AVAILABILITY-AWARE PROVISIONING IN *P*-CYCLE-BASED MESH NETWORKS

Amin Ranjbar

A thesis

in

The Concordia Institute of Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Applied Science in Information
Systems Security
Concordia University
Montréal, Québec, Canada

September 2008

Your file Votre référence
ISBN: 978-0-494-45471-8
Our file Notre référence
ISBN: 978-0-494-45471-8

Abstract

Availability-Aware Provisioning in p-Cycle-Based Mesh Networks

Amin Ranjbar

Survivability of high-capacity optical wavelength-division multiplexing (WDM) mesh networks has received much research attention for many years now. These networks are typically designed to survive single component failures. The method of pre-configured protection cycles (p-cycles), recently proposed by W. Grover's research group, promises to achieve ring-like high speed protection with mesh-like high efficiency in use of spare capacity. In such networks, which are designed to withstand only single failures, service availability comes to depend on dual-failure (or more) considerations. Hence, availability-aware service provisioning emerged as a topic of great importance in the past few years.

In this thesis, we first revisit the problem of availability analysis in p-cycle based networks and present an accurate model for availability-aware provisioning after highlighting major flaws in prior work. Our model provides a technique for allocating p-cycles to restore single link failures such that the unavailability of all the demands in the network is bounded by an upper limit. We then provide some heuristics for restricting the number of variables and constraints in an integer linear programming formulation in order to solve our problem in a reasonable amount of time.

Failure-Independent Path Protecting (FIPP) p-cycle recently has been proposed as an extension of the basic p-cycle to provide a pre-connected, failure independent, path-protecting network design. We present in this thesis the first model for availability-aware provisioning in FIPP based networks. Our study focuses on determining whether FIPP will maintain its resource efficiency advantages over span p-cycles when the network design is based on limiting the service unavailability.

Acknowledgments

I would like to take this opportunity to express my sincere thanks and gratitude to my supervisor, Dr. Chadi Assi for his continuous guidance throughout my research work. His support has always been generous and timely. He has given me invaluable insight in my research domain while his encouragement made my thesis work an extremely educational experience. I would further like to express my appreciation to Dr. Brigitte Jaumard for her helps to improve my thesis.

I thank my colleague, Mr Seyyed Mohammad Kiaei with whom I have coauthored research papers. Our discussions contributed to this thesis greatly. I enjoyed cooperating with him on various projects.

I would also like to thank all the members of Concordia Institute for Information Systems Engineering (CIISE) for creating such a dynamic and collaborative environment for research and study.

Last, but not least, I would like to express my warmest and deepest thanks to my parents for their self-giving love and support throughout my life. Without them, I would never be able to complete my study at Concordia University.

Contents

List of Figures

List of Tables

List of Publications

- Amin Ranjbar and Chadi Assi. "Insights in Using Availability-Aware Design for p-Cycle and FIPP Based Networks", *5th workshop on the Optimization Optical Networks (OON 2008), Montreal, Canada.*

- Amin Ranjbar and Chadi Assi. "Availability-Aware Design in FIPP p-Cycles Protected Mesh Networks", *IEEE proceeding, International Conference on Optical Networking Design and Modeling (ONDM 2008), Vilanova i la Geltr, Catalonia, Spain.*

- Amin Ranjbar and Chadi Assi. "An Improved Approach for Enhanced Availability Analysis and Design Methods in p-Cycle-Based Networks", *4th workshop on the Optimization Optical Networks (OON 2007), Montreal, Canada.*

- Amin Ranjber and Chadi Assi. "Availability-Aware Design in Mesh Networks with Failure-Independent Path-Protecting (FIPP) p-Cycles", under second revision for *IEEE Transactions on Reliability.*

List of Acronyms

ACK – Acknowledgement

ADM – Add/Drop Multiplexers

APS – Automatic Protection Switching

BIP – Binary Integer Programming

BLSRs – Bi-directional Line Switched Rings

CG – Column Generation

CIP – Critical Infrastructure Protection

CSP – Cutting-Stock Problem

FIPP – Failure-Independent Path-Protecting

IEEE – Institution of Electrical and Electronics Engineers

ILP – Integer Linear Programming

IT – Information Technology

LP – Linear Programming

MIP – Mixed Integer Programming

MRCP – Multi Restorability Capacity Placement

MTTF – Mean Time to Failure

MTTR – Mean Time to Repair

OADMs – Optical Add-Drop Multiplexers

OPPR – Optical Path Protection Ring

OSPR – Optical Shared Protection Ring

OXCs – Optical Cross-Connects

***p*-Cycles** – Pre-configured or Pre-connected Protection Cycles

RMP – Restricted Master Problem

SBPP – Shared-Backup Path Protection

SCA – Spare Capacity Assignment

SEACP – Selectively Enhanced Availability Capacity Placement

SHR – Self-Healing Ring

SLA – Service Level of Agreements

UPSRs – Unidirectional Path-Switched Rings

WDM – Wavelength- Division Multiplexing

Chapter 1

Introduction

Survivability of high-capacity backbone mesh networks based on wavelength division multiplexing (WDM) technology has been an issue of great practical importance in recent years; this is due to the fact that any service interruption with connections operating at high bit rates (typically 10 Gbit/s or higher), even for a short period of time, may lead to a tremendous waste of (critical) data information. Network operators of these new generation of backbone networks have to face their customers' requirements, which are usually very strict on the outage period. Therefore, preventing service interruption and minimizing the loss of service must be considered in designing telecommunication networks.

Optical networks, using WDM technology, have enabled the provisioning of fast and high-quality connections to end users. Using optical fiber as a transmission medium achieves many advantages such as low loss, light weight, electromagnetic immunity, high bandwidth and low cost [2]. Currently, optical fiber is widely deployed in backbone networks and the number of users relying on this type of communications is growing rapidly. In WDM networks, where up to 100 terabits data information per second is transmitted over a single fiber, service outages can have disastrous consequences.

Previously, component failures were manually repaired by re-routing the failed connections and sending teams to repair the damaged links or nodes. However, currently, optical networks requiring manual re-routing are considered as unprotected networks and the outage periods due to traffic recover based on the human intervention are unacceptable. Hence, optical network operators are not willing to accept unprotected facilities: survivability must be always guaranteed by adopting efficient techniques of automatic recovery from failures, that is to say re-routing broken connections automatically [3].

1.1 Overview of WDM Networks

Wavelength-Division Multiplexing (WDM) has emerged as the most promising technology for taking the full advantage of the bandwidth potential of fiber and thereby satisfying the increasing demands for bandwidth. WDM is an approach that allows multiple WDM channels from different end-users to be multiplexed on the same fiber. Each of these channels operates at data rates around 10Gbps and it is likely to extend to 80Gbps. Since all components in a WDM device need to operate only at electronic speed, WDM devices are easier to implement. Extensive research and development on optical WDM networks have been done over the past few years; hence the next generation of the Internet will employ WDM-based optical backbone.

The next generation of optical networks is a complex system designed according to a layered approach. Higher layers of these kinds of network are fully managed by electronic equipments. Several protocols can be stacked one over the other in different combinations. The WDM optical layer behaves as a common platform with the ability to carry all the possible protocol combinations; hence the WDM layer has been standardized as a circuit-switching oriented multi-protocol transport level.

Accordingly, the main task of the WDM optical layer is connectivity and bandwidth provisioning to the electronic layers in a client-server relationship [4]. In such a network, end users will attach to the network through a wavelength-sensitive switching/routing node [5] and the WDM networks offer a provisioning service which consists of setting up optical point-to-point circuits in order to fulfill requests of end-users. Such requests are referred to as virtual connections and an optical circuit is named a lightpath. In other words, a lightpath is an optical connection carried end-to-end from a source node to a destination node over a wavelength on each intermediate link [6]. At intermediate nodes in the network, the lightpaths are routed and switched from one link to another. Each lightpath carries a high bit-rate digital stream. It is added and dropped by electro-optical devices interfacing the WDM layer to the higher electronic layers. In some cases, lightpaths may be converted from one wavelength to another wavelength along their paths. WDM switching is performed either by optical add-drop multiplexers (OADMs) or by optical cross-connects (OXCs), according to the type of network architectures.

1.2 Motivations and Preliminaries

Critical Infrastructure Protection (CIP) is a national program whose objective is to build capabilities to strengthen and support the robustness, reliability, resilience, availability and protection of physical and information technology (IT) facilities, networks, services, and assets, which if disrupted or destroyed would have a serious impact on the health, safety, security, economic well-being, or effective functioning of the nation [7]. Telecommunication networks, using WDM technologies, which is based on cable equipments is classified as critical infrastructure, and therefore its protection is of utmost importance. Accordingly, it is necessary to assess the vulnerabilities and provide various protection techniques, and evaluate the resiliency and availability of these kind of networks.

According to annual statistics, metro and long haul networks experience lots of fiber cuts [2]. In the first eight months of 2002 alone, 116 network outages were recognized in the United States with wide-ranging effects [2]. Similarly, on February 13, 2002 in Yadkinville, NC, workers severed a Sprint cable during repairing a water line and cut around 52 trunk groups for over 5 hours [2]. In Maryland, fire in a power transformer melted a Verizon fiber cable and affected more than 5000 customers for over 9 hours only a week after. In addition, on August 14, 2003, Ontario and much of the northeastern U.S. were hit by the largest blackout in North America's history. Electricity and all communication services were cut to 50 million people, bringing darkness and service outages to customers from New York to Toronto up to North Bay. While some communication services' consumers had service restored early the next morning, many areas remained in the same situation for the next day and even the following one. Statistics show that metro networks annually experience 13 cuts for every 1000 miles of fiber and long haul networks experience 3 cuts for 1000 miles fiber [2].

Most of the previous work on protection techniques for survivability of optical networks has focused on 100% restorability against all single span failure. Obviously, single failure restorability is of great overall benefit to network integrity, but that qualitative recognition does not guarantee that the availability of the service path will be 100%. This makes the dual failures as the main factor which can contribute to the unavailability of the service. Since the probability of occurrence of triple failures is significantly less than the probability of occurrence of dual failures (i.e., $4.69 * 10^{-9}$ for triple failures [8]), it is reasonable to neglect the probability of occurrence of triple and more failures in networks. Numerical examples in [8] have shown that the effect of dual span-failures are in fact more important in determining the expected service path unavailability and considering only dual failures is sufficient to obtain a good estimate on the availability of the service.

Therefore, being able to analyze the factors that affect and determine the end-to-end availability is of great interest not only academically but also for network operators as it assists them in offering their customers some guarantees for high service availability. These kind of guarantees are commonly offered as service level agreements (SLA) which has to be honored by the network operator. That is because that a failure to provide service as specified per the SLA may force the network operators to pay certain penalties [1]. Accordingly, it is vital for the network operators to have quantitative measures of end-to-end availability and different availability-aware network designs in order to offer competitive SLAs.

1.3 Thesis contribution

In this thesis, we study service availability and availability aware provisioning principles in optical networks that employ efficient preplanned protection schemes; namely p-cycles [9] and its newer generation of failure-independent path-protection p-cycles (FIPP) [10]. Failure-Independent Path-Protecting (FIPP) p-cycles are an extension of the basic p-cycles and an alternative approach for providing fully pre-connected protection paths with end-to-end failure-independent path protection property [10]. We revisit the problem of availability-aware provisioning in p-Cycle based networks, highlight some of the flaws in the prior models [11] [1] and present a more accurate method for solving the problem.

