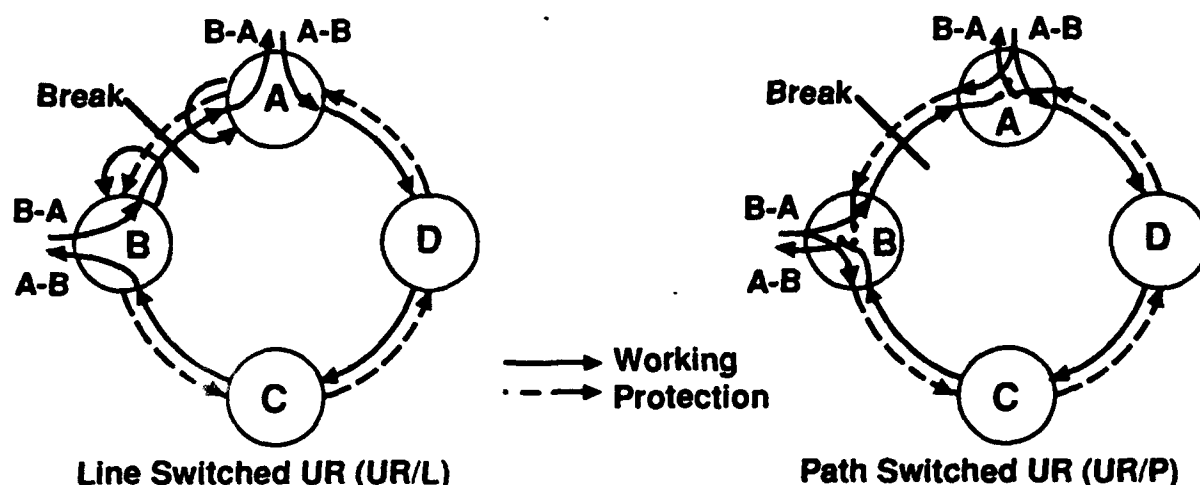


Figure 6.5: Unidirectional Ring Implementations

An example of a bidirectional, line protection switched ring is a two fiber BR (BR/2F) implementation, and is shown in Figure 6.6. In this implementation, protection capacity is provided around the ring and is available to restore disrupted traffic during protection switch conditions. This protection capacity (i.e., channels) is shared by all spans (i.e., between adjacent nodes) on the ring. In the case of a fiber cut, the nodes adjacent to the break set up bidirectional paths (loopback functions using Time Slot Interchange (TSI)) on the protection channels in the opposite direction for only those services affected by the break. The loopback function is for looping back disrupted traffic between nodes A and B from working to protection channels. As a result, all traffic that was normally on the working channels on the link between nodes A and B is now on the protection channels in the reverse direction around the ring. The protection channels can then be reused by other spans when the protection channels become available again.

¹¹ Synchronous Payload Envelopes.

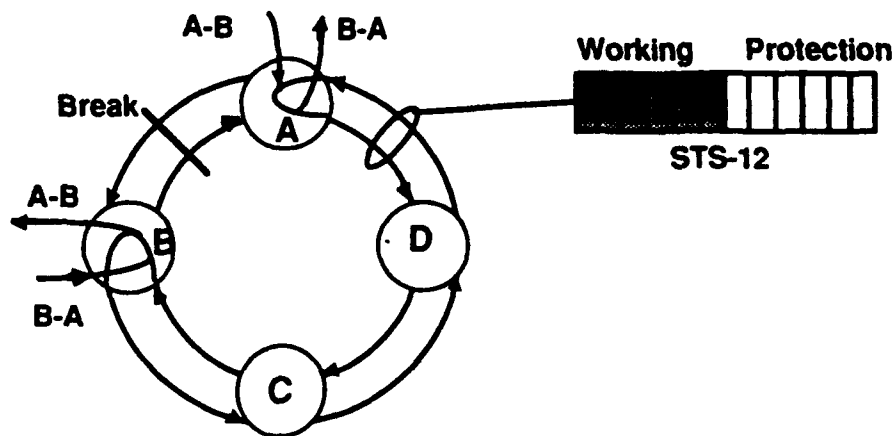


Figure 6.6: Bidirectional Ring Implementation

A four fiber BR (BR/4F) implementation uses a line loopback recovery technique, but unlike the BR/2F implementation, working and protection channels are routed on separate communications paths (fibers). Another BR/4F implementation uses a combination of line loopbacks and automatic protection switching span protection. The line loopback function protects against cable cuts, and the span protection safeguards against equipment failures.

6.2.2.4 Path Protection Switched Ring A path protection switched ring architecture uses SONET path layer indications¹² to trigger the protection switching action and may be unidirectional or bidirectional. Switching action is performed at only the STS or VT path level to recover from failures, and does not involve line level indications. Path level indications include path level failure conditions and, if necessary, signaling messages that are sent between nodes to effect a coordinated path protection switch. Path switching of a specific path is independent of other paths' status.

In some unidirectional, path protection switched rings (UR/P, see Figure 6.5), both directions of working traffic between any node A and any other node B travel on a single fiber ring in the same direction (e.g., clockwise), but arrive at their destinations by different paths (i.e., the two directions of traffic use the opposite portions of the single ring). A second protection fiber ring carries a duplicate copy of the information signals in the opposite direction. This is considered a "dual-fed" or "1+1 protection" approach when duplicate information is transmitted in both directions on the ring. Thus, at a receiving node (e.g., node A), two signals (one from each direction) are available for signal selection. In the case of a break in the ring (between node A and B), receiving node A switches to the protection path for

¹² DS-1 or DS-3 Alarm Indication Signals (AISs) are also used for this purpose.

path restoration. Receiving node B was not affected by the break, so no path switching is necessary. For each disrupted path, path switching requires identifying and selecting the valid (or protection) path for path restoration. The information for the switching control is conveyed within the path signal itself (e.g., Path AIS). It should be noted that, unlike common approaches to ring implementations which maintain their ring characteristics following a fault, this implementation ceases functioning as a ring if a break in the ring occurs, i.e., it does not remain in a ring topology (no loopbacks) following a break in the ring or the loss of a node. However, communications among the nodes is maintained following such a break.

6.3 Logical Layer The logical layer reestablishes active demands that route over failed network links or nodes, as contrasted with a circuit switched network in the service layer, where existing demands are lost (and must redial). This requirement imposes a more difficult network survivability routing task on the subnetworks at this layer, as well as more critical restoration intervals. Restoration switching, as implemented in DCS based reconfiguration strategies, is described below.

6.3.1 DCS Reconfiguration Strategies The routing capability of DCSs can be used to restore service in case of network failures. By changing cross-connects, DCS reconfiguration methods restore service by routing existing, failed demands (e.g., DS-1, VT 1.5, or STS-1 systems) on alternate routes. At present, DCS reconfiguration methods operate on a network of either wideband DCSs or a network of broadband DCSs. However, because of the smaller number of demands to reroute, most DCS reconfiguration schemes are targeted for broadband DCS networks.

There are many methods being developed for DCS reconfiguration. In this section we generically describe the steps of a DCS reconfiguration scheme as well as present the general characteristics of the different approaches.

DCS repair and reconfiguration processes follow the basic steps identified at the start of Section 6 in response to a network failure. For each of those steps, there are a variety of implementation techniques. However, a fundamental characteristic that applies to each step is the method of control: under *centralized* control, the step is principally done by an OSS or "controller," which we will refer to generically as a "centralized system," while under *distributed* control, the step is done by the individual DCSs. Note that there are methods that mix some steps under centralized control with other steps under distributed control (e.g., distributed control of the path selection step with centralized control of the normalization step).

Another fundamental characteristic for DCS reconfiguration is the method for data communications. For centralized control this involves communication between the DCSs and the centralized system(s). For distributed control this involves communication among the DCSs. The most critical aspect of data communications for DCS reconfiguration is the guarantee of a secure data channel. This can be described succinctly by the following requirement: for a given demand, if there exists a surviving, alternate route, then data communication is required among all DCSs along that route (to enable rerouting of the demand). Therefore, in general, communication between DCSs and the centralized system(s) must occur over physically redundant routes. For distributed control, communication between DCSs can use SONET overhead (or related signal overhead) along all the possible inter-DCS "logical" links. This aspect guarantees that if there exists a surviving, alternate route between two DCSs, then communications will occur along this route and, thus, the requirement is satisfied.

We discuss various general characteristics of DCS reconfiguration schemes for each step below:

6.3.1.1 Detection and Notification Under centralized control, DCSs (or, possibly, higher level network elements) detect and notify the centralized system of the failure, possibly via another OSS for network surveillance. It is possible that these alarms are received at a variety of multiplex levels and the centralized system has to "locate" and analyze the fault, i.e., deduce which channels are affected (e.g., which STS-1s or DS-1s are affected). Under distributed control, the DCS monitors each of its channels and analyzes the faults; although, for repair purposes, an alarm may still be sent to the network surveillance OSS.

6.3.1.2 Identification Under centralized control, the centralized system identifies the fault and instigates the reconfiguration process. It may have various thresholds to screen out false or intermittent alarms. Under distributed control this step means identifying the failure to the DCSs and triggering the event. When a failure is identified between a pair of DCSs, this implies some sort of "handshaking" or prearranged understanding so that only one of the DCSs instigates the reconfiguration process.

6.3.1.3 Path (Route) Selection . Methods for implementing this step can generally be described by the three following characteristics:

i) Control: centralized or distributed. Centralized control means that a centralized system calculates the alternate routes, i.e., the affected DCSs and channel assignments. Distributed control means that software in the DCSs calculates the routes, often in a parallel manner among all the DCSs. It is possible to have combinations of both: the centralized system determines higher level routes or routing guidelines and the network elements choose specific routes. For example,

the DCSs can use routing tables that are populated by a centralized system, an architecture that is similar to that used in switches in a circuit switched network.

ii) *Type of alternate routing: link or point-to-point*¹³. With *link* rerouting, all demands routed over the failed link have the segment of their route corresponding to the failed link replaced by an alternate route. Link rerouting is illustrated for a wideband DCS subnetwork with DS-3 links in Figure 6.7. Here the route of a sample DS-1 demand is shown by a dashed line. When the B-C DS-3 link fails, alternate DS-1 routes are found between nodes B and C. The "segment" of each DS-1 route that originally routed through the failed link is cross-connected to the alternate route, sometimes called a "patch". For the sample DS-1, the alternate route goes through nodes B-A-D-C. The result is that at the DS-1 level there is retracing of the route. With point-to-point rerouting, constituent demands are rerouted from end node to end node. This is illustrated in Figure 6.8 where the DS-1 is routed directly over the A-D DS-3 link.

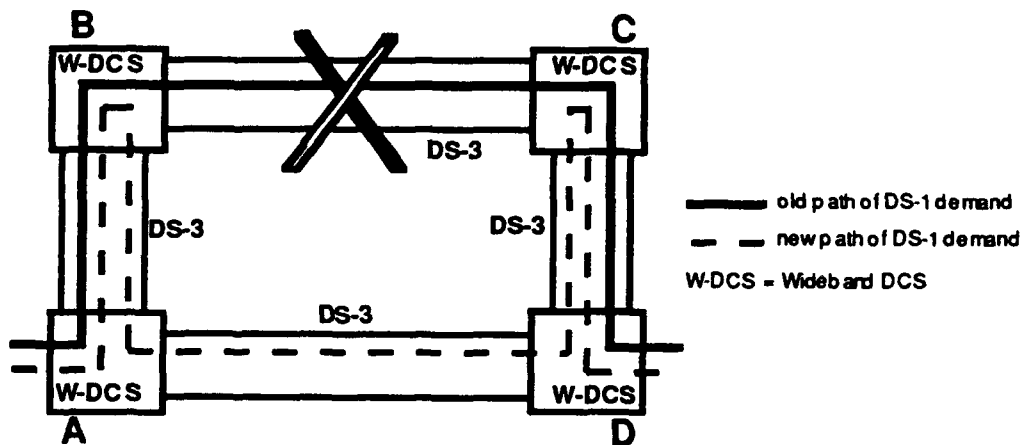


Figure 6.7: Link Rerouting from Link B-C to B-A-D-C

The principal advantages of link rerouting are speed and ease of implementation because the DCS needs to find alternate routes between only one node pair, e.g., nodes B and C in Figure 6.7, and the various end points of the demands that route through the link do not have to be determined. However, it is complex to adapt some link rerouting methods for multiple link failures and, especially, node failures.

¹³ Many other references refer to point-to-point rerouting as "path" rerouting (for example, see [16]). However, since these references also use the term "alternate path," even for link rerouting methods, "point-to-point" is a less ambiguous term.

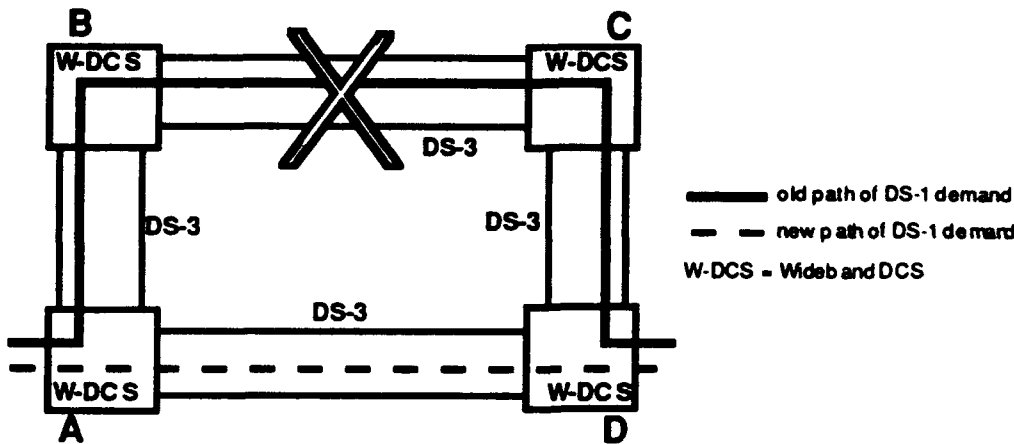


Figure 6.8: Point-to-Point Rerouting of A-D DS-1 Demand over Spare Facilities

Point-to-point rerouting requires determination of individual demands and their end points and an associated route selection process for each such end point pair. Point-to-point rerouting schemes tend to be more efficient from the standpoint of utilization of network capacity, i.e., they require less network spare capacity to achieve the same level of network survivability. Another advantage is that some implementations allow the capability to selectively choose which demands to reroute in case of failure. Also, because demands are rerouted between their end points, point-to-point rerouting techniques can be more readily adapted to handle node failures. However, the complexity of point-to-point rerouting techniques tends to make them slower than link rerouting techniques.

iii) Route calculation process: precalculated or dynamic. With precalculated methods the alternate routes are stored in a database (either in a centralized system or locally in the DCS) and retrieved when the restoration process is instigated. If a precalculated method stores only one route, then for each demand that must be rerouted, a single, alternate route is identified and reserved. If multiple routes are stored, then they can be tried in sequence until a functioning route (with spare channels) is found. Precalculated methods can suffer from the difficult task of satisfying physical diversity requirements when preplanning the alternate routes, as well as database integrity issues. Precalculated schemes also suffer from the problem of keeping the data current under a changing network. The main advantage of precalculated schemes is simplicity of implementation and speed of execution.

Dynamic routing methods use current network status knowledge, i.e., link/node operational status (failed or working) and associated link spare (also called "occupancy" or "utilization") to calculate the alternate routes at the time of failure. An advantage of status based methods is that failed links or nodes can be readily

identified at the time of failure and avoided in the selection of alternate routes. This simplifies the computational burden of diversity requirements faced by predetermined methods. However, dynamic routing methods are more complex to implement and can be slower to execute.

Most methods using distributed control for the path selection step implement dynamic routing via various adaptations of a method called "flooding." In this method for a given pair of DCSs (say, A and Z) between which alternate paths are sought, one of the DCSs instigates the restoration process (called the "sender"; say DCS A) by sending out requests for alternate routes to each of its neighboring DCSs, which in turn send out requests to each of their neighboring DCSs, etc. Thus, routes are pieced together a link at a time. As route requests reach the destination (DCS Z), confirmations are sent back along the route. To control the process, guarantee a good selection of routes, and avoid cycling, the DCSs use information such as link counts to retain efficient routes and discard the others.

6.3.1.4 Rerouting Under centralized control, demands are rerouted when the centralized system sends specific cross-connect commands to the DCS over the data communication network (indicating which channels to cross-connect). Under distributed control, the DCS instigates its own cross-connects.