The new model exhibits some non-linearities in the unavailability constraints which are not desired in Integer Programming problems. In order to resolve these non-linearities, new variables have to be introduced which impose new constraints and eventually result in a huge ILP formulation. Consequently, some simplifications are required to obtain near optimal solution in a reasonable amount of time. Then, we address some heuristics for restricting the number of variables and constraints in the resultant ILP formulation. The bottom line of our work is providing an exact model for availability-aware design of p-cycle based networks.

In addition, we study the unavailability of end-to-end traffic in p-cycle and FIPP based mesh networks, which are designed to protect against single failures and present an availability-aware network design method. Our design method aims at allocating cycles such that the end-to-end unavailability of the protected demands is bounded by an upper limit and the upper limit can be varied as desired.

Allocating cycles to different working paths can be done as follows: The pre-processing program finds candidates or eligible cycles using a depth-first search and pre-selected by hop count lengths, then identify a subset of spans for p-cycle method or a subset of working paths for FIPP method between end nodes that are on the cycle in which they satisfy availability constraints. This process is done before hand and the pre-selected cycles can be utilized for both p-cycle and FIPP protection techniques.

Our study also focused on determining whether FIPP maintains its resource efficiency advantages over span p-cycles when the network design is based on limiting the unavailability. Our results first showed that the length of the FIPP cycle plays a vital role in determining the availability of the working path(s). Similar to span p-cycles, higher service working path(s) availability is obtained when the FIPP cycles contain fewer hops. Results also indicated the important role of the number of demands protected by the same FIPP. We noticed that the higher the desired availability is, the less efficient FIPP method becomes. This is due to the fact that in order to achieve higher service availability, the design will limit the number of demands sharing the same FIPP cycle.

Accordingly, we affirmed that when the network design limits the service unavailability, FIPP tends to be less efficient and its redundancy is 8-13% more than basic

p-cycle. Additionally, we observed that when we do not limit the unavailability, the average availability for span p-cycle tends to be more than the FIPP p-cycle method. In this thesis, we presented our analysis and discussions on these findings.

1.4 Thesis Outline

The rest of this thesis is organized as follows: Chapter 2 focuses on reviewing the most commonly existing strategies for providing survivability and various optimization techniques utilized for the design of WDM mesh networks. An exact model for availability-aware provisioning method is provided in Chapter 3. In Chapter 4, we analyze the availability of mesh networks using FIPP p-cycle based designs and propose a novel method (availability-aware provisioning) for allocating FIPP cycles to restore single link failures such that the availability of all demands in the network is bounded by a lower limit. Chapter 5 concludes and provides direction for future work.

Chapter 2

Background and Related Work

In general, survivability refers to the ability of a network to recover from failures and is considered as one of the most basic requirements of modern telecommunication systems. In real telecommunication networks, frequent occurrence of failures and their corresponding cost give enough motivation to work on improving the survivability of these networks. The impact of network outages can be normally measured in terms of customer-minutes, defined as the outage in minutes multiplied by the number of affected customers [4].

Two basic types of network failures are considered (in the study of survivability): link and node failures. The former is usually caused by cable cuts and the latter is caused by equipment failures in different nodes. Another less considered type of failure in optical networks based on wavelength division multiplexing (WDM) technology is channel failure, which is usually caused by the failure of transmitting and/or receiving equipment (e.g., lasers, photodiodes, etc.) operating on that channel [12].

In this chapter, we present an overview of the various strategies currently proposed for managing survivability in telecommunication networks and in particular, networks based on WDM technologies. In the following sections, also we describe survivable designs and various types of optimization techniques used in optical networks.

2.1 Classification of Survivability Schemes

A general classification of frequently used survivability techniques is presented in [13]. These techniques are generally classified into two main large categories: namely, protection and restoration [14] [2].

Protection mechanism provides a pre-planned scheme where the details of allocation and reservation of the back-up resources are known in advance. In this scheme, some resources are reserved for recovery from failures at either connection setup or network design time. Since reserved resources are kept idle when there is no failure, protection mechanisms are not efficient in terms of capacity; nevertheless, it can provide a quick and hundred percent failure recovery on pre-planned protection paths. The two most common protection schemes used in optical networks are Automatic Protection Switching (APS) and Self-Healing Ring (SHR). The latter is more flexible than APS in that it can handle both link and node failures. Also, SHR utilizes high speed add/drop multiplexers (ADM) and its control mechanism is simple which makes it a very attractive way to provide survivability.

On the other hand, restoration is a survivability technique where the details of responses on the occurrence of a failure in the network are not known a priori. Unlike protection scheme, here the spare capacity in the network is dynamically discovered to restore the affected services upon occurrence of a failure. This means that there are no reserved resources for recovery at the time of connection establishment. The back-up or protection paths are determined by using of real time availability of resources when the failure occurs; hence the recovery time is usually longer than protection scheme. Since sufficient spare capacity may not be available at the time of failure, hundred percent service recovery cannot be guaranteed. However, this technique would be more efficient in terms of resource utilization or spare capacity as compared to protection scheme.

Most studies in the field of survivability in WDM optical networks are contributed to protection schemes. Protection against single failure in WDM networks can be broadly classified into two paradigms: path-based protection and link-based protection (deployed in either a shared or a dedicated manner).

In path-based protection [14] [2], after the primary (or working) path is computed, the source node computes a link disjoint secondary (or backup) path before setting up the call. When a link on the primary path fails, the source and destination nodes of the path are informed and the service is switched over to the back-up path. Since all connections on any of the links in the active path should be restored, the back-up must have sufficient bandwidth for restoration. When a set of spare resources is reserved exclusively to a light path in a path protection, the scheme is designated as dedicated path protection. In dedicated path protection, two non-disjoint protection paths (for two different service paths) must use different wavelengths even if their corresponding working paths are disjoint [12]. On the contrary, in shared path protection method, the same wavelength is shared on the common links of two non-disjoint protection paths, given that their corresponding working paths are link-disjoint and hence do not fail simultaneously. The advantage of dedicated path protection is that in some cases it is able to protect multi-link failures; meanwhile, the shared path protection is more efficient in terms of spare capacity.

In the link-based protection scheme [15], a protection path is computed separately for each link in the primary path. On the occurrence of a link failure, the end-nodes of the failed link on the service path redirect the traffic along the pre-determined protection path. Similar to path-based protection, there are dedicated and shared link protection schemes. In dedicated link protection, each working channel on a

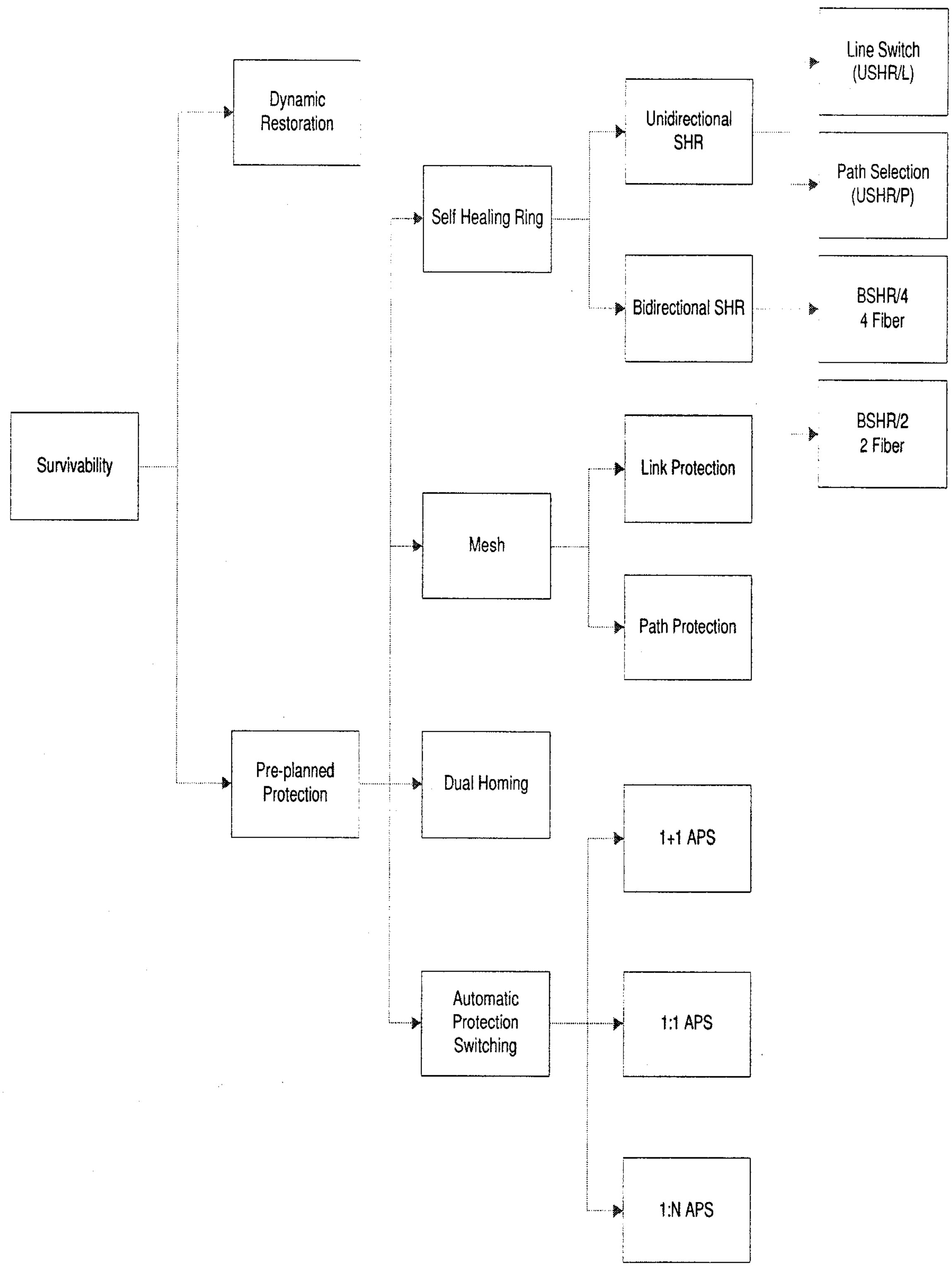


Figure 2.1: General classification of survivability schemes in WDM network

link has its own dedicated protection wavelength path. In this scheme, non-disjoint protection paths for two different links must have different wavelengths on the overlapping portion. Similar to path-based protection, in shared link protection the same wavelength can be used on the common links of two non-disjoint protection paths as long as their corresponding working channels are on different links [12]. Therefore, shared link protection is more capacity efficient than dedicated link protection, and can provide 100 percent recovery from single-link failures.

In general, path protection scheme is more efficient in terms of utilizing resources than link protection scheme, however, it is much slower. This is due to the fact that path protection scheme requires a larger time in signaling and allocating back-up resources. In addition, when cross connect configuration time is small (10 ms), protection switching time is lowest for shared link protection and highest for shared path protection. Whereas, when cross-connect configuration time is large (500 ms), dedicated path protection has the lowest and shared path protection has the highest protection switching times [16].

A general classification of the survivability schemes in the WDM optical networks is shown in Figure 2.1. In the following sections of this chapter, the most important pre-planned protection techniques, ring and mesh architectures, will be described in details.