6.3.1.5 Normalization This is one of the most difficult aspects of DCS reconfiguration methods. The major difficulty is that after the failure is repaired (e.g., a fiber cut is re-spliced — possibly days after the initial failure), it is difficult to reroute demand back onto the failed link without incurring further interruption of service. In fact, if not carefully executed, the normalization step could potentially incur a longer disruption of service than the initial failure event. Where adverse impact associated with normalization is unavoidable, this impact can be minimized by planning and coordination with the users of the services.

There are some arguments to leave the rerouted demands alone; i.e., unchanged. However, this option is not likely to gain support from network planners because the reconfiguration process tends to use up most of the network spare capacity and, in addition, uses long and inefficient alternate routes. If the demands are not rerouted after the failed link is repaired, new connections would also tend to be routed inefficiently in the attempt to use up the idle spare on the newly repaired link. Since one of the advantages of DCS reconfiguration is its capacity efficiency, this practice would obviate this advantage. To understand the potential magnitude of this, consider a 144 fiber cable that carries OC-48 systems. A full cut on this cable can easily involve hundreds of active STS-1 channels (a maximum of 3,456 STS-1 channels).

6.3.1.6 Summary of DCS Reconfiguration Methods Beyond implementation issues, the various methods can be evaluated in terms of speed of

reconfiguration and the efficiency of path selection. In terms of speed of reconfiguration, while there are simulation studies of the times of the path selection step for various methods, at present there are few observed execution times in real networks. However, a generalization for broadband DCS reconfiguration is that methods that use distributed control for the rerouting step have restoration time on the order of seconds while those that use centralized control have a restoration time on the order of minutes.

Methods that use centralized control for all steps are found in [10] and [11]. Methods that use distributed control for all steps (except normalization) and dynamic route calculation are given by [12-16]. A method that uses distributed control for the detection, signaling and identification steps, centralized control for the path selection step (precalculated routes), and distributed control for the rerouting steps is given in [17].

6.4 Service Layer The service layer provides public and private data and voice services. The survivability techniques of this layer mitigate the impact of uncovered failures at the logical, system, and physical layers as well as failures and degradation at the service layer itself (e.g., switch failures).

6.4.1 Circuit Switching

6.4.1.1 Size Limits A preventative technique to mitigate the severity of service outages is to limit the size or capacity of a part of a circuit switched network to ensure that its failure will not overly diminish the overall grades of service. Limiting capacity loss typically involves limiting the number of lines carried by (and thus user dependence on) any one switch or transmission system.

6.4.1.2 Dynamic Routing Dynamic Routing is the evolutionary result of traditional switched-network hierarchical-routing structures. Using dynamic routing, a service network can reroute workloads around failures in the network without requiring changes in the underlying logical network. It is most appropriately based upon service demands, and is an enhancement of simple alternate routing. A key attribute of dynamic routing techniques is that the path choice is determined at the time of the call attempt. The choice is based upon a relatively short network status update interval calculation which determines networking path strategy using link occupancy, optimum route sets and changing network conditions. In the event of a service outage, dynamic routing reroutes new service requests around the affected area. The established service connections in the affected area are disrupted and not rerouted.

Dynamic routing has the potential of providing the survivability advantages of link failure and tandem node failure protection in addition to network routing efficiency

benefits. Descriptions of various dynamic routing methods for circuit switched networks can be found in [29-32].

6.4.1.3 Reconfiguration Reconfiguring the service network includes options such as dual-homing and internal rearrangements in private networks. After a failure, dual-homing reconfigures a service network by redefining the home switch for both incoming and outgoing traffic destined to the affected service area.

6.4.1.4 Network Management When two or more network elements in an exchange carrier and/or interchange carrier network are experiencing degraded service, affecting performance for a set threshold time, manual network management control methods such as call gapping (where a limited number of calls is allowed in a given time interval), CAnCel To (CANT, which disallows calls to a designated switch), CAnCel From (CANF, which disallows calls from a designated switch), SKIP (which causes affected traffic to skip the controlled trunk group) and reroute can be used. See [49].

6.4.1.5 Multi-Serving Techniques that enhance survivability at the Service Layer include:

- wireless technology,
- dual homing (two or more serving central offices for a given serving area) or split homing (providing service from two or more nodes)
- alternate access to other ICs, and
- EC bypass.

6.4.2 Packet Switching Packet networks have their own layered structure for data communications within the seven layer Open Systems Interconnection (OSI) standard of the International Standards Organization (ISO) [25]. Generally, in this structure, communication at higher OSI layers assumes secure communication at lower OSI layers. Additional data communication protocols such as Frame Relay or other non-OSI standard implementations may use the OSI or other layering structures.

Techniques for survivability of packet networks at the service layer include some of the following actions within the transmission protocols and routing schemes:

- retransmission of packets or data frames that are lost or corrupted,
- dynamic routing of packets for routing around links and nodes that are failed or heavily congested,
- priority routing of messages, and

- congestion control; for fixed routing schemes — rerouting of traffic around failures or heavy congestion, queuing of packets or rejection of packets between certain point-to-points.

6.4.3 Common Channel Signaling CCSNs (Common Channel Signaling Networks) are dedicated signaling networks which transport both circuit associated (call setup) and non-circuit associated (transaction capabilities) signaling messages, between exchanges or other signaling nodes, separately from the voice/data path, using the SS7 (Signaling System Number 7) protocol¹⁴. CCSNs support both Integrated Services Digital Network (ISDN) and non-ISDN or POTS calls in a circuit switched network. The SS7 protocol consists of the following five parts: the Message Transfer Part (MTP), the Signaling Connection Control Part (SCCP), the ISDN User Part (ISDN-UP or ISUP), the Transaction Capabilities Application Part (TCAP), and the Operations, Maintenance and Administration Part (OMAP). The MTP and SCCP provide the basic transport messages between nodes in the CCSN, management of mated pairs of Signaling Transfer Points (STPs), duplicate Service Control Points (SCPs) and databases, and load balancing within and among link sets. The ISDN-UP provides call handling signaling functions, such as call setup and release, for non-ISDN as well as for ISDN voice and/or data calls. The TCAP supports non-circuit related activities, such as establishing forwarding arrangements for an 800 Service call. The OMAP provides operations and management of the logical and physical resources.

Survivability techniques are needed for each of the five parts of the SS7 protocol. Some survivability techniques for the MTP and the SCCP are described here. Software diversity, which can be used as a survivability technique for any part is also described. The TCAP relies mainly on SCCP for survivability. Some survivability techniques for circuit-switching networks using the ISDN-UP for call control are covered in Section 6.4.1.

6.4.3.1 Service Layer Architecture for MTP CCSNs as designed today include a high level of network element redundancy, intended to avoid interruption of service for the users of the traffic networks which rely on CCSNs for signaling. Common Channel Signaling (CCS) network element arrangements are used as service layer survivability techniques. The effectiveness of these arrangements depends on:

- proper engineering of CCS network elements to satisfy the desired performance objectives, and

¹⁴ The SS7 protocol, a variant of CCITT SS No. 7 [41], is modeled after the Open System Interconnection (OSI) seven-layer reference model [25], although it does not strictly adhere to it. Figure 5 of [41] illustrates this relationship.

Technical Report No. 24

- support of CCS network element redundancy and arrangements by the interoffice facilities (physical, system and logical layers) network, which should have an appropriate level of diversity of equipment sites and facilities (see sections 6.1, 6.2, 6.3, and Table 5.1).

There are two categories of signaling nodes, or Signaling Points (SPs):

- a. Signaling End Points (SEPs), consisting of:
 - CCS Switching Offices (CCSSOs), which have the capabilities to provide call control on the CCSN and may also interact with centralized, on-line databases to support query-oriented services such as 800 Service and Alternate Billing Service (ABS). CCSSOs that have both call control and database access capabilities are called Service Switching Points (SSPs).
 - SCPs, which provide access to databases and query processing systems for call control and customer-related information, and decision-making for intelligent services. SCPs may be duplicated, or may operate as stand-alone units.
- b. STPs, which are high-speed, high-reliability, large-capacity packet switches for CCS message routing and transport. The American National Standard (ANS)-based reference architecture (ANSI T1.111.5-1992) [43] has STPs deployed in mated pairs. Figure 6.9 shows a typical STP hierarchy that consists of:
 - Regional (remote) STPs, or RSTPs, which are directly connected to regional SCPs.
 - Local (home) STPs, or LSTPs, which route traffic between switching offices, and forward traffic to RSTPs for routing to an SCP.
 - Gateway STPs, either regional or local STP pairs, which interconnect different carriers' CCSNs and route signaling traffic between them.

Note that an STP can serve both RSTP and LSTP functions. For simplicity, not all possible connections are shown in Figure 6.9.

According to ANSI T1.111.5-1992 [43], there are six categories of CCS link sets, Access (A), Bridge (B), Cross (C), Diagonal (D), Extended (E) and Fully-associated (F):

1. A-link sets connect SEPs (SSPs, CCSSOs and SCPs) to home STP pairs.
2. B-link sets connect STP pairs of the same hierarchical level within a network (e.g., local STP pair to local STP pair).

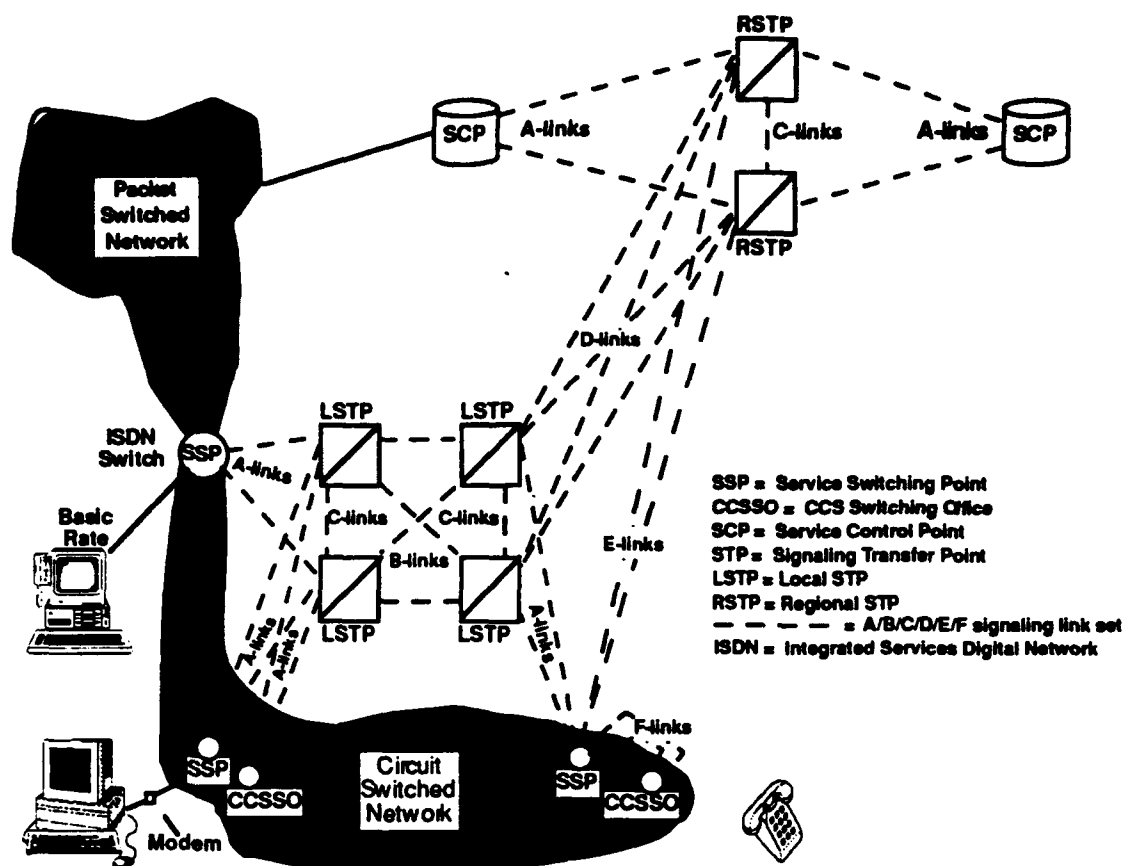


Figure 6.9: Relationship Between a CCSN and Other Traffic Networks

3. C-link sets connect the two mates in a mated STP pair. Under normal (no failure) conditions, they carry only network management messages.
4. D-link sets connect STP pairs of different hierarchical levels within a network (e.g., local STP pair to regional STP pair) or interconnect STP pairs of one network to STP pairs of a different network¹⁵ (e.g., EC STP pair to IC STP pair).
5. E-link sets directly connect SEPs with remote STP pairs.
6. F-link sets directly connect SEPs which have a high community of interest.

A-link sets are deployed in pairs (such a pair is called a combined link set) from an SEP to its home mated STP pair. E-link sets are deployed in pairs (such a pair is called a combined link set) from an SEP to one or more remote mated STP pairs. B- and D-link sets are deployed in quads (one combined link set from each STP) between two sets of mated STP pairs, as per the ANS-based reference architecture.

¹⁵ This terminology is based on agreements by T1S1.3 at its November 1992 meeting [50].

In such cases, the CCSN traffic load is balanced among the available (non-failed) links within each of the A-, B-, D-, E- and F-link sets. Upon STP or link set failure, the entire load is routed onto the A-, B-, D-, E- and F-link sets which remain connected, or the C-link sets. In order for service layer diversity to be supported at the physical layer, (i) mated STPs and duplicated SCPs should be deployed at physically diverse equipment sites and (ii) paths using physically diverse equipment sites and facility routes are needed between the link sets in a combined A-link set, between the link sets in a combined E-link set, and within B- and D-link set quads.

Some networks may not comply with the ANS-based reference architecture (e.g., STPs in stand-alone, unmated configurations). Performance implications for such configurations are discussed in Section 7.4.1.3.

The basic mesh network segments used in CCSNs are [22, 43]:

- *user interface segments*, consisting of the SEP equipment which contains shared CCS message handling capabilities.
- *network access segments*, consisting of the SEP interfaces¹⁶ with the A-link sets, the A-link set transport facilities, the STP interfaces with the A-link sets, and the portion of the STPs which contain shared CCS message handling capabilities.
- *backbone network segments*, consisting of the STP interfaces with the B-/D- and C-link sets, and the B-/D- and C-link set transport facilities.

Figure 6.10 illustrates a typical MTP arrangement of CCS basic mesh network segments. Given that sufficient link capacity has been provisioned so that all surviving links can carry the offered loads, such arrangements provide service layer survivability and limit service disruption due to: (i) any single failure in the network access segments or backbone network segments, and (ii) any double failure in the backbone network segments¹⁷.

Some Survivability Techniques in the MTP layer include:

- dynamic rerouting after link and STP failures,
- congestion control mechanisms,
- graceful MTP restart procedures, and
- cluster routing procedures.

¹⁶ Interfaces are those CCS message handling capabilities dedicated for one link.

¹⁷ This implies that the four link sets composing a quad are provisioned with three-way diversity.

The MTP layer dynamically changes message routing to account for the changing availability of links and STPs. Congestion control procedures reduce traffic loads to links and STPs by throttling message generation at the sources. MTP restart procedures allow an STP to become available at the MTP layer without becoming overloaded with service and network management messages and processing. Cluster routing reduces the amount of routing information required in an STP, the amount of network management processing, and network management message transmission required if routes change their status (e.g., available to unavailable). Cluster routing may also benefit SEPs.