2.1.1 Ring-Based Survivability

Ring-based protection schemes are the most basic survivability techniques for optical networks. These schemes utilize two fibers to propagate the signals in opposite directions. Path protection is designed using both fibers to set up two lightpaths around the ring. The source node splits the signal in two identical copies transmitting them on the two different lightpaths at the same time. The receiver node selects the signal with the best quality. This scheme is defined as 1+1 dedicated protection and the architecture is known as WDM self-healing ring-based protection. In self-healing ring, when a failure occurs, no optical switching device has to be reconfigured and no signaling is necessary. Hence, recovery is very fast.

Self-healing rings can be categorized into two general types: bi-directional line switched rings (BLSRs) (shown in Figure 2.2) and unidirectional path-switched rings (UPSRs) (shown in Figure 2.3). Similarly, their optical version are optical path protection ring (OPPR) and optical shared protection ring (OSPR).

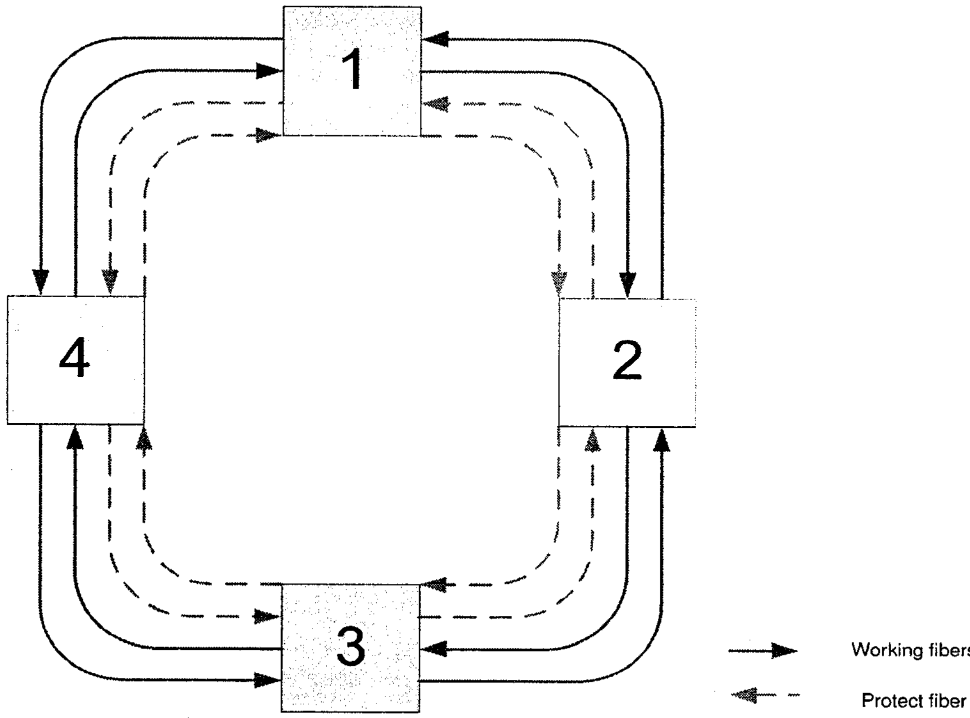
In general, BLSR and UPSR methods are generalizations of 1:1 and 1+1 APS respectively [12]. In the BLSR, nodes adjacent to a span failure test the status of the protection channel. If it was free, the transmitted signal will be switched to the protection channel in the reverse direction from the failure. Each of these nodes receives a replacement signal copy for their receiver in the reverse direction on the protection channel. BLSR can be used more efficiently than UPSR, because any two nodes can make similar use of the shared standby capacity around a BLSR [12].

2.1.2 Mesh-Based Survivability

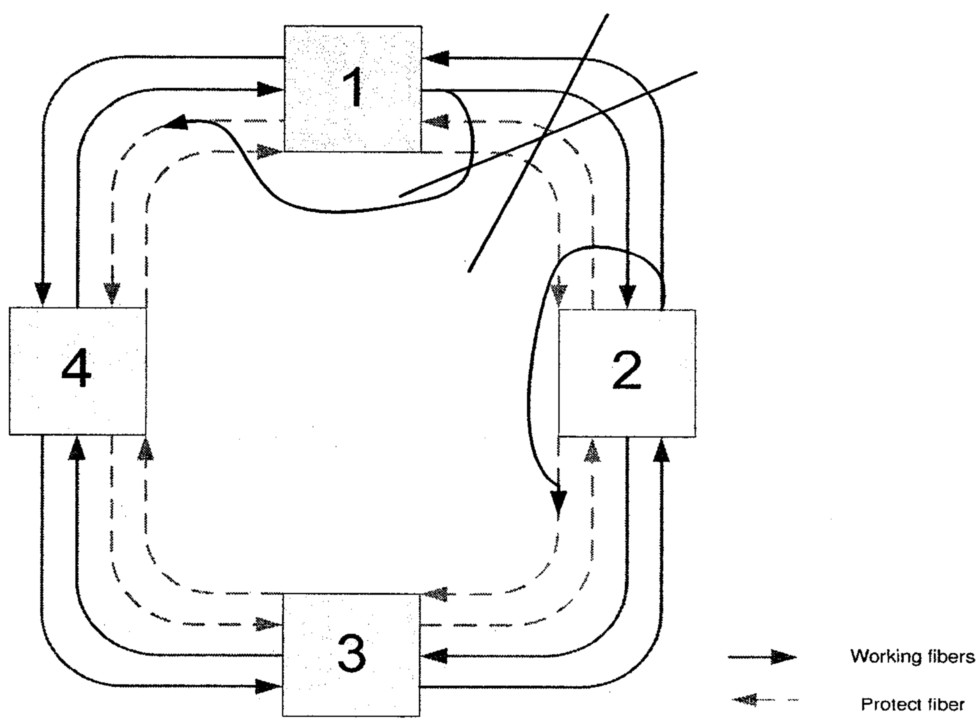
Although the ring structure is the most common physical topology nowadays, WDM mesh networks are obtaining considerable interest. In a mesh network, survivability is more complex problem than in a ring topology because of the greater number of routing and design decisions that have to be considered [17].

Mesh-based survivability schemes can be viewed as the extension and automation of 1:N or 1+1 APS [12]. In mesh-based networks, for the same investment in capacity, more demands can be served in more diverse patterns as compared to a corresponding set of rings. In addition, mesh is less costly in long distance networks where bandwidth, size, and geographically diverse path connectivity are highly demanded [2].

A challenging issue is to decide between these two schemes; i.e, ring-based or

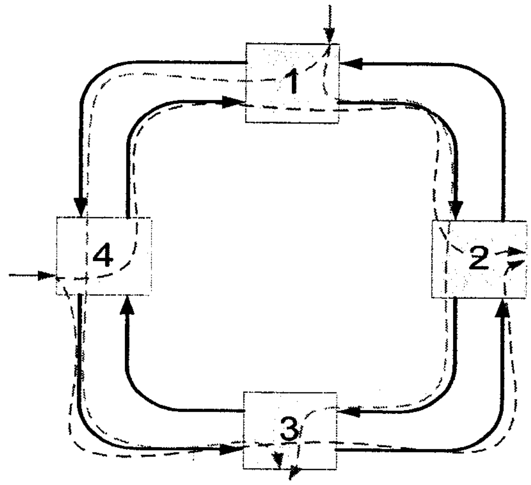


(a) 4-BLSR

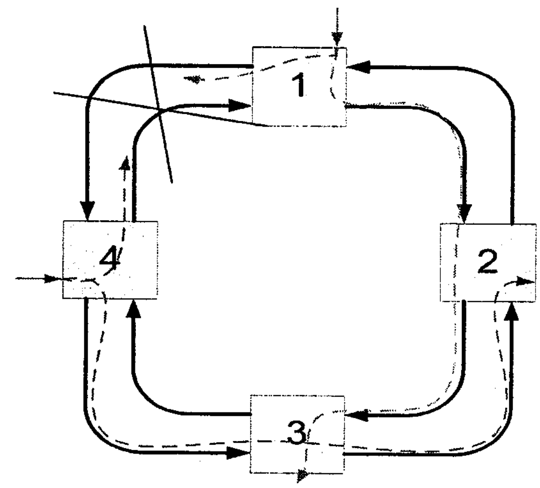


(b) 4-BLSR-Failure

Figure 2.2: BLSR



(a) UPSR



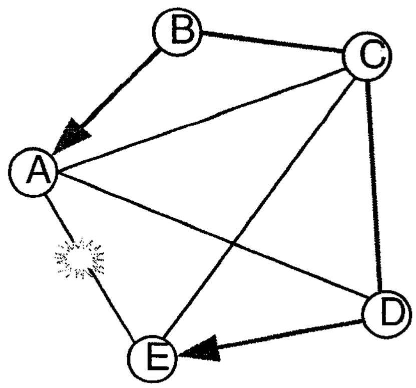
(b) UPSR path recovery

Figure 2.3: Unidirectional Path Switched Rings

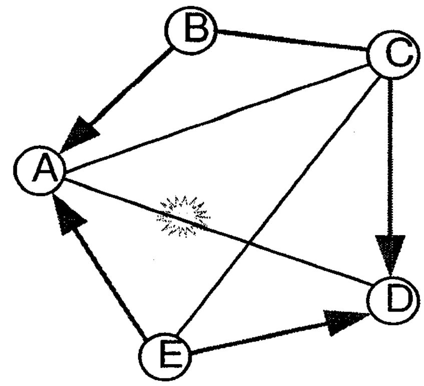
mesh-based survivability. The main advantage of rings is its low cost and high speed when compared to mesh-restoration schemes that required a sophisticated central system with a separate signaling network [12]. This is the reason that despite the need of over 100% redundancy, rings are still preferred in metropolitan networks. On the opposite, mesh offers greater flexibility, efficiency, and support for multiple service classes. Mesh-based networks are also able to organize survivability in response to time-varying patterns of demand [12]. Hence, it is highly desirable to develop a method that is as fast as rings but have the flexibility and capacity efficiency of mesh.

2.2 The Concept of p-Cycle

Pre-configured or pre-connected protection cycles (p-cycles) method proposed by W. Grover's research group [18] [19][9], is a fully pre connected cyclic network architecture which can achieve ring-like high speed protection with mesh-like efficiency in terms of spare capacity. This method utilizes the ring protection function (originally



(a) On-cycle failure



(b) Straddling failure

Figure 2.4: p-cycle example (p-cycle: A-B, B-C, C-D, D-E and EA)

used in SDH/ SONET rings) to perform fast protection and meanwhile it can protect not only the on-cycle spans but also the straddling spans in the network. A straddling span is one that has its end nodes on the cycle, but it is not part of the cycle. In addition, a p-cycle provides one back-up path for restoration of a failed on-cycle span and only a switching action at both end nodes of the failed span is needed to restore the affected path(s). Hence, this method is similar to bi-directional line switched rings (BLSR).

There are different types of p-cycles, however, we can divide them into two main categories: Link p-cycles and node encircling p-cycles [20]. Link p-cycles protect the individual channels within a link and node encircling p-cycles are routed through all neighbor nodes of a specific node and protect all the connections traversing through that node.