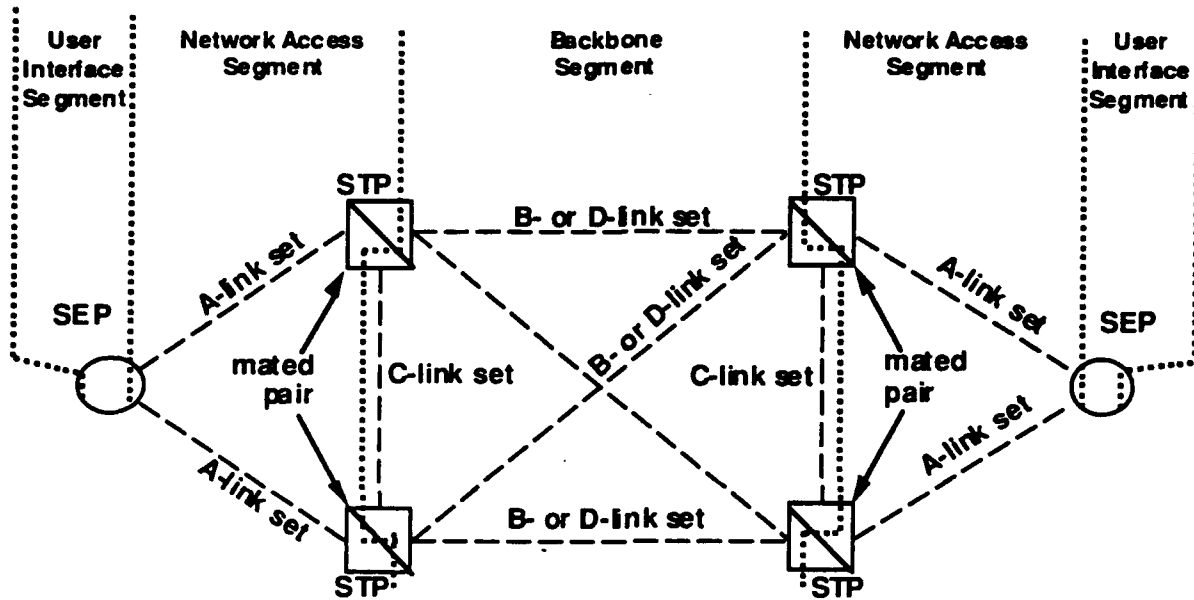


Figure 6.10: CCS Basic Mesh Network Segments

6.4.3.2 Service Layer Architecture for SCCP The SCCP provides additional network type functions to the MTP procedures. SCCP supports connectionless and connection-oriented services that are required for data transport between points. Connectionless service refers to data transport through the network without the set-up of a logical signaling connection, (datagram service). Connection-oriented service refers to data transport through the network using logical signaling connections (virtual circuits). The signaling links, STPs and SEPs may be combined in many different ways to form a "signaling network." Figure 6.11 gives an example of how the MTP signaling network of Figure 6.10 can be configured for SCCP.

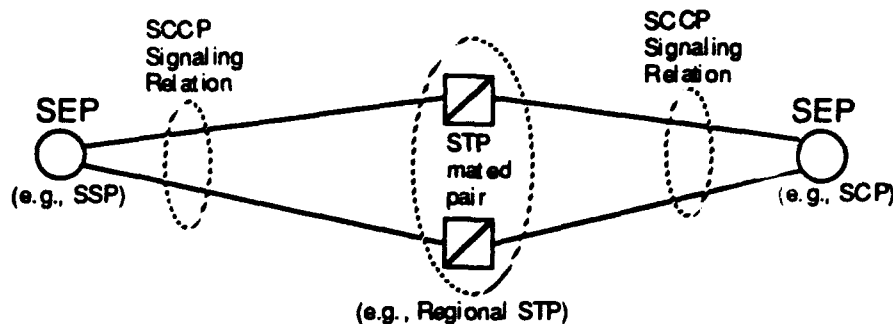


Figure 6.11: Example SCCP Signaling Network Structure

In the above figure an SCCP signaling relation exists between each SEP and the pair of STPs. The STP pair performs SCCP relay functions (e.g., Global Title Translation, or GTT).

Survivability Techniques in the SCCP layer include:

- backing up GTT to another STP pair,
- duplicating SCPs,
- load balancing between two SCPs,
- switching to backup SCPs, and
- using multiple MTP routes.

SCCP redundancy through multiple signaling relations between two signaling points can be achieved by equipping several STP pairs for GTTs. Load balancing between two SCPs spreads the impact of any traffic surges. When a data base fails in an SCP, SCCP can switch to use backup data bases in other SCPs. The impact of any changes to the MTP signaling network structure is diminished through the use of multiple MTP routes for some SCCP signaling relations.

6.4.3.3 Manual Traffic Management Controls In situations where the SS7 protocol may not provide protection against some failures or implementation error conditions, several manual traffic management controls are available and others have been proposed [20]. These manual network traffic management controls provide or will provide capabilities to render link sets unavailable to all message traffic, and to selectively block message traffic based on a number of selected MTP and SCCP attributes.

For example, one control would allow a network operator to cause links to become unavailable. This control could be used to isolate a signaling point from other signaling points or isolate one network from another.

6.4.3.4 Relationship of Software Diversity to CCSN Reliability With the increasing amount of software deployed in the CCSNs and its increasing complexity, software and its failure effects on the CCSN have become a major

Technical Report No. 24

concern. Because software engineering is a process susceptible to human error, no program can be guaranteed to perform as expected. The software “bug” contributing to the 1991 CCSN outages, which affected a large number of customer lines, has underscored the vulnerability of CCSNs to software failures.

Potential alternatives related to CCSN reliability are included in the following paragraphs. These potential alternatives are included here because they are being discussed in the industry; however, the validity of these and other potential alternatives is not known and requires further investigation. In addition, some of these potential alternatives may not be viable.

Four alternatives are identified here: i) multiple software developments in STPs, ii) different software generics for backup, iii) mixed-supplier STP pairs, and iv) E-links to different suppliers' STPs. This Technical Report considers mixed-supplier STP pairs to be the least viable of these alternatives.

The *advantages and disadvantages (or concerns)* of these alternatives to ensure software diversity *should be weighed* by individual telecommunications network providers. For example, the prevalence of correlated software failure modes, the cost of deploying these alternatives, as well as the service and revenue lost by correlated software failures should be carefully considered and analyzed. These and the other concerns expressed below are not exhaustive listings, but rather catalysts for further discussion and standards contributions.

i) Multiple software developments in STPs: Software generics meeting the same requirements could be developed by two different teams from a single supplier (i.e., not identical copies). These two developments would reside on the two STPs of a mated pair and would help ensure the software failure mode independence between the two STPs in the mated pair.

Following are some of the concerns related to this alternative:

- The additional resources required for multiple software development and the associated costs concern both suppliers and network providers. These additional costs would probably be reflected in the prices network providers pay for these products, and would also continue for each new generic.
- Although the processes for multiple software developments may be separate, many faults may result from the fact that the requirements/specification stage is usually the same.
- Uniformity in features, capabilities, user interfaces, etc. between the two software developments is extremely important. In other words, multiple developments of the software have to be totally transparent to the users.

ii) Different software generics for backup: The same software could be used in both STPs as the current implementation, and a different software generic (e.g., an older version) could be used for backup. For example, when certain thresholds are exceeded or phenomena observed, an older generic would be loaded into the STPs.

The following are some of the concerns with this alternative:

- Based on the current understanding of the behavior of the CCS network, the measurements to be used and the phenomena to indicate major problems can not be clearly defined. It is therefore extremely difficult to decide what conditions should exist before loading the STPs with a backup generic.
- The procedures to reboot and reload the STPs under the failing conditions and the problems which may occur in transition need to be carefully evaluated. Without detailed procedures to ensure smooth transitions, many problems may occur.
- Incompatibilities may exist between newer and older generics and less capabilities may be offered by the previous generics.

iii) Mixed-supplier STP pairs: It has been suggested that mixed-supplier STPs in a mated pair configuration may protect against failure conditions propagating between the mates, because the failure modes of the two mates may be independent for both hardware and software.

The following are some of the problems to be resolved with this alternative:

- The lower-capacity STP in the mated pair becomes the limiting factor. Furthermore, the CCSN supported by the mixed-supplier STP pair would suffer the shortcomings (capacity, recovery, etc.) of both pairs.
- If all interoperability issues are not resolved and thoroughly tested, and deployment is not well coordinated, introducing a different supplier's STP to an existing mated STP pair would risk end-office isolations or mate STP outages.
- OA&M and Provisioning:
 - the OA&M and provisioning functions of each mate STP would have differing requirements that would create problems within the respective support areas. These problems include: link growth and activation, making routing data changes and, in particular, systems support during critical network outages.
 - If all of the activities and differences between each STP in the mated pair are not well understood, then additional and exceptional (e.g., interoperability) training of operations personnel is required. Lack of such training would jeopardize the survivability of the CCSN.
 - Additional, separate data administration may be required, and any inconsistencies between mate STPs would cause confusion.

Technical Report No. 24

- Troubleshooting mixed-supplier STP mated pair failures would be extremely difficult with the two suppliers' technical support organizations.
- System initialization procedures may have designed capabilities that facilitate operation when the mate STP is from the same supplier.
- There may be feature differences between mixed-supplier STP mated pairs which would cause inconsistencies. Timing of feature introduction would also be different with the two suppliers.
- Comprehensive conformance testing needs to be undertaken (e.g., to ensure that the STPs do not send or depend on proprietary C-link messages during recovery).

iv) E-links to different suppliers' STPs: In addition to the two A-link sets to the home STP pair, E-links could be deployed to a remote STP pair of a different supplier [28] (see Figure 6.12). This architecture provides alternate access for SEPs to the rest of the CCSN and also helps ensure software and hardware failure mode independence. This architecture will improve the availability of the network access segment as well as bypass problems that occur in the home STP pair or the quad links. Various levels of diversity implemented in E-links (e.g., diversity between E- and A-links) may lead to different levels of reliability improvement. The benefits of E-links, which increase the diversity and redundancy of the CCSN, do not depend on the provisioning of different supplier STP pairs in the network.

The following are some of the concerns with this alternative:

- In the case where the E-links are deployed to a pair of STPs from a different supplier, all of the functions and features available at the home STP pair may not be available at the remote STP pair. This may cause capabilities which might otherwise be offered on one STP pair to be delayed until the supplier of the other STP pair develops the software generics to support those capabilities.
- If E-links are deployed, more routing options exist and therefore, routing tables will be more complex in both switches and STPs. Additional effort may be needed to administer and maintain the routing table databases. Furthermore, much more memory will be needed at the switches and STPs to support the additional routing and translation table information.
- The remote STP pair would have to handle the entire traffic load if the home STP pair becomes inaccessible. As a result, additional throughput and link ports are required at the remote STP pairs to handle the traffic and links from the SEPs when E-links are deployed. This may force rehomings and reengineering the network more frequently.

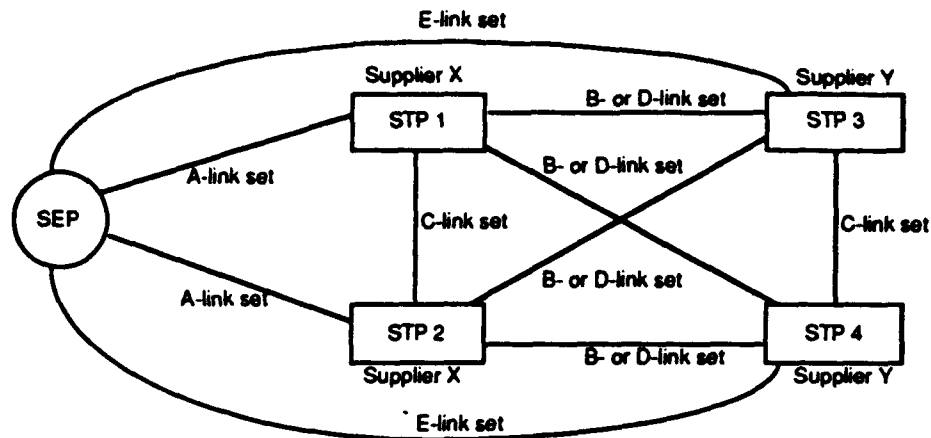


Figure 6.12: E-Link sets to a Remote STP Pair of a Different Supplier

The above four alternatives, in addition to others, should be considered in future discussion and standards contributions addressing software diversity and CCSN reliability improvement. These alternatives are intended to ensure software failure mode independence of the STPs so that correlated software failures will not result in a major network outage.

6.5 Integrated Techniques Some integrated techniques include networks that have combinations of DCSs, automatic protection switching and self-healing rings. A few of these are summarized below:

- A subnetwork combining DCS restoration and point-to-point systems using APS with diverse routing, where the spare bandwidth for the DCS rerouting method can access the spare paths of the APS systems that are reserved for their protection channels (e.g., Figure 6.13). Among other methods, this can be accomplished by having direct optical interfaces into the DCSs for the APS systems.
- Combinations of many intersecting (including overlapping) self-healing rings covering the network:
 - traffic between rings is cross-connected via the low-speed ports of DCSs and Add-Drop Multiplexers (ADMs, see Figure 6.14), (A device which processes signals for combination and decombination purposes) and
 - the self-healing rings intersect via DCSs with integrated high-speed or fiber interfaces (see Figure 6.15). The DCS plays the role of ADM for all the intersecting rings at that office, as well as cross-connecting traffic between rings.
- Self-healing rings lie on the perimeter of the network and intersect with the core, mesh subnetwork of DCSs (see Figure 6.16). Failures within the mesh are protected by DCS restoration methods. The rings may either intersect the DCSs with integrated high-speed or fiber interfaces or the traffic among

Technical Report No. 24

the rings and DCS mesh are cross-connected via low-speed ports of the DCSs and ADMs.

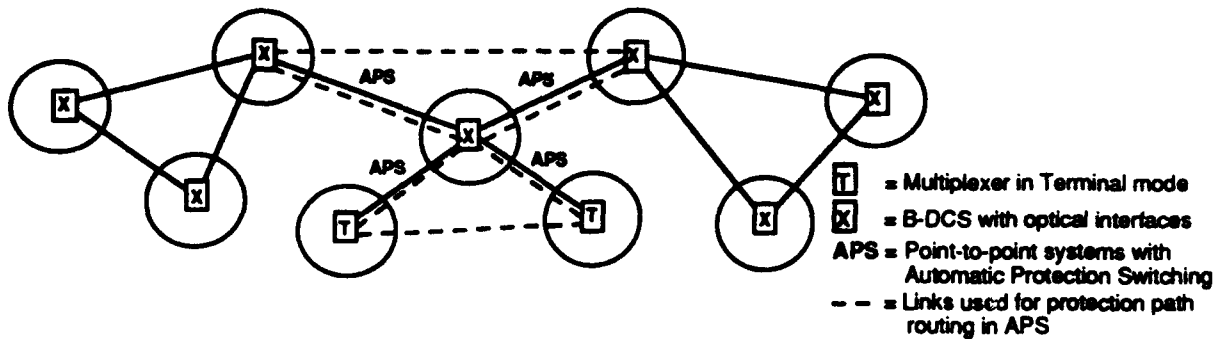


Figure 6.13: Integrated Point-to-Point Systems and DCSs

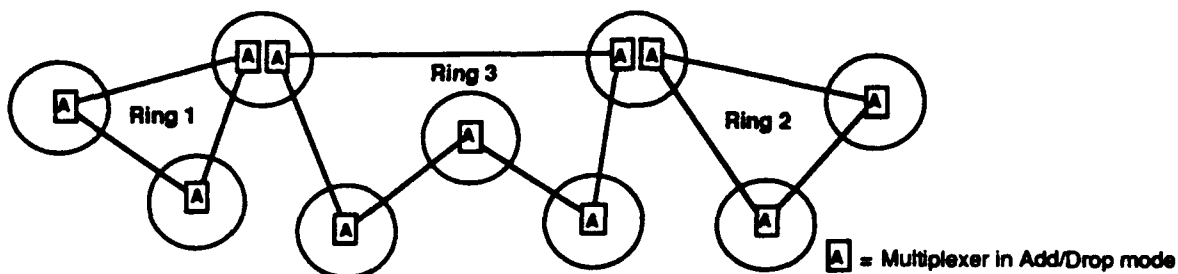


Figure 6.14: Intersecting Self-Healing Rings

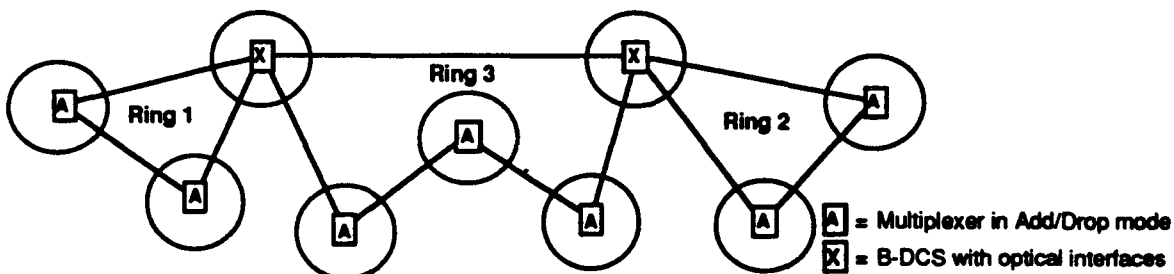


Figure 6.15: Integrated Self-Healing Rings and DCSs

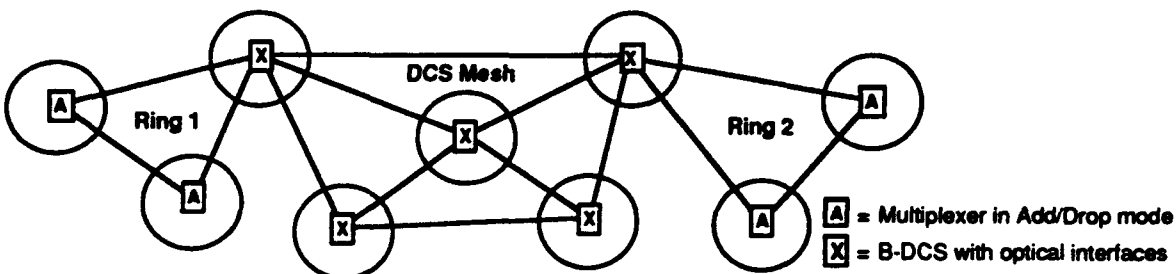


Figure 6.16: DCS Mesh Ring

7. Network Survivability Performance Analysis

7.1 Network Survivability Characterization For the purpose of a network survivability performance study, the general framework defined in Section 5 can be used and the network can be partitioned into the following four network survivability layers: service, logical, system and physical. For each layer, a set of measurements can be defined and their characteristics can be modeled.

There are many ways to characterize network survivability and define survivability measurements. Using survivability measurements, one can formulate models and then define relevant measures such as traffic survivability ratio, service connectivity ratio, average network downtime, etc., that can be used to estimate the survivability performance of a real or hypothetical network. Table 7.1 lists some network survivability measurements for the four network survivability layers.

Layers	Examples of Survivability Measurements
Service	end-to-end grade of service, number of calls, number of connected subscribers, traffic volume, carried load, packet throughput, utilization
Logical	number of surviving lower-rate transmission channels (DS-1s, or DS-3s), connectivity
System	number of surviving transmission systems
Physical	building integrity, connectivity and number of surviving cables

Table 7.1: Examples of Network Survivability Measurements

7.2 Network Survivability Analysis Models There are two basic approaches to survivability analysis defined here. The first, the Given Occurrence of Failure (GOF) survivability analysis model, uses a conditional approach and defines survivability measures for a network assuming that given failure(s) have occurred. This GOF model may either use probabilistic weighting of the resulting states of the network and resulting network restoration and/or repair after the failure or it may use deterministic analysis of these states. Both approaches can be used to evaluate different restoration, repair or preventative methods depending on which type of comparison characteristics are critical.

The second survivability analysis model, the Random Occurrence of Failure (ROF) model, uses probability of network failure(s) and, possibly, rates of repair and/or restoration to calculate various probabilistic measures of network unservability or loss (e.g., the expected amount of time a network is unservable).

7.2.1 Given Occurrence of Failure Survivability Model In the GOF survivability model, it is assumed disaster has already occurred and a given

element (or elements) of the network has already failed. Hence, from a modeling point of view, the occurrence of a given failure is assumed. Users often phrase their survivability requirements in terms of which types of failure they want their traffic protected from, and what proportion of the traffic should survive. The general procedure for evaluating GOF measures is as follows (see also [26]):

1. define a survivability measurement (see Table 7.1),
2. identify the sample space (i.e., the failures that can occur),
3. choose the failures of interest, and
4. calculate the network survivability measures.

7.2.2 Random Occurrence of Failure Survivability Model The ROF survivability model is a general form of the well known availability model. This model is based on the assumption that failures can be characterized by random variables with given probability distribution functions. The general procedure for evaluating availability based network survivability measures is as follows:

1. obtain observed rates of failure and repair/restoration,
2. define network survivability measurements of interest,
3. identify the network state space (i.e., the various states in which a network can reside concerning whether its components are working or failed),
4. determine the survivability measurement for each network state,
5. determine or assign the transitional probability from each state to another, and
6. calculate the network survivability measures (e.g., the expected units lost or unservable over time).

Note: Step 6 is often a very complex task and sometimes impossible to evaluate exactly in networks.

7.2.3 Application of GOF and ROF Models The above two survivability models can be applied to develop performance measures for all four survivability layers. For example, the fraction of buildings remaining after a hurricane of a specified strength and extent is a survivability attribute for the physical layer. The measures used in system and logical layers are usually based on a static traffic model, i.e., the traffic is assumed to change over a long interval of time, an assumption that may be inappropriate for higher network layers (e.g., a service layer network).

In both the ROF and GOF models, it is important to first specify what measurements of the network to capture, then to obtain the survivability measures

using procedures summarized in Section 7.2. To clarify the application of the models, more discussion on their measures follows.

7.3 GOF Network Survivability Measures In the GOF model, a selected measurement, say x , of the network can be quantified, such as in Figure 7.1, failure of x occurs at time t_0 and the following survivability attributes can be defined:

- S_a the fraction of x that remains after failure has happened and before restoration starts ($1-U$),
- S_u the fraction of x unservable after failure has happened and before restoration starts (U),
- S_r the fraction of x that is restored at t_1 ,
- t_0 the time the failure occurs,
- t_r the duration required until fraction S_a+S_r of x is restored, and
- t_R the duration required until all of x (t_2) is restored (D).

In general, S_a , S_u , S_r , t_r and t_R are random variables depending on network topology, type of failure, restoration techniques, etc.

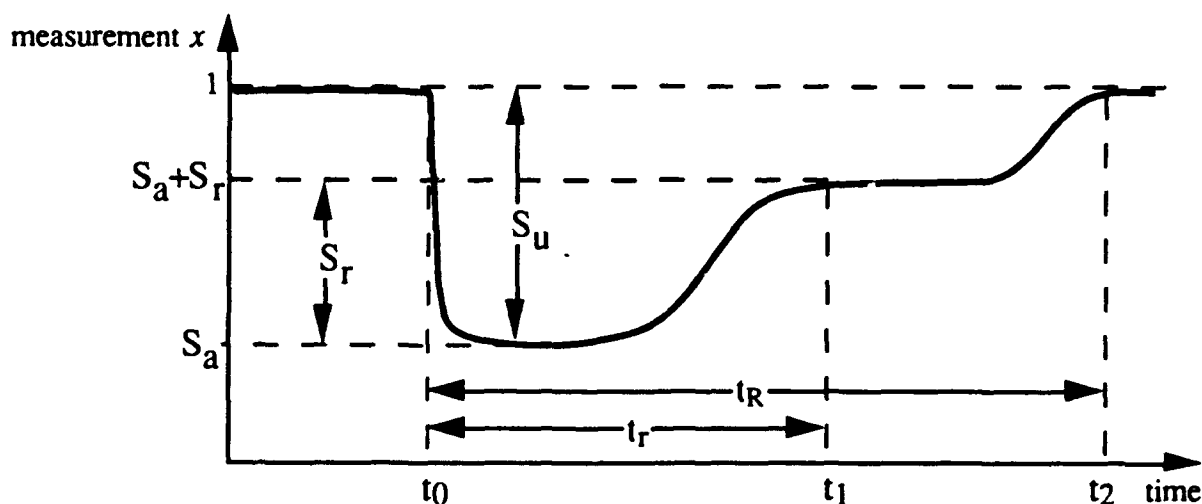


Figure 7.1: Survivability Attributes

Using the three main survivability attributes, S_a , S_r , and t_r defined above, the survivability of a network for a given "measurement" can be characterized. In general, these attributes are random variables, and each one has a probability (frequency or distribution) function.

A number of different quantities of interest can be derived, each capturing a particular network characteristic. For instance, for S_a , S_r , and t_r , we can obtain the

expected values, the worst-case values, the p -percentile values, and the probability of zero. Note that each measure captures a different aspect of network survivability. Two examples on network survivability analysis and assessment based on the GOF model are given in Appendix B.

7.3.1 Service Layer Examples

Service Outage Performance Based on Traffic Blocking

Traffic blocking level can be used as a survivability measurement in circuit switched network survivability analysis. In this case, the survivability parameter can be the "relative average traffic blocking level." Many grades of availability could be defined, in terms of various blocking levels and the duration of such blocking. The basic two states are defined as Acceptable (no service outage) and Unacceptable (service outage) as shown in Figure 7.2.

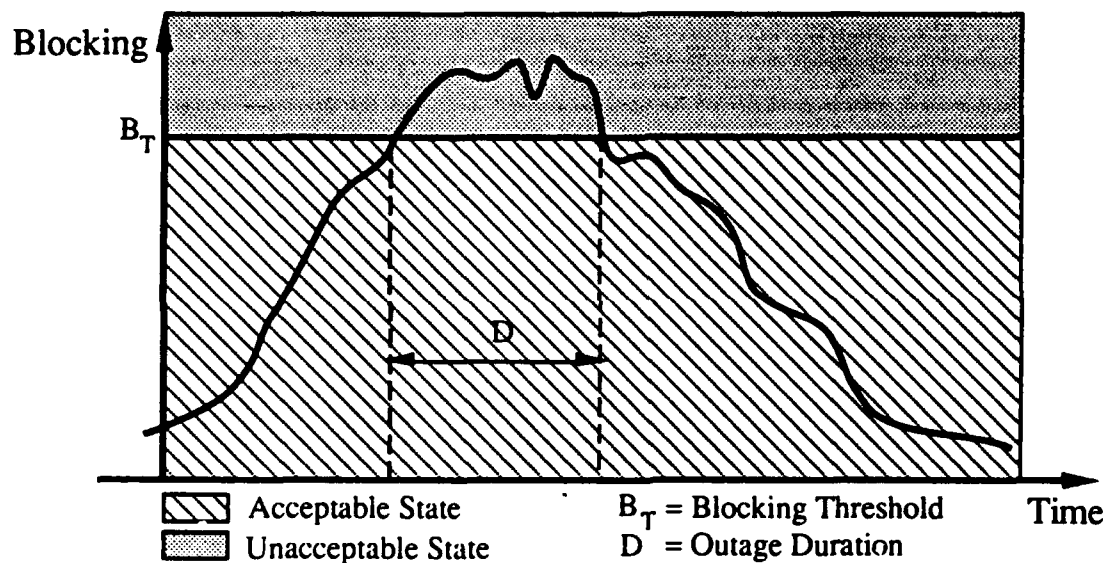


Figure 7.2: Two-State Traffic Model

The acceptable state can be defined as the case where the blocking is less than some specified level, B_T . The threshold of B_T could be the engineered blocking, or some higher level. The cause of blocking may be network failure conditions (capacity loss) as well as overload situations (demand increase).

The unacceptable state can be defined as the case where blocking exceeds the blocking threshold B_T . The Service Outage Time, D is defined to be the interval between the point where blocking exceeds the threshold B_T , and the first point where blocking drops below B_T . The blocking threshold which defines the transition between the two states can be determined on the basis of specific user group requirements and objectives. Although a two-state model is defined here, other states can be defined, including marginal states of acceptability. For

example, a network manager may define other states which will allow more refined actions.

We will relate the blocking survivability parameter to the network failure impact categories in Section 5.2. Given a failure, three parameters, U, D and E are used to classify the magnitude of the service failure. Here, U is the load-weighted percentage of point-to-points whose blocking exceeds the threshold B_T , D is the maximum duration for which all point-to-points included in that percentage exceed the threshold, and E is the set of point-to-points considered. For example, if the combination of U, D and E is within the proper qualifying region as defined in Section 4, then a catastrophic service outage has occurred for this user group.

User Lost Erlang

The User Lost Erlang (ULE) measure proposed in [33], is a single dimensional measure of service outages. Single dimensional measures are not adequate to capture the three essential parameters of an outage (unservability, duration and extent). For example, under the ULE measure a 30 million access line Interexchange network with a loss of one million Erlangs over six minutes is equated with a one million access line EC (Exchange Carrier) network with a loss of ten thousand Erlangs over ten hours.

7.3.2 Logical and System Layer Examples

Restoration Time

Given a network failure, the restoration time is the time from loss of first working signal until the time the last signal is restored by the restoration method. Note that depending on various types of failures (e.g., node failure) and network configurations, it is possible that not all lost signals can be restored. Also, restoration times of different methods must be compared at the same level of restoration, i.e., over similar networks with the same number of lost and restored signals.

Consistent with Section 6, restoration time can be generally broken into the following phases:

1. *Detection time*: the time from loss of first signal until the restoration method detects the loss or alarm.
2. *Notification time*: the time required to notify the control architecture of the failure.
3. *Identification* of failed signals.
4. *Path selection time*: the time required to obtain alternate paths for lost signals (by data look-up or algorithm).

5. *Rerouting time*: the time required to complete all cross-connects or switching to alternate paths.

Impacts of restoration time on various network services can be found in Appendix D.

Verification Time

In addition to restoration time, practical implementation of most restoration methods will verify their operation. Verification time is defined to be the time required to verify successful cross-connection or switching by an operations system, network controller, or network element. In some methods, the verification process starts after the last signal is restored, while in others it occurs in parallel with the restoration process.

Verified Restoration Time

While the verification time is not strictly included in the restoration time definition, its implementation can vary widely across restoration methods and its measure can be important for the performance evaluation of different restoration methods. Therefore, a combined measure is defined here. Given a network failure, verified restoration time is defined to be the time from the loss of the first working signal until the time the last signal is restored and verified.

Restoration Ratio

Given a set of simultaneous network failures, the restoration ratio is x/y , where x is the number of signals restored by the restoration method and y is the number of signals that were lost because of the network failure. This measure is used to compare two similar restoration methods on equivalent networks, i.e., with the same nodes and links, the same amount of working signals, the same amount of spare capacity, and the same amount of lost signals, x . An aggregate restoration ratio can also be calculated by averaging the restoration ratio over a given set of non-simultaneous failures (e.g., all possible single link failures or all single node failures or both).

Given the failure set, a restoration method can also be compared against the optimal ratio, i.e., the maximum restoration ratio for a given network. Note that in some cases the maximum ratio may be intractable to compute. For example, with DCS reconfiguration methods, if a *link* rerouting method is used (see Section 6.3.1 for definitions), the theoretical maximum number of signals that can be restored for any single link failure by using any link rerouting method is the solution to a maximum "flow" problem (in optimization terminology) between the two end points of the failed link. This is, generally, an easy problem to solve. However, if *point-to-point* rerouting methods are used, then this theoretical maximum number is at least as large as that of link rerouting methods, but computing the theoretical

maximum number of signals that can be restored by using point-to-point rerouting methods is, in general, an intractable problem.

Efficiency Ratio

Given a possible set of non-simultaneous failures in a network (e.g., all possible single link failures), the efficiency ratio is defined to be how much spare signal capacity is necessary to guarantee maximum restoration over all possible network failures, expressed as a ratio = *number of spare signals / number of working signals*. Note that, in general, for node failures, signals that terminate/originate at the failed node cannot be restored, and hence maximum restoration is less than 100%.

Given the possible failure set in a network, a restoration method can also be compared against the optimal efficiency ratio, i.e., the minimum theoretical efficiency ratio in that network for any possible restoration method. As with the restoration ratio, the minimum ratio may be intractable to compute.

7.4 ROF Network Survivability Measures In contrast to the GOF model, the ROF model makes no assumptions that a given failure has occurred (i.e., t_0 in Figure 7.1 is random). If the probability distribution of time between failures is exponential, the failure process could be fully characterized by Mean Time Between Failure (MTBF), or Mean time Between Service Outages (MTBSO) and the Unsolvability (ν), could be estimated in terms of Mean Time To Repair (MTTR) and MTBSO. Service unsolvability, based on traffic blocking is an important measurement in the service layer and is discussed next.

7.4.1 Service Layer Examples

7.4.1.1 Circuit Switching

Overall Average Blocking

The ROF model can be used to estimate the expected Overall Average Blocking (OAB_{avg}) when enough historical data on the long-term variation of the blocking level is available and an acceptable estimate for MTTSR (Mean Time to Service Restoration) and MTBSO (Mean Time Between Service Outages) can be obtained. It can be shown that [34]:

$$OAB_{avg} = (1-\nu)B_n + \nu B_u$$

where B_n and B_u are the blocking in acceptable (normal) and unacceptable states respectively. The Unsolvability (ν), is defined as the long-term ratio of outage time to scheduled service time. Under the assumption of exponentially distributed failure and restoration times we have:

$$\nu = \frac{MTTSR}{MTTSR + MTBSO}$$

Annual Loss of Traffic

Annual Loss of traffic is the amount of traffic that the network is expected to lose per year due to failures. This aspect of survivability takes into account the frequency and duration of failures using the ROF model.