2.2.1 The mechanism of p-Cycle

Figure 2.4 shows the basic operation of p-cycles which may be used for restoration in WDM networks that are not possible with rings. It can be seen that for an on-cycle failure, the end nodes switch the traffic to the other side of the cycle (pre-planned path). As shown in Figure 2.4(a), a failure happens on a span on the cycle ($A - E$)

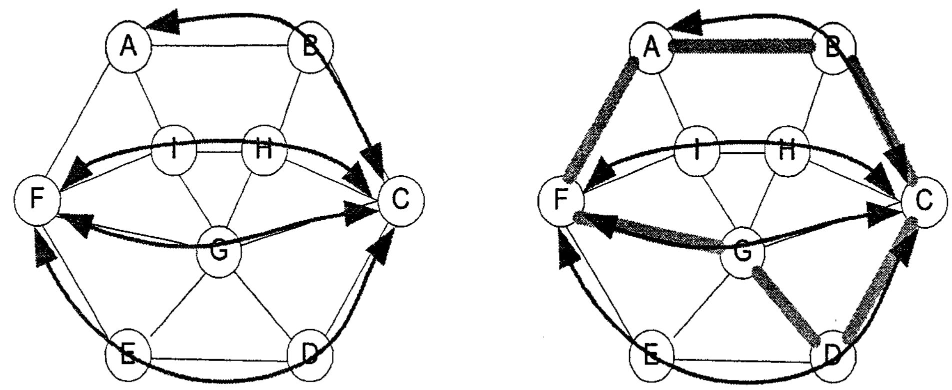
and the surviving arcs of the cycle is used for restoration ($A-B-C-D-E$) using similar method to the ring based BLSR protection mechanism. When a straddling span fails, there are two alternative protection paths on the cycle. Figure 2.4(b) illustrates how the p-cycle can also be utilized for restoration of spans that are not on the cycle. In fact, two restoration paths ($A-B-C-D$ and $D-E-A$) are possible for straddling span $A-D$. Straddling spans make a significant difference to the failure coverage provided by the same investment in spare capacity in a ring compared to a p-cycle. Straddling spans have twice capability of an on-cycle span in terms of efficiency, because when they fail, the p-cycle itself remains intact; hence it can provide two protection paths for each unit of protection capacity.

p-cycles were introduced as a system layer method where they use modular-capacity nodal elements similar to an ADM and implemented at the whole fiber or waveband level. Since the concept of p-cycle separates the routing of working path from configuration of protection schemes, p-cycle based protection technique is also considered as logical layer of implementations.

2.2.2 Failure-Independent Path-Protecting (FIPP) p-Cycle

The failure-independent path-protecting (FIPP) method can achieve p-cycle properties but yields improvements in the usage of network capacity. p-cycle has ring-like high speed protection with mesh-like high efficiency in the use of spare capacity but it is not failure-independent path-protection such as shared-backup path protection (SBPP). FIPP p-cycles support the same failure-independent end-node-activated switching of SBPP, but with the fully pre-connected protection path property of p-cycle [10]. The important property of FIPP p-cycles is that a FIPP acts as a conventional p-cycle for end-to-end paths between nodes on the cycle.

Figure 2.5 illustrates how a set of mutually disjoint working paths can use a single FIPP p-cycle; note that a single failure will only affect one of these demands



(a) Four mutually disjoint working paths

(b) FIPP p-cycle

Figure 2.5: FIPP p-cycle protecting a set of mutually disjoint working paths

and the FIPP p-cycle provides complete recovery. Any set of working routes that are all mutually disjoint is called a group of "compatible routes". Applying the mutually disjointness constraint to primary paths enables them to share a fully pre-connected protection structure, not individual spare channels that still have to be cross connected to form backup paths.

The FIPP p-cycle used on their end-nodes is shown in Figure 2.5(b). The working path between end nodes A and C in the example in Figure 2.5 is totally on-cycle. For this kind of working path, FIPP can provide only one protection path and it is the rest of the cycle itself. Some of the working paths can straddle the respective FIPP p-cycle, such as the working path between end nodes C and F $(C - H - I - F)$. In the case of totally straddling working paths, a demand with two capacity units can be protected per one unit of FIPP p-cycle capacity. Alternatively, the working path between end nodes C and F $(C - D - E - F)$ in this example is partially on-cycle; hence the cycle is limited to providing a single protection path per one unit working capacity.

The switching logic in any case is trivially realized locally within the end-nodes [21]. Each node-pair chooses a direction on the p-Cycle as the default and the other as secondary, for each working path between them. If an on-cycle path fails, then the rule simply says "switch to default backup route". If a straddling path fails, the end nodes switch the working traffic first to the default, and then if required, to the secondary. The general case, where the demand is partially on-cycle, can be also realized by adding a local rule at every node that says "if the span in which the failed path arrives is the same as the span in the default direction on the p-Cycle for that path, then use the secondary". Since this is local information for node, no special-case signaling is required.

2.3 Survivable Network Design

It is desired to design networks which are capacity-efficient while at the same time they are easy to restore under failures. Hierarchical restoration schemes are popular for designing survivable networks in practice [22]. In the first stage, a capacity-efficient solution is found for the network design problem without survivability concerns. In the second stage, given the working capacities, a minimum cost allocation of spare capacity on the links is determined so that the disrupted flow can be safely rerouted, in case of a link failure. This second stage problem is referred to as the Spare Capacity Assignment Problem (SCA) [23].

Two basic scenarios were developed and tested in [22]. In the first case, the objective is to achieve the highest level of restorability for a given set of existing mesh network spare capacity. The second scenario is the reverse of the first one; that is the minimum set of spare capacity is generated such that hundred percent restorability is ensured.

In both scenarios mentioned above, the working demands are first routed via shortest paths (or any other arbitrary path) and then a corresponding minimum spare capacity allocation (or maximum restorability) problem is solved. This is called the non-joint design problem. Another design principle is to solve the problem jointly which was firstly modeled in [2]. This principle attempts to optimize the choice of working routes in conjunction with the placement of spare capacity to achieve the objective.

The idea of optimal spare capacity design for p-cycle based restorable networks was first formulated in [24], [22] using Integer Linear Programming (ILP). In the following sections, we describe linear programming and optimal spare design for p-cycle-based networks in details.

2.3.1 Linear Programming

Linear programming entails the planning of activities to obtain an optimal result which reaches the specified goal according to the mathematical model among all feasible alternatives [25]. The most common type of application of linear programming involves allocating resources to activities. In addition, any problem with mathematical model fitted in general format of the linear programming model is a linear programming problem.

In general, the linear programming problems or LP problems, for short, is the problem of minimizing (or maximizing) a linear function subject to a finite number of linear constraints [26]. Generally, if $c_1, c_2, ..., c_n$ are real numbers, then the function f of the real variables $x_1, x_2, ..., x_n$ defined by

$$f(x_1, x_2, ..., x_n) = c_1x_1 + c_2x_2 + ... + c_nx_n = \sum_{j=1}^{n} c_jx_j \tag{2.1}$$

is called a linear function. Accordingly, if f is a linear function and b is a real number, the equation

$$f(x_1, x_2, ..., x_n) = b \tag{2.2}$$

is referred to a linear equation. Similarly, the inequalities with linear function f

and real number b

$$f(x_1, x_2, ..., x_n) \leq b \tag{2.3}$$
$$f(x_1, x_2, ..., x_n) \geq b$$

are called linear inequalities. Both linear equations and linear inequalities are entitled as linear constraints. Hence the LP problem will be in the following form:

$$minimize \quad \sum_{j=1}^{n} c_j x_j \tag{2.4}$$

subject to

$$\sum_{j=1}^{n} a_{ij} x_j \geq b_i \qquad (i = 1, 2, ..., m)$$
$$x_j \geq 0 \qquad (j = 1, 2, ..., n)$$

Note that the LP problem in this form is referred as standard form. The linear function which has to be minimized (or maximized) in the LP problem is referred to the objective function of that problem (in the standard formulation). In addition, values for variables $x_1, x_2, ..., x_n$ that satisfy all constraints of the LP problem are considered as different feasible solutions of that problem.

In real world, the decision variables make sense when they have integer values (i.e., number of people, machines and vehicles). Since requiring integer values is the only possibility in which a problem deviates from linear programming formulation, it would be an Integer Programming (IP) or integer linear programming [25]. The mathematical model for integer programming is the same as linear programming model with an additional restriction in which variables must have integer values.

Accordingly, in the case that only some of the variables are necessary to be in the integer format, the model is referred to as Mixed Integer Programming (MIP) [25]. Similarly, the IP problems containing only binary variables are called Binary Integer Programming (BIP) problems or 0 - 1 integer programming problems. Binary variables or 0 - 1 variables are referred to decision variables with just two values, 0 and 1. Hence, such variables would be presented by:

$$x_j = \begin{cases} 1 \\ 0 \end{cases} \tag{2.5}$$

2.3.2 Survivable Network Design Based on *p*-Cycle

The design of survivable networks based on *p*-cycles is called the *p*-cycle design problem which is firstly formulated as a Mixed Integer Programming (MIP) model in [9]. In this model, the working demands are routed in advance and then the protection capacity is minimized. The main problem is that the number of possible *p*-cycles grows exponentially with the number of nodes and spans in the network. In order to achieve the real optimal solution, all candidate cycles must be considered which makes the problem unsolvable in a reasonable amount of time. One alternative is just to consider a limited number of promising cycles and find the optimal solution with the restricted possibilities. However, the optimal solution of the main problem is no longer guaranteed [12].

As mentioned in [27], the design of a min-cost set of *p*-cycles to protect a given set of working flows is an NP-hard problem. Therefore, heuristic methods are required to deal with this problem. One heuristic is to develop some criteria for preselecting the most promising eligible cycles, when we design a large-scale network, so that not all the eligible cycles are required to be considered. Some methods have been proposed in the literature for eligible cycle pre-selection. In [27] a parameter called scoring credit was proposed to sort all the possible eligible cycles based on their individual credits in a descending order. For a design with a limited number of eligible cycles, the cycles would be selected starting from the head of the sequence. It was found that the proposed strategy could achieve a much better performance than the random selection.

In addition, in [27] another two different measures of the p-cycles efficiency for p-cycle pre-selection are suggested: A Priori p-cycle Efficiency AE(r) and Demand-weighted p-cycle Efficiency EW(r). The efficiency measure AE(r) counts the number of protected links, divided by the cost of the p-cycle. In EW(r) the offered protection capacity is weighted with the working capacity which needs to be protected for each link. Hence, this measure assumes that the demands are already routed. The optimal p-cycle according to the AE(r) measure is the p-cycle with the lowest average cost for link protection. The main problem is that it does not take into account the actual need for protection, i.e., the working capacity which needs to be protected. Hence, another heuristic for generating cycles dynamically is introduced in [28].

In [28], the authors introduced an approach using Column Generation method to produce cycles dynamically. The idea of column generation is to only generate the variables when needed. Initially, the column generation algorithm is started with a set of dummy cycles. A dummy cycle is a cycle which has the ability of protecting just one link for p-cycle protection techniques or one working path for FIPP techniques. Hence, this dummy cycle is so costly that it will never show up in the optimal solution. Then, in each iteration of this algorithm, we try to improve the cycles according to maximum unavailability constraints, hence this process continues until no improving cycles are found. The column generation method can be used for both protection techniques. In the following section, we will describe column generation in details.

2.3.3 Column Generation (Exact Method)

Column generation techniques [26] [29] propose exact solution methods for linear programming problems. The combination of Column Generation and "Branch-and-Bound" methods which is called "Branch-and-Price" gives the solution for Integer Linear Problems. Column Generation can deal with problems with a very large number of variables by implicitly expressing the constraints. [28]. Column generation methods are based on decomposition of initial linear problem into two sub-problems: Master problem and Pricing problem. The general idea for the decomposition of the initial problem is to deal with smaller number of variables and to independently address and solve the sub-problems.