A measure for the annual loss of traffic is Expected Loss of Traffic (ELT), measured in units such as carried load (e.g., in units of Erlangs) per year.

7.4.1.2 Packet Switching

Packet Delay (Transport Time)

Survivability of packet switching networks can be measured using packet delay or transport time. Others measurements are: message loss probability, undetected error probability, message out of sequence probability, and availability¹⁸. Packet delay can be computed on a point-to-point basis and averaged over time, i.e., point-to-point packets from A-Z see an average delay over a given interval of time. Aggregate measures are computed by averaging the weighted point-to-point delays over all point-to-points, where the weighting is the proportional relative packet traffic (load) to the whole traffic in the network of interest.

Failure can be related in terms of Section 4, by analyzing a distribution of point-to-points whose delay exceeds a given threshold. This threshold would depend on the grade of service objectives for the particular network. An example of an aggregate unservability measure could be the weighted proportion of point-to-points whose delay exceeds the threshold, e.g., 50% of traffic exceeds the threshold (the point-to-points whose delay exceeds the threshold constitute 50% of the total traffic of the network).

7.4.1.3 Common Channel Signaling

Performance Evaluation

Most performance characteristics commonly used for packet networks also apply to CCSNs. Performance of CCSNs can be evaluated using measurements such as availability (or downtime), transport time (delay), message loss probability, undetected error probability, message-out-of-sequence probability, and signaling link bit error ratio [21]. The first two measurements have direct impact on planning and engineering survivable CCSNs, and are discussed below.

Performance measurements such as delay or utilization can be specified as peak (maximum) over a given time interval, or mean (average) over a given time interval, or else are relative to the completion of a given type of call. Whether

¹⁸ CCITT Recommendation I.355 [47] on ISDN availability identifies the wide range of ISDN network performance parameters (circuit-switched and packet-switched) and their thresholds for defining unacceptable states.

peak or mean delay/utilization measures are appropriate, and how long the time intervals over which these statistics should apply need to be determined to derive appropriate CCSN traffic engineering parameters. For more detailed information on SS7 monitoring and measurements, see [48].

For a circuit switched network, for example, traffic engineering is based on performance objectives such as maximum busy-season busy-hour blocking. More study of the characteristics of CCS traffic which results from voice, ISDN or other services (this traffic is likely to be more bursty than traffic in a circuit switched network) is needed, to determine the relevant time scales for the CCSN.

Downtime Objectives

Downtime objectives are intended to control the amount of time a CCSN (or a segment thereof) is unable to perform its required signaling functions. They can be represented by a single number equal to the long-term percentage of time a CCSN — or segments thereof — are expected to be “down.” As such, downtime objectives can significantly influence end user perception of service quality. The expected percentage of downtime for a network element can be interpreted either as:

- the average downtime over many years for this network element, or as
- the average downtime over one year for a population of the network elements.

According to American National Standard (ANS) ANSI T1.111.6-1992, Section 5.1.2 [43], the MTP downtime objective for the CCS basic mesh network shown in Figure 7.3, corresponds to (an average of)¹⁹ no more than 10 minutes downtime per year for the signaling paths between two SEPs, and is broken down as follows (see Section 6.4.3.1):

- each user interface segment should be down (an average of)¹⁹ no more than 3 minutes per year
- each network access segment should be down (an average of)¹⁹ no more than 2 minutes per year, and
- the backbone network segment should be down a negligible amount of time (i.e., close to 0 minutes downtime per year). Note that downtime for this segment includes failures that prevent use of the backbone segment but do not by themselves disable any other segment(s).

¹⁹ The original text in [43] refers to “nominal requirements” that are interpreted in this Technical Report as “average downtime numbers.”

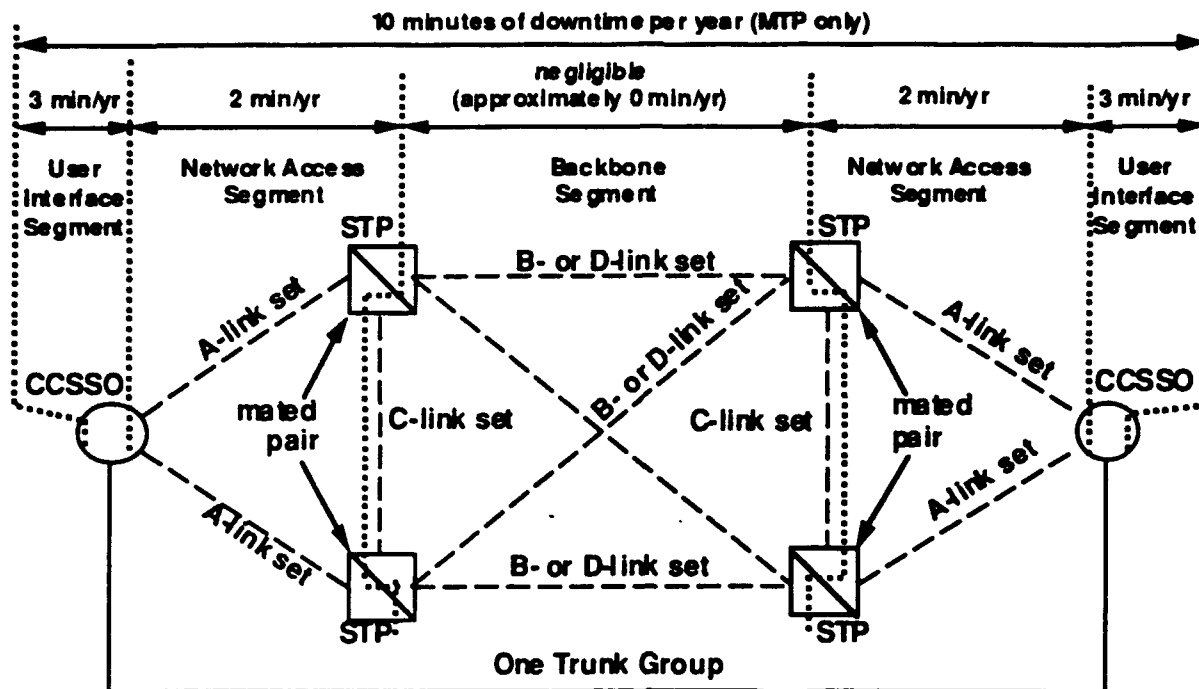


Figure 7.3: ANSI (T1.111.6) Downtime Objectives for CCS Basic Mesh Network Segments (MTP only)

To meet the above allocation, the ANS-based (ANSI T1.111.5-1992, Section 7.2.1) reference architecture uses two-way diversity for the A-link sets, mated pairs of STPs, and three-way diversity for the B-/D-link sets [43]. If three-way diversity is not achieved in the backbone segment, the downtime of that segment may not be negligible. Also, if mated STP pairs are not deployed in the CCS architecture, then the downtime objectives of the network-access segment and the backbone segment may not be met. Hence, the 10-minute end-to-end downtime objective may not be achievable.

When the SEPs in Figure 6.10 (Section 6.4.3.1) are both CCSSOs, the 10-minute end-to-end downtime objective and the above allocation to network segments correspond to a single trunk group with its terminating CCSSOs interconnected using the ANS-based reference architecture (see Figure 7.3). It also applies when instead of CCSSOs we have SSPs or SCPs.

One or more switches homed on an STP pair cannot communicate with all switches homed on another STP pair (but all switches on any one STP pair can still communicate with other switches homed on that pair) if any of the following occurs:

- the entire B-/D-link set quad fails,
- one of the STPs in either mated pair and the B-/D-link set pair of the other STP in the mated pair fail, or

Technical Report No. 24

- one or more A-link set(s) to one of the STPs in either mated pair and its C-link set and the B-/D-link set pair of the other STP in the mated pair fail.

Delay Objectives

Delay objectives have the most direct impact on CCS traffic engineering. Three major types of message transfer delay measures (as well as hybrids of these types) are usually considered: peak (or maximum), percentile, and mean (or average). A peak (or maximum) delay requires that messages be transmitted within a given time period. A percentile delay requires that a certain percentage of all messages be transmitted within a given time period. A mean (or average) delay requires that, on the average, messages be transmitted within a given time period. Some measures for cross-STP delays are provided in [23]. No objectives are defined in the SS7 standards for either link set delays or end-to-end delays²⁰.

Even when end-to-end delay objectives are completely defined, allocating them to individual cross-network element delays is difficult. When a hybrid delay objective involving the peak, percentile and mean criteria is considered as in the FCC Docket No. 86-10 case²¹, finding an effective allocation policy is a yet more complex difficult task. For example, it is difficult to determine how the 5-second call setup delay for the 800 Service mandated by FCC Docket No. 86-10 should be apportioned to the call piece parts: POTS connection to switch, intra-SSP processing, processing in network access segments of the CCSN, interconnection to ICs, access to SCPs, etc. Criteria/rules are needed to effectively perform allocation of delay objectives among individual or groups of CCS network elements. For example, one may want to increase the utilization of (thus allocating higher admissible delays to) more expensive CCS network elements.

Utilization Objectives

Signaling link set utilization is defined as the fraction of time message signal unit (MSU) packet transmission occurs over the link set. It can be computed as the ratio of the link set's carried load to the link set's capacity (or link set speed). Processor utilization can be computed in a similar fashion. Low utilization implies higher costs. High utilizations, on the other hand, may cause unacceptable delays. Even if the network element is engineered for a low utilization, CCS network element failures might cause some additional signaling traffic to be diverted to the network element, thereby increasing (usually, doubling for a single network

²⁰ Some delay objectives have been mandated by regulatory agencies. For example, average end-to-end call setup delays under 5 seconds for the 800 Service are required to comply with FCC Docket No. 86-10 [24].

²¹ FCC Docket No. 86-10 requires that (i) 97% of the 800 Service traffic involve call set-up times of 5 seconds or less by March 4, 1993, and (ii) 100% of the 800 Service traffic involve maximum call set-up times of 5 seconds or less and average call set-up times of 2.5 seconds or less by March 4, 1995.

element failure) its carried load and its utilization. Utilization upper bounds can be used to ensure both congestion control and survivability at reasonable costs.

As an example, [23] requires that a signaling link set load be limited to 0.4 Erlangs per link under normal conditions (i.e., utilization objective of at most 40%), so that if a failure occurs, an expected peak load of at most 0.8 Erlangs of CCS traffic would be carried by the surviving signaling link set. The time scale involved in this utilization objective is however left undefined in [23]. It is not clear whether hourly averages adequately capture the variations in carried loads, or whether averages over some other period (e.g., five minutes) are needed.

Based on [23], the utilization objective guideline of 40% applies to A-, B-, D- and E-link sets, which are then engineered to not exceed 40% utilization under normal (no failure) conditions. To support double failures in the backbone network segments, B- and D-link sets could be engineered to a utilization lower than 40% under normal (no failure) conditions. Currently, there are no standardized utilization guidelines for either C-link sets or F-link sets. C-link sets carry CCS traffic only in case of a combined B-/D-link set or A-/E-link set failure (they also carry network management messages between STP mates).

Setting adequate utilization objectives ensures survivability in the event of a single failure in the network access segments, or a single failure in the backbone network segments, or a double failure in the backbone network segments. Based on [43], in the event of such a failure, congestion is avoided if the utilization remains below 80%. However, a backbone network segment in a CCSN engineered according to this requirement cannot withstand a double failure such as the simultaneous failure of an STP and one of the two B- or D-link sets connecting its mate to the other mated pair of STPs. Among other alternatives [50] to lessen vulnerability to this type of double failure, the utilization objective for B-/D-link sets could be reduced.

Another requirement given in [23] is that an STP should be able to handle its mate's traffic load in addition to its normal (no failure) traffic load. No similar requirement for CCS nodal network elements other than STPs is given in [23]. SCPs are candidates for similar mating requirements.

7.4.2 Logical and System Layer Examples

Annual Loss of Traffic

Annual Loss of traffic is the amount of traffic that the network is expected to lose per year due to failures. This aspect of survivability takes into account the frequency and duration of failures using the ROF model. A good measure for the annual loss of traffic is Expected Loss of Traffic (ELT) – measured, in the logical layer, in units such as DS-1 or DS-3 minutes per year.

Technical Report No. 24

To compute this measure, for a network with say DS-3 demands, consider all the pairs of nodes in the network for which zero DS-3 demand or "traffic" exists (a DS-3 demand between node A and node Z is defined to be a contiguous transmission signal that is accessible at the DS-3 rate at both nodes A and Z and does not drop below the DS-3 rate between A and Z). For example, in local exchange networks a DS-3 signal may be provided to carry a customer demand at the DS-3 rate or it may be provided to multiplex or groom many DS-1 signals from various origination and destination pairs for transport between the node pair of the DS-3.

If the network of interest has self-healing capability, each DS-3 demand would normally be routed over a path of self-healing systems/subnetworks. Given a pair of nodes, $\{i, j\}$, and the number of DS-3 demands between them, d_{ij} , let $\{R_{ijk}\}$ denote the unique paths between i and j over which the $\{i, j\}$ DS-3 demands route. Correspondingly, let d_{ijk} denote the number of $\{i, j\}$ DS-3 demands that route over path R_{ijk} ($\sum_k d_{ijk} = d_{ij}$). ELT_{ij} is defined to be the sum over k of the product of d_{ijk}

and the expected downtime (in minutes per year) of path R_{ijk} . ELT , the quantity for the whole network, is the sum of ELT_{ij} over all pairs of nodes in the network for which non-zero traffic exists. The expected downtime for each path, R_{ijk} , is calculated by taking into account the probability and average duration of failures.

Annual Loss of Connectivity

Annual Loss of Connectivity is the propensity of the network to lose all connectivity between a pair of nodes per year. Connection between two nodes, i and j , is lost if, for all paths R_{ijk} , there exists no working or protection channels able to carry the demand, d_{ij} . This aspect of survivability also takes into account frequency and duration of failures using the ROF model.

The consequences of losing connection between two nodes can be more serious than the consequences of losing an equivalent amount of traffic throughout the network without losing connectivity. Loss of connectivity can lead to the loss of important emergency and high priority traffic or create a situation of isolation.

Two measures can be defined for this aspect of survivability, the Average Expected Downtime of Connection (AEDC), measured in units of minutes per year, and Probability Distribution of Downtime of Connection (PDDC).

The first measure, AEDC provides a summary of the loss of connectivity. The second measure, PDDC, allows a more detailed study of the likelihood of incidents of prolonged loss of connection.

Given a pair of nodes, $\{i, j\}$, $EDC_{i,j}$ is defined to be the expected value of the downtime of connection between nodes i and j measured in minutes per year. For the same pair of nodes, $PDDC_{i,j}$ takes the form of a curve. On the horizontal axis is duration T in minutes. On the vertical axis is the probability that in a year there will be at least one incident where the nodes i and j will lose connection with each other for T minutes or longer.

AEDC, the summary quantity for the whole network, is the arithmetic average of the $EDC_{i,j}$ over all pairs of nodes, $\{i,j\}$, with non-zero traffic in the network.

7.5 Qualitative Assessment of Network Survivability Techniques Other considerations beyond the GOF and ROF techniques are often important to assess network survivability techniques. This section defines qualitative assessment criteria for network survivability techniques.

Failure Types

This is defined to be the types of failures for which the network survivability technique can restore some or all signals. The general categories of failure types include:

- single link failure (i.e., single, partial, or total cable failure),
- two simultaneous (or near simultaneous) link failures,
- three or more simultaneous (or near simultaneous) link failures,
- single node failure (i.e., partial or total failure of a network element), and
- multiple node failure.

These categories can be used to assess the use of network survivability techniques.

Restoration Signal Rate

This is defined to be the rate or level at which the method can restore signals. For example, self-healing rings can restore signals at line rates (OC-48, etc.) or "path" rates, i.e., per channel, and DCS methods can restore at various signal rates (e.g., DS-1, DS-3, STS-n).

Stability

Stability is defined to be the response or ability to predict the response of the restoration method to variations or "perturbations" in the network parameters, such as type of failure, speed of data links, speed of hardware, database inconsistencies, data communication errors, or size of network.

Capacity Limitation

Capacity limitation is defined to be the inherent capacity limitations of a single "contiguous" configuration of the restoration method. For example, a SONET OC-48 self-healing ring can accommodate 48 STS-1/DS-3 signals that add/drop out of the ring before another ring is needed (each ring is in some sense "autonomous,"

although some types of dual homing rings can interact); a DCS reconfiguration method has no hard limit because more DCSs can always be added to the contiguous self-healing network.