Similar to the simplex algorithm, the column generation method is an iterative procedure. At each iteration of column generation, the process tries to find a reduced cost for current solution after solving the pricing problem. If the reduced cost is negative, it shows an improvement of the value of master objective function; hence the column generation algorithm adds one or more columns to the constraint matrix of the master problem to gain an improvement in the objective function of the master problem. On the other hand, if no solution of the pricing problem has negative cost, then the algorithm terminates and the current solution is optimal.

Master Problem

The master problem corresponds to a linear program subject to some explicit constraints and some implicit constraints with the properties of the coefficients of constraint matrix [28]. The master problem is formally denoted as the following linear program [29]:

$$z^* := \quad min \quad \sum_{j \in J} c_j x_j \tag{2.6}$$

subject to

$$\sum_{j \in J} a_j x_j \geq b \tag{2.7}$$

$$x_j \geq 0, j \in J$$

where x_j are decision variables, and z^* is objective function in minimization problem. As mentioned above, in each iteration of simplex method, the column generation technique looks for the vector $u \geq 0$ of dual variables as follows:

$$argmin\bar{c}_j := c_j - u^T a_j | j \in J \tag{2.8}$$

where c_j are corresponding cost, and u^T are dual variables corresponding to constraint 2.8. Since an explicit search of J may be computationally impractical, a restricted master problem (RMP) [29] works with a reasonable subset of columns ($J' \subseteq J$). If there is a feasible solution, then master problem computes the reduced cost $\bar{c}^*$ by:

$$\bar{c}^* := min\{c(a) - \bar{u}^T a | a \in A\} \tag{2.9}$$

and returns an answer to the pricing problem. If there is no solution with negative reduced cost in pricing problem (assuming minimization), then the current solution is indeed optimal. Otherwise, one or more columns with negative reduced cost derived from the solution of pricing problem will be added to RMP. Hence, re-optimization of RMP will be repeated to make an improvement in the objective function of RMP.

Pricing Problem

The pricing problem is defined by optimization of the reduced cost subject to the set of implicit constraints, in other words, pricing problem is identifying and generating the most promising columns for the master problem [28]. Selecting columns could be based on different criteria. According to the classical Dantzig rule [30], one chooses among all columns the one with the most negative reduced cost. Various approaches are proposed in [26] like full and partial pricing. In full pricing approach, all columns with negative reduced cost will be selected and in partial pricing approach only a subset of columns with most negative reduced cost is chosen at each iteration to generate new columns.

This technique was first put to actual use by Gilmore and Gomory (1961, 1963) as part of an efficient heuristic algorithm for solving the Cutting-Stock Problem (CSP) [29]. Various Materials (paper, textiles, cellophane, and metallic foil) are made up rolls of large widths. These rolls are referred as raws and they can be cut into rolls of small widths named as finals [26]. The CSP deals with finding the most economical way of cutting the raws into the desired finals when a complicated summary of orders has to be filled.

2.4 On Network Availability

Most of failure scenarios are considered as single failures which means single fiber cuts or cuts of single edges in WDM networks. In almost all survivability mechanisms, the main goal is to replace all the affected service paths after the occurrence of a single network failure.

Fully restorable networks to single cuts are often called "100% restorable". However, multiple failure combinations can cause such networks unable to fully recover. In other words, making a network fully restorable to single failures is not a guarantee that it has 100% availability for every service path. Hence, the availability analysis is important to help obtaining a comprehensive knowledge on the affect of multiple failures in a network.

Some papers have been published on the question of determining and managing the effects of dual span-failures on networks protected by p-cycles [31], [32]. These studies were conducted from the point of view of the restorability to dual-failures. The analytical expression for the availability of paths in a p-cycle-based network was firstly introduced in [11]. The model presented is based on the calculation of the unavailability caused by dual-failures.

2.4.1 On Availability in General

Availability is defined as the probability of the system being found in the operating state at some time t in the future given that the system started in the operating state at time $t = 0$. For repairable systems, an equilibrium is reached between the failure arrival processes and the repair processes, both characterized by the respective rates and resulting in the fraction of the total time that is up time [2]. The most widely used expression for availability in repairable systems is given [2], [33] by:

$$A = \frac{MTTF}{MTTF + MTTR} \tag{2.10}$$

where MTTF (Mean Time to Failure) is the mean time till the failure occurs and MTTR is the Mean Time to Repair. The probabilistic complement of the availability is the unavailability U, where:

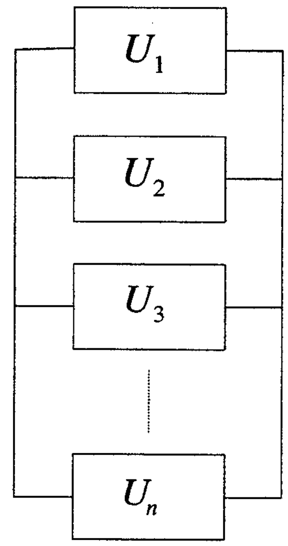


Figure 2.6: N elements in parallel structure

$$U = 1 - A. \tag{2.11}$$

Note that for the rest of the thesis, we use this definition for the unavailability of the system. A system with N elements in parallel (as shown in Figure 2.6) has the unavailability equivalent the product of the unavailabilities of the individual elements; that is all the system would be unavailable if all the elements in the system are unavailable [2], [33].

$$U_p = U_1 U_2 U_3 ... U_n = \prod_{i=1}^{n} U_i. \tag{2.12}$$

Similarly, a system with N elements in a series structure (as shown in Figure 2.7) has the availability equivalent to the product of the availability of the individual elements; that is all the elements in the system needs to be available for the system to be available [2], [33].

$$A_s = A_1 A_2 A_3 ... A_n = \prod_{i=1}^{n} A_i \tag{2.13}$$

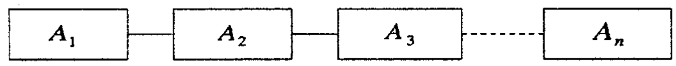


Figure 2.7: N elements in series structure

$$U_s = 1 - A_s = 1 - \prod_{i=1}^{n} A_i = 1 - \prod_{i=1}^{n}(1 - U_i) \tag{2.14}$$

If we assume that $U_i \ll 1$, we can ignore the high power of U_i, so the equation (2.14) will be:

$$U_s \approx \sum_{i=1}^{n} U_i. \tag{2.15}$$

Accordingly, the total unavailability of systems with elements in parallel is obtained by multiplying the unavailability of each element, and the unavailability of systems with elements in series equals to add up unavailability of each element, which is an approximation assuming that $U_i \ll 1$.

2.4.2 An Overall Methodology for Availability Analysis

In telecommunication networks, there is a well established framework for approximate availability analysis, based on the following overall method and built in assumption [34].

Method:

- *First step:* Obtain a good estimation of unavailability of all components or subsystems involved. This may involve summation of devices' failure rates in long time to achieve the failure rate of the system.
- *Second Step:* Understand and draw out all the parallel and series availability relationships among elements. It should be done based on a functional understanding of what is required for the system to perform its function.

- *Third Step:* For systems with elements in series, the unavailability of each element needs to be added up, and for systems with elements in parallel the unavailability of each element is multiplied to obtain the total unavailability.

- *Forth Step:* Add the unavailability contributions from any known correlated multi-element failure scenarios to be considered [34]. Hence, the final system availability is: $1 - U_{tot}$.

Assumptions:

We analyze the availability of a system (which could be a component, path, connection, etc.) in a mesh network with the following typical assumptions:

- The two-state "working - failed" model describes the status of all elements. In other words, a system is either available (functional) or unavailable (experiencing failure) [35].

- All elements (different network components) fail independently.

- For any component, the "up" times (or Mean Time To Failure, MTTF) and the repair times (or Mean Time To Repair, MTTR) are independent memoryless processes with known mean values [35].

These assumptions are accepted as sufficient for practical availability analysis of network services. The availability of a system is the fraction of time the system is "up" during the entire service time. If a connection is carried by a single path, its availability is equal to the path availability; if this connection is dedicated or shared protected, its availability will be determined by both primary and backup paths [35].

Here, the contribution of the reconfiguration time for switching traffic from the primary path to the backup path (including signal propagation delay of control signals, processing time of control messages, and switching time at each node) towards unavailability is disregarded since it is relatively small, usually of the order of a few tens of milliseconds, compared to the failure-repair time (of the order of hours) and the connection's holding time (of the order of weeks or months) [35].

2.4.3 Availability Analysis of a Path in p-Cycle-Based Networks

The most common and practical approach for finding service availability in a network is "cutsets method" [34]. In this method, failures that cause service outage are divided into categories. Then the probabilities of unavailability in different categories are added in order to obtain an estimate of the average service unavailability.

To develop the equations for path availability in a p-cycle based network, the path is divided into "protection domains". A path may cross several protection domains between its origin and destination nodes. A path is said to cross a protection domain associated to a p-cycle, if at some point that path is protected by that p-cycle.

Two different definitions are given for a protection domain in [11] and [1]. In [11], if a span on a path was protected by a p-cycle as an on-cycle span and another span on the same path was protected as a straddling span, then these two spans were counted as two different domains. However, in [1], a "protection domain" is defined as the set of spans which are protected by the same p-cycle. In other words, all spans in a path protected by the same p-cycle belong to the same protection domain and hence the two spans in the above case also belong to the same domain.

Since the protection domains of a path are in series, the unavailabilities of a path in a p-cycle protected network can be expressed as the sum of the unavailability of the path in the different protection domains crossed. Therefore the unavailability of each section of the path, which belongs to the same protection domain, must be analyzed.

In [11] and [1], all possible combinations of dual failures within the protection domain that can result in an outage on the corresponding service path are derived.

Considering all combinations for on-cycle and straddling spans, we can have 12 different dual failure scenarios [12]. Based on both definitions of protection path in [11] and [1], six of these cases for dual span-failure sequences can result in service path outage. Two of these scenarios are for on-cycle path, and the rest are for straddling path in a given protection domain. Each of these sequences is independent from the others, i.e., with respect to a given path a dual span-failure can only belong to one of these sequences. The physical unavailability of each span is assumed to be the same.

After deriving the equation for the overall unavailability of a path, different factors can be investigated in order to compare the unavailability of a path in a given protection domain. It would depend on whether the path is an on-cycle or straddling for the cycle associated to that domain. For both cases the unavailability of the path in the protection domain is directly proportional to the number of spans on which the paths crosses in the domain [12].

Results in [11] show that the unavailability of both on-cycle and straddling spans is proportional to the number of on-cycle spans. It is also shown that the unavailability of the straddling path is 25% lower than that of the on-cycle path for all values of the number of on-cycle spans. The reason of this difference is that on-cycle spans have longer protection paths. Therefore, on average they will be more vulnerable to a second failure in their protection path compared to straddling spans [11]. As the number of on-cycle spans increases, the average length of a backup path for on-cycle spans increases twice as fast as that of straddling spans. Hence, the unavailability of on-cycle spans will be higher.

According to these results, it can be inferred that the size of p-cycles is very important in determining the availability of service paths that traverse their domains. Smaller p-cycles will allow much higher availability to be offered to paths [11]. However, as we know smaller p-cycles are generally less capacity efficient than larger ones. Therefore, based on the size of candidate p-cycles there is a trade-off between capacity efficiency and availability. It would not be a good strategy to restrict the number of protected straddling spans in order to improve service availability, because the essential reason of p-cycle efficiency is its ability to protect straddling spans. Limiting the number of straddling spans gives a slight improvement for the path availability but highly decreases the capacity efficiency.