Ability to Inter-work

Each restoration method requires particular equipment and spare signal capacity to function. This ability to inter-work is defined as the ability of a given restoration method to inter-work with other restoration methods by sharing the same signal capacity or equipment.

Control Mechanisms

The general types of control mechanisms for restoration methods are *distributed* (done by the network elements or equipment) and *centralized* (done by an operations system or subnetwork controller). Each restoration method can go through different phases of execution, as broadly outlined in the beginning of Section 6. For a given restoration method, the control mechanism can differ across different phases. For example, a DCS reconfiguration method may use distributed control to find alternate paths, yet use centralized control to accomplish the "normalization" step.

Implementation Requirements

Implementation requirements are the systems, procedures, and equipment, needed to implement and maintain the restoration method in the network. These are categorized below:

- interfaces with centralized network operations support systems,
- interactions with network planning personnel, procedures, and systems,
- network design algorithms and methods (for example, a DCS reconfiguration method must determine the placement of spare signal capacity in the network to guarantee a prescribed restoration ratio over a defined possible set of failures), and
- software or hardware implementing the restoration method (including cost, size, speed, and complexity).

8. Suggestions to General Industry

1. Use the (U, D, E) triple to quantify and categorize service outages. This approach is more comprehensive than a single dimensional measure, such as the ULE.
2. The performance measures, definitions and terminology used in this document should become the standard when discussing network survivability, reliability, integrity and performance. For example, they can be used in joint IC - EC network survivability studies.

3. Use the framework for classifying network survivability techniques to classify future network survivability analyses and performance techniques.

9. Recommendations to Standards Organizations

9.1 Recommendations for Committee T1

1. Review established performance objectives (e.g., restoral time objectives for SONET rings, utilization objectives for CCS network elements) to ensure they are appropriate for future applications (e.g., Broadband ISDN, or B-ISDN).
2. Use the performance measures and terminology introduced in this report in related standards work.

9.2 Recommendations for Future T1A1.2 Work

1. Quantify the qualifying regions for the (U, D, E) triple for service outage categories. Numerical limits, such as the FCC's existing outage reporting thresholds should be studied further to establish their validity for comparing network service outages.
2. Obtain additional user expectations on network survivability performance to meet current and future needs. Analyze impacts of service outages on user categories.
3. Develop planning, engineering and implementation guidelines for network survivability.

10. Summary

This Technical Report has addressed concerns of the Telecommunications Community about the issue of network survivability. This report provides a common approach to describe and assess network survivability. Terminology has been introduced. A framework for quantifying and categorizing service outages and a framework for classifying network survivability techniques have been built. Performance analysis techniques have also been characterized. Recommendations have been included for the interpretation and use of this report for continuing industry activities.

11. Bibliography

- [1] G. Brush and N. Marlow, "Assuring the Dependability of Telecommunications Networks and Services," *IEEE Network Magazine*, Vol. 4, No. 1, pp. 29-34, January 1990.

- [2] R. Cardwell and G. Brush, "Meeting the Challenge of Assuring Dependable Telecommunications Services in the '90s," *IEEE Communications Magazine*, Vol. 28, No. 6, pp. 40-45, June 1990.
- [3] T. Flanagan, S. Oxner and D. Elkaim, "Principles and Technologies for Planning Survivability – A Metropolitan Case Study," IEEE Global Telecommunications Conference (GLOBECOM) 1989, Vol. 2, pp. 813-820, Dallas, TX, November 1989.
- [4] W. Grover, "The Selfhealing Network: A Fast Distributed Restoration Technique for Networks Using Digital Crossconnect Machines," IEEE/IECE Global Telecommunications Conference (GLOBECOM) 1987, Vol. 2, pp. 1090-1095, Tokyo, Japan, November 1987.
- [5] R. Hall and S. Whitt, "Protection of SONET Based Networks," IEEE Global Telecommunications Conference (GLOBECOM) 1989, Vol. 2, pp. 821-825, Dallas, TX, November 1989.
- [6] D. Kolar and T.-H. Wu, "A Study of Survivability Versus Cost for Several Fiber Network Architectures," IEEE International Conference on Communications (ICC) 1988, Vol. 1, pp. 61-66, Philadelphia, PA, June 1988.
- [7] E. McCarthy and E. Abdou, "Network Survivability: an Integrated Planning Approach," International Switching Symposium (ISS) 1987, Vol. 3, pp. 538-544, Phoenix, AZ, March 1987.
- [8] R. McGorman, "A Methodology for Designing Survivable Telephone Networks," IEEE International Conference on Communications (ICC) 1988, Vol. 2, pp. 1172-1176, Philadelphia, PA, June 1988.
- [9] T.-H. Wu, D. Kolar and R. Cardwell, "Survivable Network Architectures for Broad-Band Fiber Networks: Model and Performance Comparison," *Journal of Lightwave Technology*, Vol. 6, No. 11, pp. 1698-1709, November 1988.
- [10] B. Wilson and R. Doverspike, "A Network Control Architecture for Bandwidth Management," IEEE International Conference on Communications (ICC) 1992, Vol. 3, pp. 1385-1391, Chicago, IL, June 1992.
- [11] D. Doherty, W. Hutcheson, and K. Raychaudhuri, "High Capacity Digital Network Management and Control," IEEE Global Telecommunications Conference (GLOBECOM) 1990, Vol. 1, pp. 60-64, San Diego, CA, December 1990.
- [12] C. Yang and S. Hasegawa, "FITNESS: Failure Immunization Technology for Network Service Survivability," IEEE Global Telecommunications Conference (GLOBECOM) 1988, Vol. 3, pp. 1549-1554, Hollywood, FL, November 1988.

- [13] T. Chujo, H. Komine, K. Miyazaki, T. Ogura and T. Soejima, "The Design and Simulation of an Intelligent Transport Network with Distributed Control," IEEE Network Operations and Management Symposium (NOMS) 1990, pp. 11.4/1-12, San Diego, CA, February 1990.
- [14] H. Sakauchi, Y. Nishimura and S. Hasegawa, "A Self-Healing Network with an Economical Spare-Channel Assignment," IEEE Global Telecommunications Conference (GLOBECOM) 1990, Vol. 1, pp. 438-443, San Diego, CA, December 1990.
- [15] W. Grover, B. Venables, J. Sandham and A. Milne, "Performance Studies of a Selfhealing Network Protocol in Telecom Canada Long Haul Networks," IEEE Global Telecommunications Conference (GLOBECOM) 1990, Vol. 1, pp. 452-458, San Diego, CA, December 1990.
- [16] R. Pekarske, "Restoration in a Flash - Using DS3 Cross-Connects," *Telephony*, Vol. 219, No. 12, pp. 34-40, September 10, 1990.
- [17] B. Coan, M. Vecchi and L. Wu, "A Distributed Protocol to Improve the Survivability of Trunk Networks," International Switching Symposium (ISS) 1990, Vol. 4, pp. 173-179, Stockholm, Sweden, May 1990.
- [18] T.-H. Wu and M. Burrowes, "Feasibility Study of a High-Speed SONET Self-Healing Ring Architecture in Future Interoffice Networks," *IEEE Communications Magazine*, Vol. 28, No. 11, pp. 33-42,51, November 1990.
- [19] Y. Kane-Esrig, G. Babler, R. Clapp, R. Doverspike, et al., "Survivability Risk Analysis and Cost Comparison of SONET Architectures," IEEE Global Telecommunications Conference (GLOBECOM) 1992, Vol. 2, pp. 841-846, Orlando, FL, December 1992.
- [20] Bellcore Technical Advisory, "Manual SS7 Traffic Management Controls at Common Channel Signaling (CCS) Network Nodes," TA-NWT-001271, Issue 2, March 1993.
- [21] Bellcore Technical Reference, "Common Channel Signaling (CCS) Network Interface Specification," TA-TSV-000905, Issue 3, December 1992.
- [22] Bellcore Technical Reference, "Bell Communications Research Specification of Signaling System Number 7," TR-NWT-000246, Issue 2, Volumes 1 & 2, June 1991, Revision 2, December 1992.
- [23] Bellcore Technical Reference, "Signaling Transfer Point (STP) Generic Requirements," TA-NWT-000082, Issue 6, March 1993.

Technical Report No. 24

- [24] "Provision of Access for 800 Service," Memorandum Opinion and Order on Reconsideration and Second Supplemental Notice of Proposed Rulemaking, FCC Docket No. 86-10, Federal Communications Commission, Washington, D.C., August 1, 1991.
- [25] International Standards Organization, ISO 7498, "Information Processing Systems – Open Systems Interconnection – Basic Reference Model," 1984.
- [26] S. Liew and K. Lu, "A Framework for Network Survivability Characterization," IEEE International Conference on Communications (ICC) 1992, Vol. 1, pp. 405-410, Chicago, IL, June 1992.
- [27] M. Mostrel, "A Transport Span Protection Strategy for Survivability of Interoffice Facilities Telecommunications Networks," IEEE Conference on Computer Communications (INFOCOM) 1990, Vol. 1, pp. 81-93, San Francisco, CA, June 1990.
- [28] Bellcore Technical Reference, "Stored Program Control Switch and Signaling Transfer Point Generic Requirements to Support E-links," TR-NWT-001204, Issue 1, November 1991.
- [29] G. Ash, J.-S. Chen, A. Frey and B. Huang, "Real-Time Network Routing in a Dynamic Class-of-service Network," 13th International Teletraffic Congress (ITC-13), pp. 187-194, Copenhagen, Denmark, June 1991.
- [30] R. Gibbens, F. Kelly and P. Key. "Dynamic Alternative Routing – Modelling and Behaviour," 12th International Teletraffic Congress (ITC-12), Vol. 3, pp. 3.4A.3.1-7, Torino, Italy, June 1988.
- [31] F. Kelly, "Routing and Capacity Allocation in Networks with Trunk Reservation," *Mathematics of Operations Research*, Vol. 15, No. 4, pp. 771-793, November 1990.
- [32] K. Krishnan and T. Ott. "Forward-Looking Routing: A New State-Dependent Routing Scheme," 12th International Teletraffic Congress (ITC-12), Vol. 3, pp. 3.4A.4.1-7, Torino, Italy, June 1988.
- [33] J. McDonald, "Public Networks – Dependable?" *IEEE Communications Magazine*, Vol. 30, No. 4, pp. 110-112, April 1992.
- [34] A. Zolfaghari, "Network Simulation Study as a Base for Survivability Standards," 13th International Teletraffic Congress (ITC-13), pp. 377-382, Copenhagen, Denmark, June 1991.
- [35] J. Sosnosky, "Service Applications for SONET DCS Distributed Restoration," *IEEE Journal on Special Areas in Communications (JSAC) on Network Integrity* (to appear December 1993).

Technical Report No. 24

- [36] F. Ellefson, "Migration Of Fault Tolerant Networks," IEEE Global Telecommunications Conference (GLOBECOM) 1990, Vol. 1, pp. 65-71, San Diego, CA, December 1990.
- [37] Bellcore Technical Advisory, "Generic Requirements for SMDS Networking," TA-TSV-001059, Issue 2, August 1992.
- [38] Bellcore Technical Reference, "Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria," TR-NWT-000253, Issue 2, December 1991.
- [39] Bellcore Special Report, "BOC Notes on the LEC Networks-1990," SR-TSV-002275, Issue 1, March 1991.
- [40] Bellcore Technical Reference, "SONET Add-Drop Multiplex Equipment (SONET ADM) Generic Criteria," TR-TSY-000496, Issue 3, May 1992.
- [41] CCITT Recommendation Q.700, "Introduction to CCITT Signalling System No. 7," 1988.
- [42] Bellcore Technical Advisory, "SONET Bidirectional Line Switched Ring Equipment Generic Criteria," TA-NWT-001230, Issue 3, April 1993.
- [43] American National Standard ANSI T1.111, "Signalling System Number Seven (SS7) – Message Transfer Part (MTP)," 1992.
- [44] American National Standard ANSI T1.105, "Digital Hierarchy – Optical Interface Rates and Formats Specifications (SONET)," 1991.
- [45] Bellcore Technical Reference, "Public Packet Switched Network Generic Requirements (PPSNGR)," TR-TSY-000301, Issue 2, December 1988, Revision 2, September 1992.
- [46] Committee T1-Telecommunications, T1X1.5/92-087, "Rationale for Unidirectional Protection Switching," August 8, 1992.
- [47] CCITT Recommendation I.355, "ISDN 64 Kbit/s Connection Type Availability Performance," June 1992.
- [48] American National Standard, ANSI T1.115-1990, "Signalling System Number 7 (SS7) – Monitoring and Measurements for Networks."
- [49] Bellcore Technical Advisory, "Network Traffic Management (NTM) Operations System (OS) Generic Requirements," TA-TSY-000753, Issue 2, December 1988.

Technical Report No. 24

- [50] T1S1.3/92-11212R1, "Report to the Signaling Network Systems (SNS) Committee on SS7 Network Architecture Evaluations and Protocol Enhancements," November 2, 1992.
- [51] FCC Report and Order 92-58, FCC Docket No. 91-273 (7 FCC Record 2010), Federal Communications Commission, Washington, D.C., adopted February 13, 1992, released February 27, 1992.
- [52] FCC Public Notice DA-92-363, "Outage Reporting Requirements and Dedicated Fax Number Announcement," April 2, 1992.
- [53] FCC Public Notice DA-92-707, "Clarification of Interim Outage Reporting," June 2, 1992.

12. Definitions

These definitions refer to network survivability and may be somewhat different from those of other documents.

Alternate Route	A second or subsequent choice path between two points.
Call Attempt	Any attempt to set up a connection for a call.
Connectivity	Node (or link) connectivity: minimum number of nodes (or links) whose removal results in losing all paths that can be used to transfer information from a source to a sink.
End User	Those who use telecommunications services; i.e., those who either originate or terminate telecommunications.
Fault Tolerance	The ability of a network element or system to continue to function under component failure(s).
Grade of Service (Traffic)	The proportion of calls, usually during the busy hour, that cannot be completed due to limits in the call-handling capability of a component in a network. For example, service objectives are defined on a per-link (per-trunk group) basis for the last-choice groups in a traffic network. Any performance objective or measure for the network. For example, the probability of a call being blocked or delayed more than a specified interval, expressed as a decimal fraction.

Layer	One of the four network survivability layers into which network functionality is decomposed (as described in Section 5). Note, unless otherwise stated, these layers are not the same as those of the ISO OSI seven layer model [25].
Maintainability	The ability of an item under stated conditions of use, to be retained in, or restored to, a state in which it can perform a required function, when maintenance is performed under given conditions, and using stated procedures and resources.
Network Availability	The probability a network can perform its required functions.
Network Failure	A complete or partial failure of a component or components of a network because of malfunction or natural or human-caused disasters. Partial failures include degradation.
Network Failure Triple	A combination of the network's Unservability (U), Duration (D) and Extent (E) parameters exceeds a threshold.
Network Integrity	See Network Survivability.
Network Performance	The level at which a network fulfills its function.
Network Reliability	See Network Survivability.
Network Restoration	Automatic or manual methods to return a network to its normal function in response to a network failure.
Network Survivability	Network survivability is: (i) the ability of a network to maintain or restore an acceptable level of performance during network failures by applying various restoration techniques, and (ii) the mitigation or prevention of service outages from network failures by applying preventative techniques.
Network Survivability Model	The analytical processes defined in Section 7.2 for estimating how well network services will be impacted and restored with transparencies to the users as a result of a failure.

Service Outage	The state of a service when network failure(s) impair the initiation of new requests for service and/or the continued use of the service and where the service outage parameters (U, D, E) do not fall into the "no outage" qualifying region.
SONET	North American Standard For Synchronous Fiber Digital Transmission [38,44].

13. Abbreviations and Acronyms

ABS	Alternate Billing System
ADM	Add-Drop Multiplexer
AEDC	Average Expected Downtime of Connection
AIS	Alarm Indication Signal
APS	Automatic Protection Switching
AT	Access Tandem
ATM	Asynchronous Transfer Mode
CCS	Common Channel Signaling, or Centa (Hundred) Call Seconds
CCSN	Common Channel Signaling Network
CCSSO	CCS Switching Office
CO	Central Office
DCS	Digital Cross-connect System
EC	Exchange Carrier (local)
EO	End Office
FDDI	Fiber Distributed Data Interface
GOF	Given Occurrence of Failure
GTT	Global Title Translation
IC	Interexchange Carrier
ISDN	Integrated Services Digital Network
ISO	International Standards Organization
MSU	Message Signaling Unit
MTBF	Mean Time Between Failure
MTBSO	Mean Time Between Service Outages
MTP	Message Transfer Part (SS7)
MTSO	Mobile Telephone Switching Office
MTTR	Mean Time To Repair
OA&M	Operations, Administration and Maintenance
OMAP	Operations, Maintenance and Administration Part
OSI	Open Systems Interconnection
OSS	Operations Support System
PCN	Personal Communications Network

Technical Report No. 24

PDDC	Probability Distribution of Downtime of Connection
ROF	Random Occurrence of Failure
SCCP	Signaling Connection Control Part
SCP	Service Control Point
SEP	Signaling End Point
SHN	Self Healing Network
SHR	Self Healing Ring
SONET	Synchronous Optical NETwork
SPE	Synchronous Payload Envelope (SONET)
SS7	Signaling System Number 7
SSP	Service Switching Point
STE	Span Terminating Equipment
STP	Signaling Transfer Point
TCAP	Transaction Capabilities Application Part
TSI	Time Slot Interchange
ULE	User Lost Erlang
VT	Virtual Tributary (SONET)

Appendix A. Telecommunications Service Priority System

Background:

Since 1967, a system has been in existence for prioritizing the restoration of commercially supplied telecommunications for the U.S. Government. This system has been known as the Restoration Priority (RP) System.

Restoration priorities have been assigned to leased, intercity, private line circuits that are deemed critical to the National Security Emergency Preparedness (NSEP) posture of the nation. The assignment of these priorities has been administered by the National Communications System (NCS), in Arlington, Va., for the FCC (Federal Communications Commission).

In 1985, following the divestiture of AT&T, the NCS undertook to revamp the RP System. A joint effort by the telecommunications industry and the U.S. Government, under the auspices of the National Security Telecommunications Advisory Committee (NSTAC), a Presidential advisory body, developed the Telecommunications Service Priority (TSP) System.

TSP Rule:

On Dec. 1, 1989, the FCC issued Declaratory Ruling DA 98-1524 that set the effective date of the beginning of the TSP System to be Sept. 10, 1990. This date is referred to as the Initial Operating Capability (IOC) date by the government. This IOC date also marks the end of RP-type code assignments and the beginning of a transition period, during which all end users will be required to submit their current circuits with RP codes for recertification under the new TSP criteria. All

Technical Report No. 24

existing RP circuit designations will remain in effect until March 1993 or until recertification by the user under the TSP criteria.

The TSP System officially replaces the RP System under Part 64, Appendix A, of the FCC Rules and Regulations, Chapter 1 of Title 47 of the CFR.

TSP Documents:

The supplemental documents that address the total TSP System include the TSP Service Vendor Handbook (NCSH 3-1-2), the TSP Service User Manual (NCSM 3-1-1) and NCS Directive 3-1 that covers the government's participation in the system.

TSP Requirements:

The TSP System requires all common carriers to provision new services and restore designated services on a priority basis. Since the FCC has approved and codified the TSP System, common carriers may now provide preferential treatment to specially designated user requests for service.

Priority levels will be assigned to NSEP telecommunications services that will specify the order in which provisioning or restoration of the services is to occur relative to other NSEP and non-NSEP services. The authorized priority levels are designated as follows, highest to lowest: E (Emergency), 1, 2, 3, 4 and 5 for provisioning and 1, 2, 3, 4 and 5 for restoration. The responsibility of any common carrier is to:

- a. Provision NSEP telecommunications services before non-NSEP service requests, when NSEP treatment is properly invoked and a 12-character, TSP Authorization Code is provided by the user/contract officer. Allocate resources to ensure the company's best efforts to provide NSEP services according to and in order of the provisioning priority.
- b. Restore NSEP telecommunications services before non-NSEP services and in order of priority level. Allocate available resources to restore as quickly as practical. Broadband or multiple service facilities restoration is permitted even though it might result in restoration of services with no priority or lower priority level.

Impacts of TSP:

The following is a list of the impacts of TSP:

- a. Provisioning:
 - The customer must provide a TSP Authorization Code to receive preferential treatment.
 - A TSP Provisioning Priority must be assigned as the eleventh digit of the Authorization Code to receive accelerated service installation.

Technical Report No. 24

Provisioning priorities are either EMERGENCY (designated by an "E") or ESSENTIAL (designated by a 1, 2, 3, 4 or 5). The FCC Order requires that service suppliers provision Emergency services before any Essential TSP service or non-TSP service.

- All common carriers must provide a 24 hour point-of-contact to receive properly-invoked provisioning requests.

b. Restoration:

- Network Operations must dispatch service personnel outside normal business hours to restore services
 - assigned priority 1, 2 or 3,
 - assigned 4 or 5, if the next business day is more than 24 hours away.
- Operations may preempt non-TSP or lower priority TSP circuits to restore higher priority TSP circuits.

c. Administrative:

- The TSP System allows state and local governments, together with private entities, to request TSP priority assignments through the use of Federal sponsorship.
- Industry must provide a separate notice to the NCS upon completion of a service installation, modification or cancellation in addition to any notification which is supplied directly to the customer/contract officer.
- All TSP records must be reconciled with the NCS annually and every three years with subcontractors.
- The TSP System provides the basis (via tariff or contract) for carriers to charge for record management, accelerated provisioning and restoration services.

Appendix B. Example Network Survivability Analyses and Assessments

Appendix B.1 Bidirectional Ring Survivability Example As an example at the GOF model, consider a bidirectional ring network with $N=6$ nodes connecting to a central node as shown in Figure B.1. Here S_a is the fraction of remaining nodes connected to a central node (CN) under a disaster causing N_u nodes to be destroyed. Depending on which nodes are failed, the value of S_a may be different.

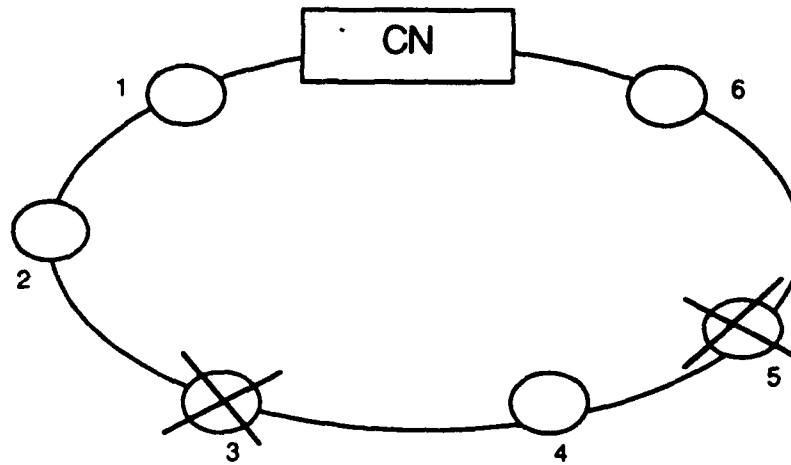


Figure B.1: Bidirectional Ring Example

It is clear that N_U is a random variable with values $N_U=0, \dots, N$. In this example, the number of failed nodes and the number of ways $N_U=2$ can occur is tabulated in Table B.1. The set of nodes connected to the central node, $N_a=0, \dots, N-N_U$, is also shown. Using this information, the fraction of nodes connected given two failed nodes can be easily obtained.

Assuming the two-node failure case (all equally probable), there are 15 ways of choosing the failed pair. For five failed pairs, the number of connected nodes is four, and for four failed pairs, the number of connected nodes is three, and so forth. The probability frequency function $P[S_a=y]$ is shown in Figure B.2 for $N_U=2$.

$\{N_U\}$ number of failed nodes	number of ways to occur	$\{N_a\}$ number of connected nodes (number of occurrences)
2	15	4(5), 3(4), 2(3), 1(2), 0(1)

Table B.1: Failure Categories

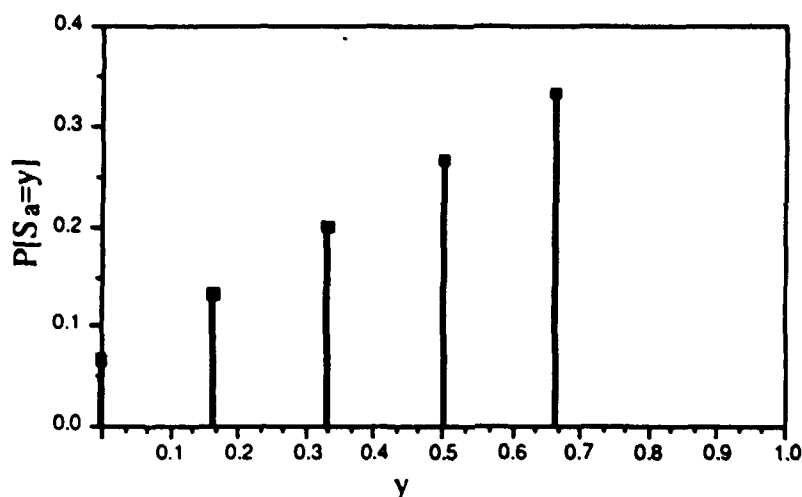


Figure B.2: Probability Frequency Function for $N_u=2$

The probabilistic nature of S_a is well demonstrated in Figure B.2. For the bidirectional ring network shown in Figure B.1, when two nodes are destroyed, there is a 33% probability that 66% of the nodes are connected to the CN, and there is about a 6% chance that no nodes are connected to the CN.

Appendix B.2 Network Survivability Assessment Example In an 8-node, 7-link, fixed route network, shown in Figure B.3, total $D_t = 56$ point-to-point DS-3 demands are distributed and routed over the links of the network prior to failure, as indicated in Table B.2. There are $C_t = 10$ demand pairs in this network.

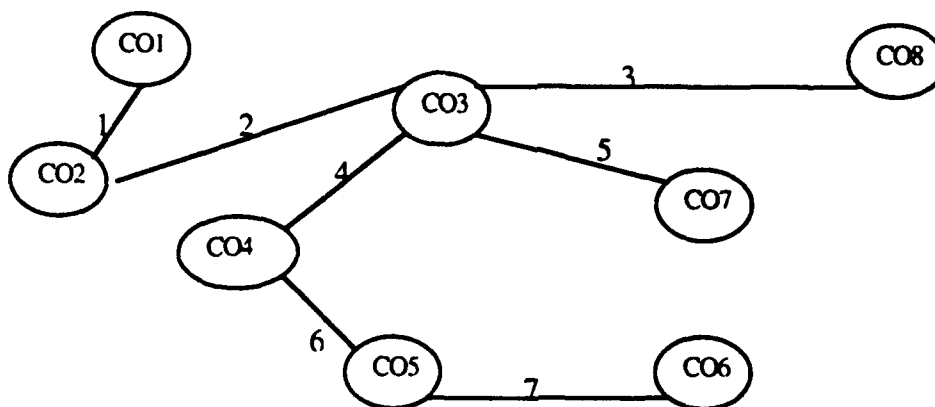


Figure B.3: 8-node, 7-link network Example

For simplicity, we assume a sample space of a single link failure and that the set of states after failure is limited to $\{x_i\}$, where event x_i denotes the failure of link i . We want to assess the survivability of this network, in terms of three survivability measures, (i) average fraction of traffic (DS-3s) that remains after failure (or

Technical Report No. 24

Average Traffic Survivability Ratio), S_a , (ii) average fraction of subscribers that remain connected after failure (or Average Connectivity Ratio), C_a and (iii) Average Time for single link Restoration, t_R .

Node Pair	Demand (DS-3s)	Routing plan
CO1 - CO3	5	1-2
CO1 - CO6	11	1-2-4-6-7
CO1 - CO8	8	1-2-3
CO1 - CO4	4	1-2-4
CO1 - CO7	5	1-2-5
CO6 - CO8	6	7-6-4-3
CO6 - CO7	4	7-6-4-5
CO5 - CO8	6	6-4-3
CO5 - CO7	3	6-4-5
CO4 - CO8	4	4-3

Table B.2: Point-to-point demand and routing table for Example B.2

By considering the network routing plan, aggregated load, L_i , on the i^{th} link can be found. If a given link fails, L_i DS-3s on that link would be lost, or

$$L_i = \text{sum of all demands going through link } i$$

and we have:

$$S_a(x_i) = (D_t - L_i) / D_t$$

Where $S_a(x_i)$ is the Traffic Survivability Ratio, assuming the i^{th} link has failed. For example for link 1, $L_1 = 33$ DS-3s and if this link fails 33 DS-3s out of 56 DS-3s would fail and only 23 DS-3s would survive. Thus $S_a(x_1) = 23/56 = 0.41$. The value of $S_a(x_i)$ for all links ($i=1,2,\dots,7$) are tabulated in Table B.3.

As indicated in Table B.2, there are $C_t = 10$ point-to-point "connections" in this network. Each link carries n_i of these connections and we have

$$C_a(x_i) = (C_t - n_i) / C_t$$

where $C_a(x_i)$ is the Connectivity Ratio, assuming the i^{th} link has failed. For example $n_1 = 5$ connections go through link 1. Therefore if link 1 fails, half of all network connections would be disconnected, or $C_a(x_1) = 5/10 = 0.5$. The values of $C_a(x_i)$, $i=1,2,\dots,7$ are also tabulated in Table B.3.

Technical Report No. 24

Similar to example B.1, we can now attach a probability to each event, x_i . If we assume that each event, x_i , has equal probability, the expected values of the random variables are

$$E[S_a] = 1/K \sum_i S_a(x_i),$$

$$E[C_a] = 1/K \sum_i C_a(x_i),$$

$$E[t_R] = 1/K \sum_i t_r(x_i),$$

where K = total number of links ($K = 7$ in this example). $t_r(x_i)$ is the estimated Restoration Time for the i^{th} link using normal restoration methods and can be obtained using long term link outage data. Note that $t_r(x_i)$ is a function of the length of link i . Using Table B.3, we have:

$$E(S_a) = 0.51, \quad E(C_a) = 0.54 \quad \text{and} \quad E(t_R) = 70.7 \text{ minutes}$$

Link No.(i)	L_{gi} (DS-3s)	n_{ci}	$S_a(x_i)$	$C_a(x_i)$	$t_r(x_i)$ (minutes)
1	33	5	0.41	0.5	15
2	33	5	0.41	0.5	160
3	24	4	0.55	0.6	55
4	38	7	0.32	0.3	45
5	12	3	0.78	0.7	100
6	30	5	0.46	0.5	35
7	21	3	0.62	0.7	85

Table B.3: Link information for Example B.2

Appendix C. User Expectations

This section discusses the expectations of certain users or user groups with respect to service continuity and survivability.

Appendix C.1 U.S. Government User Expectations Government agencies have diverse and unique user expectations such as Telecommunications Service

Technical Report No. 24

Priority²² (TSP), security, interoperability, compatibility, redundancy, diversity, hardening, prevention of loss or corruption of critical data, measurable Reliability/Maintainability/Availability (RMA) performance parameters²³, and responsiveness. For example, the U.S. Federal Aviation Administration (FAA) and Department of Defense (DoD) define and require several types and scopes of communications services and are concerned about the causes of probable network failures shown in Table C.1. The users of the service types defined in Table C.1 depend upon these availability and restoration time requirements.

Service Types	Service Scope	Failure Causes	Expected Availability	Expected Restoration Time
Critical	Functions or services which, if lost, would <u>prevent</u> the network/system from exercising safe operation and control for the end users.	Earthquakes, floods, hurricanes, tornadoes, acts of war, hardware/software failures, electromagnetic pulse, etc.	0.9999987	< 6 seconds
Essential	Functions or services which, if lost, would <u>reduce</u> the capability of the network/system to exercise safe operation and control for the end users.	Terrorism, fires, power/battery failures, etc.	0.999	approximately 10 minutes
Routine	Functions or services which, if lost, would <u>not significantly</u> degrade the capability of the network/system to exercise safe operation and control for the end users.	Node and link failures, subsystem failures, equipment failures, etc.	0.99	approximately 2 hours

Table C.1: Government Communications Services²⁴

²² See Appendix A for details on TSP.

²³ Definitions of terms used here and in government acquisitions are provided in this report and by U.S. Federal Standard 1037B, Telecommunications Glossary.

²⁴ The expected availability values recorded here require information as to the population and interval over which they have been calculated.