2.4.4 Availability-Aware Design in p-Cycle-Based Networks

Former studies show that the amount of spare capacity required to protect all demands against any dual span-failure is typically in the order of three times the amount required to protect against single failures, and the total capacity cost for the whole network would often increase by more than 50% [36]. Obviously, this is too costly for most network operators. Therefore, in [11] an alternative approach is presented, which consists of improving the availability of only selected service paths instead of trying to improve the availability of all paths.

As mentioned before, the availability of paths on straddling spans is higher than that of paths on on-cycle spans. Therefore, to achieve higher availability in p-cycle networks, one way is to design the relationship between cycles and paths so that priority services only pass through straddling spans or put the protection capacity only on straddling spans [11].

A more practical and economical strategy is to improve only the restorability of selected"priority" service paths. Two new models for p-cycle networks are introduced for joint routing of path and spare capacity taking the priority of service paths into account [12]. The service paths are divided into three classes:

- Gold class includes the normal paths which are allowed to be routed either on on-cycle spans or on straddling spans

- Gold-plus class where the demands are restricted to travel merely on straddling spans
- Platinum class which restricts the path to straddling spans. In addition, they will be offered two protection options instead of just one.

The first model is called Selectively Enhanced Availability Capacity Placement (SEACP) which jointly optimizes the routing of demands and the allocation of spare capacity to find the minimal cost capacity placement [11]. In addition to the usual routing constraint to guarantee the protection of working capacity against single link failures, SEACP takes the Gold-plus demands into account and guarantee that selected priority paths will be routed exclusively on straddling spans, therefore enjoying an availability improvement, whereas other paths are routed either on straddling spans or on on-cycle spans. Higher availability is achieved by using a different routing strategy for service paths of Gold-plus class.

The second strategy is called multi restorability capacity placement (p-cycle MRCP) [11]. In this strategy, demands of the Platinum class are considered which are offered a much higher availability improvement. This model is logically protected against dual failures instead of just single failures. MRCP offers two protection options to selected priority paths by routing them on straddling spans and allowing them to access either sides of the cycle they straddle [12]. Results show that in MRCP, capacity requirement increases rapidly by increasing the priority demands. However, the availability of priority paths with MRCP is expected to be very much higher than that of priority paths with SEACP.

In Platinum class, since in two restoration paths are offered for each working path, it is expected to enjoy extremely high availability. In fact, the only dual failure scenario which results in an outage for platinum class is when both failed crossed spans are straddling and both have a Platinum-class working link protected by the same p-cycle [12]. In this case, only one of the two Platinum-class working links can be survived.

The availability of a service path in any of these classes is influenced by many factors such as the statistics of network element failures, the statistics of repair times, mean restoration time, etc. Numerical examples in [8] have shown that for the determination of expected service path availability in long-haul networks, the effects of dual span-failures are in fact much more important than other failure scenarios, and considering dual span-failures only is sufficient to obtain a good estimate of the expected availability of service. This makes the dual failures as the main factor which can contribute to the unavailability of the service. Since the probability of occurrence of triple failures is significantly less than the probability of occurrence of dual failures, it is reasonable to neglect the probability of occurrence of triple and more failures in networks. Moreover, in networks where pre-connected protection schemes have been used to achieve restorability against single failures, the recovery speed is expected to be very fast. In such networks the unavailability of service during the recovery time is also insignificant. Hence, availability-aware network design based on analysis of dual-failures has emerged as an important topic for achieving some level of robustness against multiple failures.

Availability-aware network design is based on the calculation of the unavailability caused by dual-failures for end-to-end working paths. This method applies cut-sets method which is the most practical approach for finding service availability in a network [2]. In the availability-aware design method, failures that cause service outage are divided into categories. Then, the probability of unavailability in different categories is added in order to obtain an estimate of the average service unavailability. Accordingly, the availability-aware provisioning method ensures that the average service unavailability of all working paths is less than the desired upper bound.

Few papers have been published on the question of determining and managing the effects of dual span-failures on networks protected by p-cycles and FIPP [31][32]. These studies were focused on the restorability of dual-failures. The analytical expression for the availability of paths in a p-cycle-based network was introduced in [11][1] and for FIPP-based network was developed in [37]. The models presented in both is based on the calculation of the unavailability caused by dual-failures.

2.5 Related Work

A number of papers have been published on the p-cycle scheme, mostly related to the issue of optimizing the spare capacity placement required to support restoration of the network against single failure [18], [19], [9], [20] and [38]. Also, more recently, some papers have been published on an extension of p-cycle concepts to achieve the property of full pre-connection paths, while adding property of end-to-end failure-independent path-protection (FIPP) against span or node failures [10], [39]. In these two papers, it is shown that FIPP p-cycles can have the conventional property of p-cycles with having the end-to-end failure-independent path protection property of SBPP.

Mathematical models for path availability and provisioning resources required in various strategies for realizing high availability service paths in bidirectional line switched rings and shared protection WDM rings has been developed in [40]. More related works on availability have been done by Clouqueur and Gorver [8], [11], [41] and [42].

Determining and restoring dual span-failures on networks protected by p-cycles have been studied recently [31], [36] and [32]. In [31], it is presented that the dual failure restorability and the protection capacity can vary significantly for cycle configurations with different numbers of developed p-cycles. Clouqueur and Grover explain the relation between the path availability and the restorability of a network to dual failures in [11].

The availability of service paths has been studied for ring-based networks in [40], and for mesh-based survivable networks referred to as span restoration in [8], p-cycle protection in [11] and [1] and also for FIPP technique in [37]. The availability analysis of a span restorable network is based on the computational analysis of the restorability of a network to all possible dual-failure scenarios. In [11] the concept of availability analysis in span restorable network is extended to p-cycles. The paper computes the unavailability of spans regarding all dual failure sequences that result in an outage of a path protected by p-cycles. The authors consider the working path to consist of one or more domains where each domain is one or more spans in the working path, which is protected by the p-cycle and they compute the unavailability of a domain. The authors showed the importance cycle size plays in terms of availability and suggested strategies for achieving high availability of paths in a network protected by p-cycles.

In [1], the authors design a method for allocating p-cycle, such that the end-to-end unavailability is bounded by a desired upper limit. The authors of [1] showed that since the length of the working path along which a demand is routed also determines the unavailability of that demand, a smaller working path could be allocated longer p-cycles to protect its spans and still limit the unavailability of the working path to a desired upper bound. In [37], authors used the same approach to define availability-aware design method for FIPP p-cycle-based networks. The authors of [37] showed that the FIPP protection method requires more network capacity to obtain the same level of availability that basic p-cycle method achieves.

Availability-aware design method is also used in [35] and [43] to provide different level of availability for different customers. The authors of [35] and [43] proposed this method to use connection availability as metric to provide differentiated protection services in WDM mesh network. In addition, in [44], authors proposed the theory to estimate a suitable safety factor for SLA availability guarantee. In [44], the authors showed that their theory can be used by network operators to show customers why they must be more than proportionately conservative when being asked to enter an SLA including an availability guarantee over a notably short term period. Finally, the authors of [45] tried to find out the trade-off between MTTR reduction to shorten outage times and adding additional capacity to increase the dual-failure restorability for preventing outages from occurring.

Chapter 3

Availability-Aware Design in p-Cycle-Based Networks: An Exact Method

3.1 Introduction

In this chapter we will revisit the problem of availability-aware provisioning in p-cycle based networks and highlight some of the flaws in the previous models [11] [1] and present an exact method for solving the problem. The new model exhibits some non-linearities in the unavailability constraints which are not desired in Integer Linear Programming (ILP) problems. In order to resolve these non-linearities, new variables have to be introduced which impose new constraints and eventually result in a huge ILP formulation. Consequently, some simplifications are required to obtain a near optimal solution in a reasonable amount of time. We will address some heuristics for restricting the number of variables and constraints in the resultant ILP formulation. Here, the objective of this work is to provide an exact model for availability-aware design of p-cycle leveraging the analysis presented in [1].

3.1.1 Outline

The rest of this chapter is structured as follows. In Section 3.2 we present a computational analysis of the unavailability of a working path protected by different p-cycle. Section 3.3 presents some critical issues required for obtaining the exact model which have not been considered in previous work. In Section 3.4, we construct our model upon the new insights given in Section 3.3. The practical concerns and some heuristics will be addressed in Section 3.5. The experimental results will be discussed in Section 3.6. Finally, Section 3.7 concludes our findings with a discussion on future direction.

3.2 Unavailability Analysis Background

In prior work, analyzing the unavailability of a set of spans protected by a p-cycle is done by using the concept of protection domains [11], [1] as explained in Chapter 2. In [1], a protection domain is defined as the set of spans which are protected by the same span p-cycle; however, in [11], if a span on a path was protected by a p-cycle as an on-cycle span and another span on the same path was protected as a straddling span, then these two spans were considered as two different domains. In this work, we will analyze the different categories of all dual span-failures that can result in an outage for a given service path in a p-cycle based network using the same concept as in [1]. In [1] all possible combinations of dual failures are found and categorized such that all these combinations are independent from each other and a dual failure can only belong to one of these categories. This analysis assumes that each span has the same physical unavailability (U). According to the protection domain, we have partitioned a p-cycle (p) protecting the span/spans along a path (r) into different sets by using the following notations:

O_r^p: The set of on-cycle spans in p-cycle (p) that are on the working path (r) and protected by p.

$O^p_{\bar{r}}$: The set of on-cycle spans in p-cycle (p) that are not on the path (r) and also those spans traversed by path (r) but not protected by p-cycle (p).

S^p_r: The set of straddling spans in p-cycle (p) that are on the working path (r) and protected by (p).

$S^p_{\bar{r}}$: The set of straddling spans in p-cycle (p) which are not on the working path (r) but protected by (p) and also straddling spans traversed by (r) and not protected by (p), but traversed by another service path which is protected by (p).

We have to note that $O^p_{\bar{r}}$ and $S^p_{\bar{r}}$ are not complement of O^p_r and S^p_r respectively.

We now categorize all dual failure scenarios that may cause service outage in a protection domain into one of the following categories:

Category-1: Dual failure scenario in which one of the failed spans belongs to O^p_r and the other failed span belongs to $O^p_{\bar{r}}$.

Category-2: Dual failure scenario in which one of the failed spans belongs to O^p_r and the other failed span belongs to $S^p_{\bar{r}}$.

Category-3: Dual failure scenario in which one of the failed spans belongs to O^p_r and the other failed span belongs to S^p_r.

Category-4: Dual failure scenario in which one of the failed spans belong to S^p_r and the other failedspan belongs to $O^p_{\bar{r}}$.

Category-5: Dual failure scenario in which both spans belongs to S^p_r.

Category-6: Dual failure scenario in which one of the failed spans belongs to S^p_r and the other failed span belongs to $S^p_{\bar{r}}$.

We will be referring to these categories throughout the rest of the chapter. This analysis will allow us to compute the total unavailability of the service path (path of concern r).

Category-1: The failure sequences involving $s_1 \in O^p_r$ and $s_2 \in O^p_{\bar{r}}$. In such a scenario, the order in which the failures occur is not important, hence both cases are certain to cause a service outage. Obviously, the number of possible combination of dual failures in this category is $|O^p_r| \cdot |O^p_{\bar{r}}|$. In other words, the number of failure sequences involving span s_1 in this category is $|O^p_{\bar{r}}|$. Hence the contribution of dual failures to the unavailability of the protection domain is:

$$U_{Category-1} = |O^p_r| \cdot |O^p_{\bar{r}}| \cdot U^2 \tag{3.1}$$

Category-2: In the second category, one span failure (s_1) belongs to O^p_r and the other span failure (s_2) belongs to $S^p_{\bar{r}}$. Here, the order in which failures occur is important. There will be a service outage if a failure occurs first on span s_2 (straddling span) and assuming that the p-cycle is fully loaded. Since a straddling span may fail first with a probability of 50%, we can denote the unavailability due to a dual failure in this category as:

$$U_{Category-2} = \frac{1}{2}|O^p_r| \cdot |S^p_{\bar{r}}| \cdot U^2 \tag{3.2}$$

Category-3: The dual failure of an on-cycle span ($s_1 \in O^p_r$) with a straddling span ($s_2 \in S^p_r$) on the working path. Irrespective of the order in which the failures occur, a service outage is guaranteed. Hence, the number of combinations of such failures is $|O^p_r| \cdot |S^p_r|$ and the contribution to the total unavailability as a result of dual-failures in this category is:

$$U_{Category-3} = |O^p_r| \cdot |S^p_r| \cdot U^2 \tag{3.3}$$

Category-4: In this category, the order in which the failures occur is important. The probability of the on-cycle span ($s_1 \in S^p_r$) fails first followed by a second failure on the straddling span ($s_2 \in O^p_{\bar{r}}$) is half. This sequence of dual failure will certainly lead to service outage on the working path. On the other hand, the first failure could happen on the straddling span (s_1) and the second failure could occur on the on-cycle span (s_2). This scenario will cause an outage if the failure on the on-cycle span affects the back-up path used by the straddling span. Since there is a 50% chance that the back-up path would be affected, the probability of an outage for a working path in this category is:

$$U_{Category-4} = \frac{3}{4}|S_r^p| \cdot |O_{\bar{r}}^p| \cdot U^2 \tag{3.4}$$

Category-5: In this category, both failures belong to $S_r^p(s_1, s_2 \in S_r^p)$. Here, we have assumed that the p-cycle is fully loaded; hence, the p-cycle is completely utilized when the first failure occurs on the straddling span (s_1). Therefore, the second failure on another straddling span (s_2) will definitely result in an outage for the working path. The number of combinations of dual failures in this category is given by $\frac{1}{2}|S_r^p| \cdot |S_r^p - 1|$; hence the unavailability contribution by this type of dual failure can formally be written as:

$$U_{Category-5} = \frac{1}{2}|S_r^p| \cdot (|S_r^p| - 1) \cdot U^2 \tag{3.5}$$

Category-6: Here, only if the first failure occurs on the straddling span ($s_2 \in S_{\bar{r}}^p$) which is not traversed by the working path (r), a service outage will be guaranteed. In this case, the second failure on the straddling span ($s_1 \in S_r^p$) traversed by the working path will cause an outage. Since, there is a 50% chance to have this order of dual failure, the unavailability contribution due to failures in this category will be:

$$U_{Category-6} = \frac{1}{2}|S_r^p| \cdot |S_{\bar{r}}^p| \cdot U^2 \tag{3.6}$$

Without much elaboration it can be stated that the total unavailability of a working path can be given by the sum of the contribution towards the unavailability by each category. Formally, we have:

$$
\begin{aligned}
U_{domain} = |O_r^p| \cdot |O_{\bar{r}}^p| \cdot U^2 + \frac{1}{2}|O_r^p| \cdot |S_{\bar{r}}^p| \cdot U^2 + |O_r^p| \cdot |S_r^p| \cdot U^2 + \\
\frac{3}{4}|S_r^p| \cdot |O_{\bar{r}}^p| \cdot U^2 + \frac{1}{2}|S_r^p| \cdot (|S_r^p| - 1) \cdot U^2 + \\
\frac{1}{2}|S_r^p| \cdot |S_{\bar{r}}^p| \cdot U^2
\end{aligned}
\tag{3.7}
$$

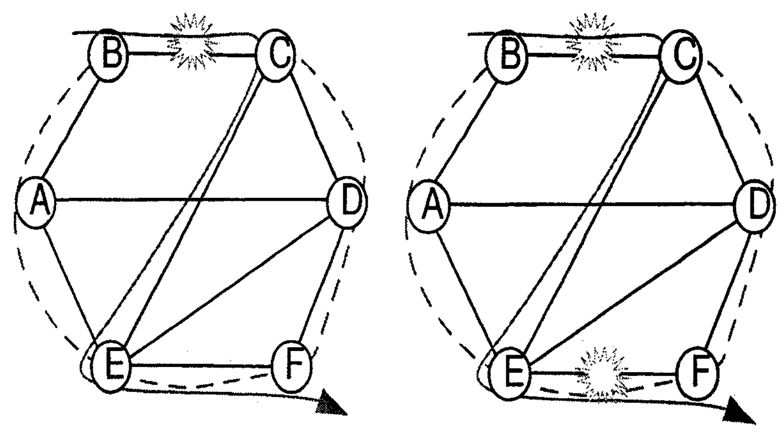
In conclusion, since the protection domains of a path are in series [2], the unavailability of a path in a p-cycle based network can be expressed as the sum of the unavailability of the path in the different crossed protection domains:

$$
U_{path} = \sum_{domains} U_{domain} \tag{3.8}
$$

3.3 Some Insights on the Availability Analysis in p-Cycle-Based Networks

As discussed in the previous section, six dual span-failure scenarios which result to service outage have been considered in [1]. Then, the expression of the overall unavailability of a path is derived based on these six possible scenarios. In this section, we will illustrate two critical issues which have to be added to the model in [1] in order to achieve the exact availability-aware network design.

The first problem with the model in [1] is that the six dual span-failure scenarios do not cover all cases that lead to the loss of traffic in one protection domain. There is one additional dual failure scenario that yields to a service outage and is overlooked in [1]. This scenario consists of dual failures on two on-cycle spans which both belong



(a) first failure (b) second failure

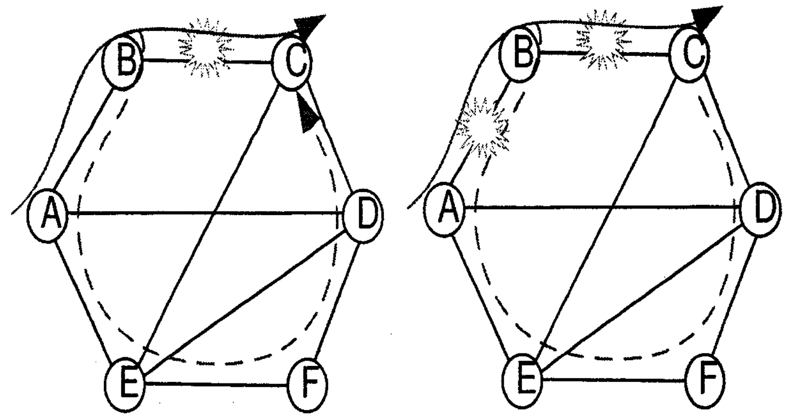
Figure 3.1: New category of dual failure scenarios which result in service outage

to the path of concern (r). This scenario which is hereafter called "category 7", causes an outage regardless of the order and location of the failures. The number of combination of dual failures in this new category is given by $|O_r^p| \cdot (|O_r^p| - 1)$; hence the unavailability contribution by this type of dual failures can formally be written as:

$$U_{Category-7} = \frac{1}{2}|O_r^p| \cdot (|O_r^p| - 1) \cdot U^2 \quad (3.9)$$

In the following, we illustrate this new category with two examples. The example given in Figure 3.1 shows a path of concern consisting of two on-cycle spans and one straddling span all being protected in one protection domain. Upon the occurrence of the first failure on span $B - C$, the end nodes switch the traffic to the back up path $B - A - E - F - D - C$. A second failure on span $E - F$, affects the backup path of the service restored from the previous failure. Meanwhile, the second failure can not be recovered, because the first failure has already affected its backup path. Therefore, this dual failure scenario results in service outage regardless of the order of failure occurrence.

Figure 3.2 shows that a dual failure involving two on-cycle spans from the path



(a) first failure (b) second failure

Figure 3.2: dual failure in two adjacent on-cycle spans

of concern will cause service outage even if the spans are adjacent to each other. The first failure on span BC invokes the end nodes to switch the traffic to the backup path $B - A - F - D - C$. At first glance, it seems that when the second failure takes place on span $A - B$, the connection can be successfully recovered through the backup path $A - E - F - D - C$. However, this is not true, because when the optical cross connect (OXC) switch at node A tries to loop back the traffic, it finds that the backup path is already in use due to the switching after the first failure. The OXC would be able to do this job if the backup path was preempted after the occurrence of the second failure. However, this preemption requires signalling which is not desired in p-cycle networks [2]. In summary, whether the on-cycle spans are adjacent or they belong to a Z-shape path which consists of non-adjacent on-cycle spans, the dual failure leads to a service outage.

In the following, we address another subtle issue about the "protection domain" definition which has to be carefully considered in order to avoid over-estimation of service path unavailability. According to the assumption in [1], the p-cycles are "fully loaded", which means they *provide restoration to two units of working capacity in all straddling spans and one unit of working capacity to all on-cycle spans.* It should be noted that one span can traverse several protection domains. However, the failure in

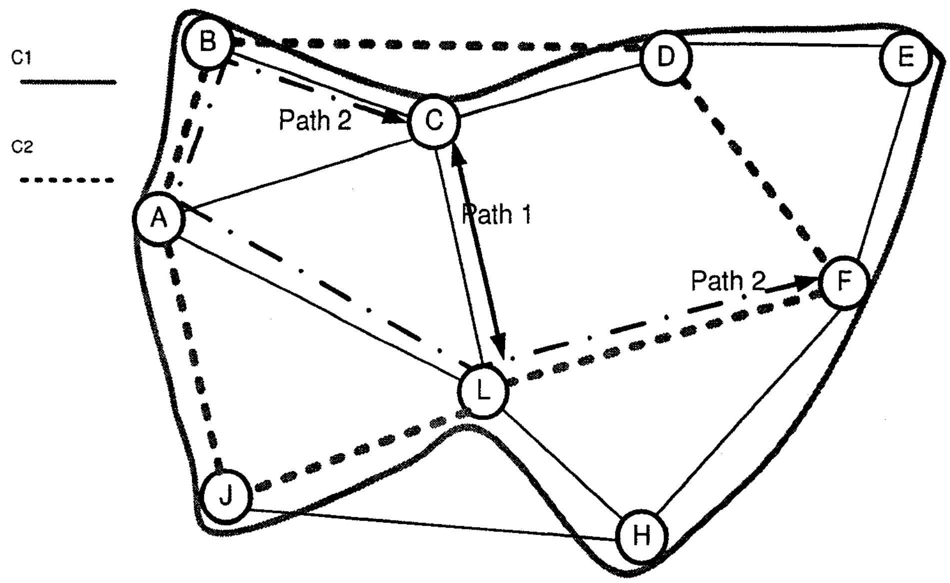


Figure 3.3: Spans traverse two protection domains

a span only affects the availability of its corresponding protection domain. In other words, when dealing with dual failure scenarios only the cycles should be considered which really protect the working channels on a failed span.