In national and international emergencies, the government expects networks/systems to be capable of providing survivable communications for national decision makers, executing crisis management control, offering distributed control of information, and restoring networks and systems. To provide for these expectations, networks/systems should, at a minimum, have the following survivability features and capabilities:

Connectivity: The DoD requires connectivity to provide the capability of connecting several networks. In order to react to rapidly changing environments and to be able to easily and quickly reconfigure networks, the ability to communicate facility status to network management centers must be provided.

Security: Systems that provide telecommunications services must be made secure from both internal and external threats such as unauthorized access, sabotage, hackers, and terrorism. Survivability must address and counter these threats through the use of security services and mechanisms such as: identification, authentication, and authorization of users, operators, and maintainers of the system; access control; database integrity; and encryption of network control messages. Multi-level security must not be impacted by network outages.

Endurability: The communications network must be capable of operating using commercial power, uninterruptable power, and back-up battery systems during failures. In addition to a network's capability to withstand hardship, stress, and adversity during given periods of time and levels of network performance. Network designs should also include counter measures for survivability against threats of nuclear and High-Altitude Electro-Magnetic Pulses (HEMP).

Interoperability: The communications network must be capable of operating with all agencies and commercial providers.

Compatibility: The communications network must have the ability to coexist and not interfere with the operations of other networks/systems.

Survivability Performance Parameters: Numerical values for these parameters should be provided because they are relevant to the network management of government long-haul and base-level networks. These parameters are a function of the following factors:

- the type of switching and transmission technology to be employed in the network under consideration (i.e., SONET, ISDN, ATM Switching, FDDI),
- the robustness of the network management technology used to control these network resources,
- network topology, and

- the required level of network throughput efficiency, based on economic constraints.

Appendix D. Tolerance Categories for Restoration Times

To design networks to meet user expectations, information is needed about the impact of failures on network services. This information can be used in standards activities for determining restoration times for network facilities as well as by network planners to deploy appropriate network survivability techniques for each category of customer. One of the most crucial parameters for determining the impact of a service outage is the outage duration (the outage duration here is defined to be the time interval between first loss of a particular unit of usage until that unit of service is fully restored). Note that actual service outage time will often exceed facility restoration time. More detail on this topic can be found in [35]²⁵.

Different services have different outage duration tolerances. For example, voiceband call tolerances (i.e., the interval in which calls in progress are abandoned) can vary anywhere from 150 milliseconds (see Table D.2) to two seconds while data (packet) session timeout can vary from two to 300 seconds. Today, session dependent applications such as file transfer commonly use System Network Architecture (SNA) and TCP/IP protocol architectures. With SNA, in particular, a session dependent application has a software programmable session time-out of from 1.1 to 255 seconds [36] and can be specified by users based on file size, for example. Lost data from shorter interruptions is retransmitted, a process which in some applications, is triggered by receiver time-outs. Receivers may begin to time-out based on twice the round-trip delay.

X.25 packet networks often incorporate idle channel state condition timers that are settable from one to 30 seconds, in one second increments (with a suggested time of 5 seconds) [45]. When links are lost between switches, these timers may expire, and if they do, they will trigger the disconnection of all virtual calls that were up on those links. The customers must then restart their sessions. It is common (but not required) in these networks to feature a virtual circuit reconnect capability, through alternate routing around the failed links, but only after the call is reestablished.

In Asynchronous Transfer Mode (ATM) networks (e.g., those supporting Switched Multi-megabit Digital Service (SMDS) [37]), normal cell routing processes are interrupted and rerouting processes may be started after a certain interval. This interval can be estimated at 200 milliseconds based on SMDS objectives in [37].

²⁵ A list of references is in Section 11.

Technical Report No. 24

The objective in SMDS supporting networks for a switch/router to update its routing tables upon notification of a link set change through lower-layer OSI protocols is less than 100 milliseconds. It is reasonable to expect that within 200 milliseconds from the onset of the link outage, that the rerouting process should have been started, allowing time for topology updating information to propagate to all switches (50 milliseconds), and delay (50 milliseconds) for waiting on physical level protection, if any (e.g., point-to-point systems using Automatic Protection Switching). However, in large or more complex networks this interval may be longer than 200 milliseconds. The recovery of any lost data is handled through higher layer OSI data protocols.

Figure D.1 shows the impact on various network services as a function of restoration times for SONET-based architectures due to a network failure. Outage durations include restoration times at the SONET level. The restoration times are classified into intervals as illustrated in this figure. Restoration times within a given interval will have roughly the equivalent impact on services. Restoration times under 50 milliseconds will be transparent to most services. These time intervals are tabulated in Table D.1.

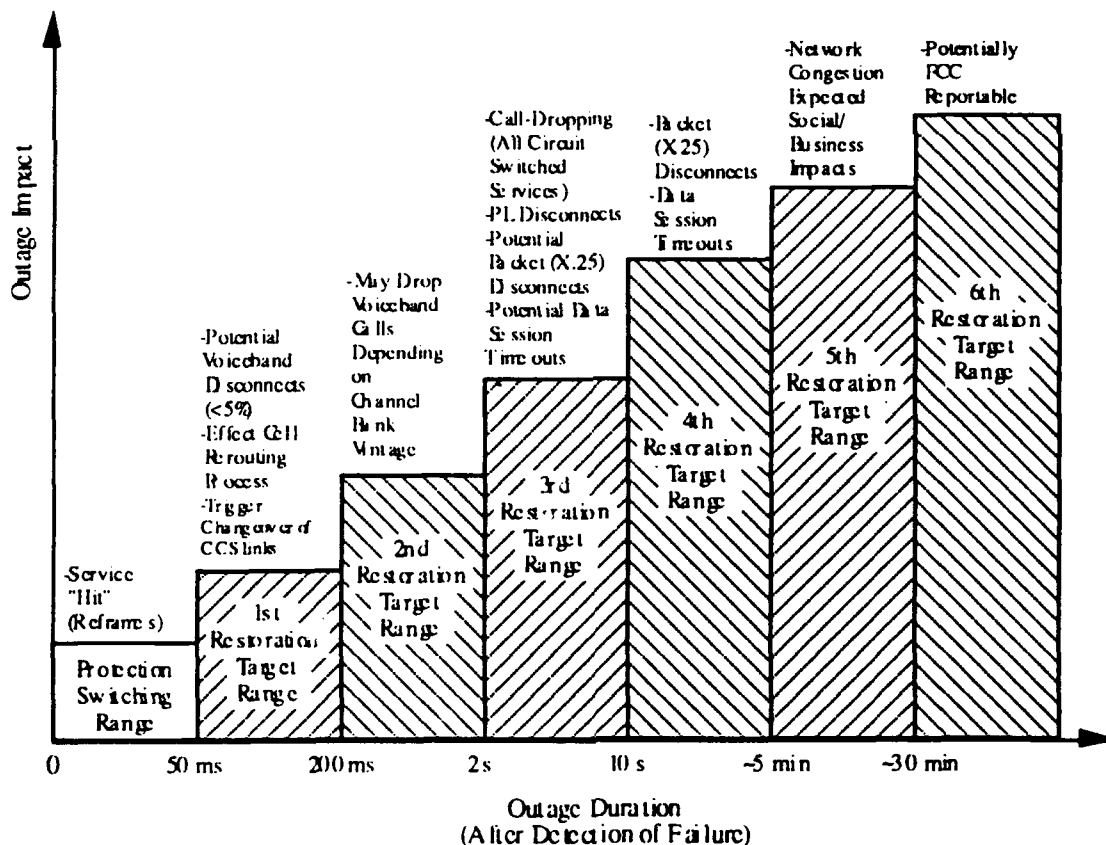


Figure D.1: Restoration Time Impact on Customers

Technical Report No. 24

Time Interval	Range	
	From	To
1	50 milliseconds	200 milliseconds
2	200 milliseconds	~2 seconds
3	~2 seconds	~10 seconds
4	~10 seconds	~5 minutes*
5	~5 minutes*	~30 minutes ²⁶
6	~30 minutes ²⁶	unspecified

*These restoration time intervals and their impacts should be studied.

Table D.1: Restoration Time Intervals

Restoration times below 50 milliseconds will meet network protection switching time requirements at the SONET level [38,44]. Recovery within this time frame will be without interruption of service (a service "hit"). A "hit" is a temporary interruption of service that causes only a reframing of a distant terminal (e.g., digital channel bank) off of the SONET backbone. A "hit" does not cause a transmission failure alarm to be issued at the distant terminal.

For a circuit switched network, the minimum disconnect timing interval for a possible false disconnect to occur at a downstream switch trunk interface is 150 milliseconds [39] (i.e., 150 milliseconds on-hook supervisory signal state). A failure in the interoffice trunk network may cause the supervisory signal state to appear as on-hook, rather than off-hook, to the switch. The network allocation of this 150 milliseconds is as follows:

	+ 10 milliseconds	failure detection time
	+ 40 milliseconds	reframing of lower-rate multiplexes
	<u>+ 50 milliseconds</u>	reframing of distant terminal
Sub-total	100 milliseconds	
	<u>+ 50 milliseconds</u>	switching time (SONET level)
Total	150 milliseconds	

Table D.2: Minimum Trunk Disconnect Time

The time frame between 50 milliseconds to just under 200 milliseconds will have minimal impact to services. Any affected voiceband switched calls (voice or data) will have less than a 5% probability of being dropped (related to signaling freezing limitations with in-band signaling). Keeping the restoration time to under 200 milliseconds will prevent any affected voiceband circuit-switched calls that were

²⁶ This value is based on FCC outage reporting requirements and should be reviewed.

on trunks associated with older channel banks from being dropped. Note that a 200 millisecond restoration time appears as a 300 millisecond outage to a channel bank (when including detection time and reframe times). Old channel bank types (e.g., D1A, D1D, D2, D3) exhibit carrier group alarm times of 300 milliseconds to 500 milliseconds. In this same time frame, the normal cell routing process in cell-relay networks (e.g., supporting SMDS) should have been interrupted, and the cell rerouting process started. Again, this estimated 200 millisecond interval for cell-relay networks may be longer in large or more complex networks.

The CCSN will provide better performance by virtue of being out-of-band and separate from the path of the call. The CCSN uses DS-0 circuits for signaling links that are bunched together on a DS-1 trunk (carrier). Calls only in the process of being set-up may be lost with an outage on the signaling links; established calls should not be affected. Note that according to SS7 specifications [22], a signaling link is considered failed and taken out of service when a loss of alignment (i.e., no signaling flags) lasts for approximately 146 milliseconds at the STP. A changeover to the alternate signaling link is then initiated. A restoration time in the time frame between 50 milliseconds to just under 200 milliseconds will trigger this changeover.

The second time frame is between 200 milliseconds to just under two seconds. Performance of the voice service is only slightly degraded in this range because only affected voiceband circuit-switched calls that are on trunks associated with old channel bank types will be dropped. Network data from one local exchange provider indicates that roughly 12% of DS-0 circuits that are carried on channel banks (including DS-0s that carry trunks of circuit switched networks as well as non-switched circuits) are in fact carried on older type channel banks. Thus, the total number of trunks in circuit switched networks that are carried on older type channel banks is less than 12% because significant numbers of DS-0 trunks do not route over channel banks (for example, DS-0 trunks that directly interface into digital switches at the DS-1 rate). Video services can become degraded in just 1/30 second (loss of one frame). Furthermore, restoration times greater than 100 milliseconds have been shown to have significant impact on the video codec reframe process and can become customer reportable impairments [46]. Nevertheless, an actual outage of less than two seconds is not critical for entertainment/educational type video services (one to two seconds is close to human reaction time). This observation applies to video services with no protocol for recovery of lost frames.

The third time frame is between two seconds to just under ten seconds. The ten second upper bound is consistent with the SONET carrier group alarm requirements [38,44]. When outages exceed two seconds, DS-0, n×DS-0, and DS-1 call-dropping will occur. Private line disconnects will also occur. Voiceband data modems will time-out (typically two to three seconds) after detecting a loss of

Technical Report No. 24

incoming data carrier. The modems must then be resynchronized (usually a 15 second procedure) after the data carrier is restored.

B-ISDN calls in the future may have alarm thresholds set higher than two seconds (e.g., five seconds), but B-ISDN calls would still be dropped in this time range, assuming the outage lasts longer than five seconds. In this time range there is a potential for packet (X.25) virtual calls to be dropped depending on the idle link timer system setting. Disruption of data communications sessions (connection oriented, e.g., SNA) may occur in this range depending on the session termination time. High-speed, interactive applications (for example, video image) may be impacted in this time range since these applications are sensitive to delay. This time range is also unattractive for high-priority data applications, for example, transfer of funds using automated teller machines. Automated teller machines use private data communications networks based mostly on SNA and TCP/IP protocol architectures. Services that are less sensitive to delay such as electronic mail (e-mail) should not be greatly impacted. Frame relay data services may be sensitive to service outages because of their high-speed and lack of ability of the frame relay network layer protocol (roughly equivalent to the third layer of the OSI stack) to recover lost data.

The fourth time frame is between ten seconds to just under five minutes²⁷. In this fourth time range, packet (X.25) calls and data communications sessions will be disrupted and customers would attempt to reinitialize their connections. The five minute upper bound is not meant to be a firm number based on particular software or hardware specifications, but rather an estimate based on general planning goals for various ECs. For example, the five minute restoration time is the goal for completing service restoration by some network management systems and is considered responsive to most customers. Also, the five minute upper bound covers the maximum SNA session timeout of 255 seconds.

In the fifth time frame, digital switches in switched networks may experience a buildup of network congestion. There may be minor social and business impacts with this outage event. The sixth time frame is based on the 30-minute FCC requirement on carriers that major outages (affecting at least 30,000 or 50,000 customers²⁸ and exceeding 30 minutes) be reportable.

²⁷ Based on industry tradition and management systems. This value should be validated.

²⁸ The FCC required 50,000 lines [51,52], with 30,000 lines on a trial basis [53]. The unanimous recommendation of the Threshold Reporting focus team of the NRC on December 15, 1992 was 30,000 lines.

Technical Report No. 24

Index

ABS 34
Add-Drop Multiplexer 42
AIS 26
APS 21
ATM 75
Automatic Protection Switching 21
B-ISDN 59, 80
Bidirectional Ring 22
BR 22
Broadband DCS 16
Call Attempt 10
CCS 33
CCSNs 4, 17, 33
CCSSO 34
Centralized Control 26
Compatibility 75
Congestion Control 33
Cross-connect 4
Data Communications Network 14
DCS 12
Detection 18
Digital Cross-connect System 12
Disaster 7
Disaster Recovery 8
Distributed Control 18
Diversity 6
DS-n 15
Duplex 21
Duration 3
Dynamic Path Generation 19
Dynamic Routing 29
EDSX 12
Efficiency Ratio 50
Engineered Blocking 10
Erlang 17
Exchange Carrier 8
Extent 8
Fault Isolation 18
FDDI 75
Flooding 30
Frame Relay 32
GOF 44
GTT 38
Integrated Techniques 42
Interexchange Carrier 8
ISDN 33
ISO 32
Link Rerouting 28
Link Set 42
Loss of Connectivity 56
Loss of Traffic 51
Media 14
MSU 54
MTBF 50
MTBSO 50
MTP 33
MTTR 50
Network Integrity 1
Network Management 32
Network Reliability 1
Network Survivability Analysis Model 44
Normalization 19
NRC 5
OC-n 15
OMAP 33
Open Systems Interconnection 32
Operations Support System 18
OSI 32
OSI Layer 32

Technical Report No. 24

OSS 18	TSP 68
Outage Duration 76	UDE or (U, D, E) Triple 8, 10, 48, 58, 66
PDDC 56	Unidirectional Ring 22
Point-to-point Rerouting 28	Unservability 10, 50
POTS 54	UR 22
Probability Distribution of Downtime of Connection 56	User Lost Erlang 48
Reconfiguration 17	Verification Time 49
Reliability 7	Verified Restoration Time 49
Restoration Ratio 49	Virtual Tributary 16
Restoration Switching 26	VT 16
Restoration Time 7	Wideband DCS 16
Ring 21	
ROF 45	
SCCP 33	
SCP 33	
SEP 34	
Service Unservability 50	
Signaling Connection Control Part 33	
Signaling End Points 34	
Signaling System Number 7 (or SS7) 33	
Signaling Transfer Points 33	
SONET Line 23	
SONET Path 25	
SONET Section 12	
SPE 24	
SS7 33	
SSP 34	
STP 33	
STS-n 16	
Survivability Measure 1	
TCAP 33	
Telecommunications Service Priority 67, 68	
Time Slot Interchange 24	
Traffic Holding Time 16	
TSI 24	