This issue is illustrated in Figure 3.3. In this figure, span $A-L$ is fully loaded by two unit working routes, Paths 1, which is protected by C_1 $(A-B-D-F-L-J-A)$. Although, both working channels on the straddling span $A-L$ are protected in the same protection domain, it clearly straddles another cycle C_2 $(A-B-C-D-E-F-H-J-A)$, as shown in the figure. Nevertheless, when deriving the model, it should be noted that a failure on span $A-L$ only affects the availability of protection domain C_1 even though it straddles another cycle C_2.

Similarly, on-cycle span $A-B$ which is protected by p-cycle C_2 should only be considered in protection domain C_2 not C_1, and although span $F-L$ in Path 2 is on-cycle span for C_2, this span is protected as straddling span in p-cycle C_1 and belonged to this protection domain. This subtle issue needs to be considered in the model to avoid over counting, which results in higher service unavailability. Therefore, a variable is needed to distinguish each working channel on a span as well as its corresponding protection domain. Accordingly in Table 3.1, we show the spans belonging to different availability sets in protection domain (C_1) and compare them with [1].

Table 3.1: Comparison between availability sets in current work and prior work[1]

Availability Sets	Current work	Prior work [1]
O_r^p	$B-C$	$A-B, B-C$
$O_{\bar{r}}^p$	$A-B, C-D, D-E, E-F, F-H, H-L, L-J, L-A$	$C-D, D-E, E-F, F-H, H-L, L-J, L-A$
S_r^p	$A-L$	$A-L, F-L$
$S_{\bar{r}}^p$	$C-L$	$A-C, B-D, D-F, H-J, C-L$

3.4 Developing An Exact Model

In this section we propose an exact formulation for availability-aware design based on p-cycles. This formulation optimizes the allocation of spare capacity in order to find the minimal cost capacity placement that allows us not only to guarantee that every working path is protected against single span failure but to also ensure that the availability of any service path is not less than the desired design requirement.

The routing of the demands is done using a standard K-shortest path algorithm and ahead of the placement of the p-cycles. All the working paths are provided as inputs for the ILP model. Thus, the optimization is a sequential non-joint optimization problem.

3.4.1 Notations

For this formulation we use the following notations:

Sets

S = Set of spans.

P = Set of simple cycles eligible for allocation.

R = Set of working paths.

Input Parameters

C_k = Cost of a span k.

$$\delta_k^p = \begin{cases} 2, & \text{if span } k \text{ straddles cycle } p; \\ 1, & \text{if span } k \text{ crosses cycle } p; \\ 0, & \text{otherwise.} \end{cases}$$

l^p =Number of hops in p-cycle (p).

MU = Maximum Unavailability of any working path after the allocation of p-cycles.

Intermediate Parameters

$$N_{r,k}^p = \begin{cases} 1, & \text{if } p\text{-cycle } p \text{ is allocated to protect span } k \text{ on the working path } r; \\ 0, & \text{otherwise.} \end{cases}$$

$$\alpha_k^p = \begin{cases} 1, & \text{If straddling span } k \text{ is protected by the } p\text{-cycle } p; \\ 0, & \text{otherwise.} \end{cases}$$

O_r^p = Number of on-cycle spans in p-cycle (p) that are on the working path (r) and protected by (p).

$O_{\overline{r}}^p$ = Number of on-cycle spans in p-cycle (p) that are not on the path (r) and also those spans traversed by path (r) but not protected by p-cycle (p).

S_r^p = Number of straddling spans in p-cycle (p) that are on the working path (r) and protected by p.

$S_{\bar{r}}^p$ = Number of straddling spans in p-cycle (p) which are not in working path (r) but protected by (p) and also straddling spans traversed by (r) and not protected by (p), but traversed by another service path which is protected by (p).

U_r^p = Unavailability of a protection domain in working path (r) protected by p-cycle (p).

Decision Variables (Output)

N^p = Number of protection units per p-cycle (p).

S_k = Number of spare units placed on span (k).

U_r = Total end-to-end unavailability of working path (r).

3.4.2 ILP Model for Single Failure protection with Limitation of the End-to-End Unavailability

In our model, the **Objective** is to minimize the total spare capacity cost:

$$\min \quad \sum_{k \in S} C_k S_k \tag{3.10}$$

Constraint (3.11) guarantees that exactly one p-cycle will be allocated to protect each span; that is, this constraint finds a p-cycle that has to be placed to ensure that every span (k) traversed by working path (r) is being protected against any single failure.

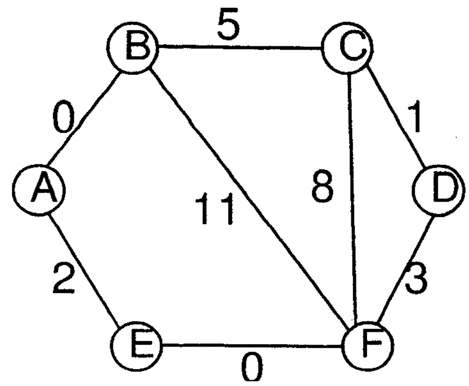


Figure 3.4: A p-cycle and the corresponding number of protected working channels on each span

$$\sum_{p \in P; \delta_k^p > 0} N_{r,k}^p = 1 \qquad r \in R : k \in r \tag{3.11}$$

The number of required copies of each cycle (N^p) is identified by the maximum number of working channels on all spans which are being protected by cycle (p) as a part of some working routes. Recall that each working channel on a span needs one copy of p-cycle if it is on-cycle span, whereas it will need one half of the cycle if it straddles the cycle. Therefore, the number of required copies of each cycle needs to be considered separately for on-cycle and straddling spans as shown in constraint (3.12), where the values of $N_{on-cycle}^p$ and $N_{straddle}^p$ are given in (3.13) and (3.14) respectively.

$$N^p = \max\{N_{on-cycle}^p, N_{straddle}^p\} \qquad p \in P \tag{3.12}$$

$$N_{on-cycle}^p = \max_{k \in S: \delta_k^p = 1} \{ \sum_{r \in R: k \in r} N_{r,k}^p \} \qquad p \in P \tag{3.13}$$

$$N_{straddle}^p = \lceil \frac{1}{2} \max_{k \in S: \delta_k^p = 2} \{ \sum_{r \in R: k \in r} N_{r,k}^p \} \rceil \qquad p \in P \tag{3.14}$$

Figure 3.4 illustrates the way of finding the number of required copies of one cycle

$(A-B-C-D-F-E)$ which is used to protect a set of on-cycle and straddling spans. The number of working channels which are protected by this cycle is given for each span. The on-cycle spans $A-E$, $B-C$, $C-D$ and $D-F$ have 2, 5, 1 and 3 working channels respectively. Spans $A-B$ and $E-F$ do not include any working channels which are protected by the cycle $A-B-C-D-F-E$. Therefore, the maximum number of protected working channels on on-cycles spans is 5 which corresponds to span BC. Hence, 5 copies of the cycle are required to protect all on-cycle spans. Similarly, 5.5 copies of the cycle is required to cover all working channels on straddling spans $B-F$ and $C-F$. Since N^p has to be an integer, the number of copies of the cycle will be 6 in this example.

Combining constraint (3.12) with identities (3.13) and (3.14), we derive the equivalent constraints (3.15) and (3.16) which can be used in the ILP model.

$$N^p \geq \sum_{r \in R: k \in r} N^p_{r,k} \qquad p \in P, k \in S : \delta^p_k = 1 \tag{3.15}$$

$$N^p \geq \frac{1}{2} \sum_{r \in R: k \in r} N^p_{r,k} \qquad p \in P, k \in S : \delta^p_k = 2 \tag{3.16}$$

The total spare capacity in each span is given by equation (3.17) and this spare capacity is optimized as per the objective (3.10) where:.

$$S_k = \sum_{p \in P: \delta^p_k = 1} N^p \qquad k \in S. \tag{3.17}$$

The intermediate parameter α^p_k is introduced for enumerating the number of spans in the set $S^p_{\overline{r}}$. This parameter indicates whether the straddling span (k) is protected by p-cycle (p) or not. In other words, if the p-cycle (p) is allocated for the protection of straddling span (k) in any working path (r), then α^p_k should be equal to one.

Therefore the value of this parameter can be found by the following identity.

$$\alpha_k^p = \max_{r \in R : k \in r} N_{r,k}^p, \qquad p \in P, k \in S : \delta_k^p = 2. \tag{3.18}$$

The identity (3.18) can be translated to the following set of constraints which are useful in the ILP model.

$$N_{r,k}^p \leq \alpha_k^p, \qquad k \in r : \delta_k^p = 2, p \in P, r \in R \tag{3.19}$$

On the other hand, when a straddling span (k) is protected by the cycle (p), there should be at least one working path (r) that traverses span (k) and is protected by cycle (p). This condition is satisfied by the following constraint.

$$\sum_{r \in R : k \in r} N_{r,k}^p \geq \alpha_k^p, \qquad k \in S : \delta_k^p = 2, p \in P. \tag{3.20}$$

Now, we constrain the unavailability of the working path to a desired upper limit which is our main objective. The upper limit is an input parameter to our ILP. The program will define p-cycles, such that this constraint is satisfied. That is, for a lower value of the constraint, p-cycles with less hop-count will be allocated and for a more relaxed value of the constraint the optimizer will tend to allocate longer p-cycles. Obviously, if the desired value for unavailability is too low, a solution may not exist.

$$U_r \leq MU, \qquad r \in R. \tag{3.21}$$

The end-to-end unavailability of a working path (r) is computed by adding all U_r^p in all domains of the working path (r):

$$U_r = \sum_{p \in P} U_r^p, \qquad r \in R; p \in P. \tag{3.22}$$

Accordingly, we can compute U_r^p which is the unavailability of the protection domain in working path (r) protected by p-cycle (p). This could be done by using the output of equations (3.24) to (3.27) and using equation (3.7):

$$\begin{aligned} U_r^p = \{ & O_r^p \cdot O_{\bar{r}}^p + \frac{1}{2} O_r^p \cdot S_{\bar{r}}^p + O_r^p \cdot S_r^p + \\ & \frac{3}{4} S_r^p \cdot O_{\bar{r}}^p + \frac{1}{2} S_r^p \cdot (S_r^p - 1) + \\ & \frac{1}{2} S_r^p \cdot S_{\bar{r}}^p + \frac{1}{2} O_r^p \cdot (O_r^p - 1) \} \cdot U^2, \\ & r \in R; p \in P. \end{aligned} \tag{3.23}$$

By using the intermediate output ($N_{r,k}^p$), we can determine the variables of interest: O_r^p, $O_{\bar{r}}^p$, S_r^p, and $S_{\bar{r}}^p$. Equations (3.24) to (3.27) calculate them.

$$O_r^p = \sum_{k \in r; \delta_k^p = 1} N_{r,k}^p, \qquad r \in R; p \in P \tag{3.24}$$

$$O_{\bar{r}}^p = l^p - \sum_{k \in r; \delta_k^p = 1} N_{r,k}^p, \qquad r \in R; p \in P \tag{3.25}$$

$$S_r^p = \sum_{k \in r; \delta_k^p = 2} N_{r,k}^p, \qquad r \in R; p \in P \tag{3.26}$$

$$S_{\bar{r}}^p = \sum_{k \in R \setminus \{r\}; \delta_k^p = 2} \alpha_k^p + \sum_{k \in R; \delta_k^p = 2} (1 - N_{r,k}^p) \cdot \alpha_k^p \qquad r \in R; p \in P \tag{3.27}$$

As can be seen, the second summation in equation (3.27) contains quadratic terms

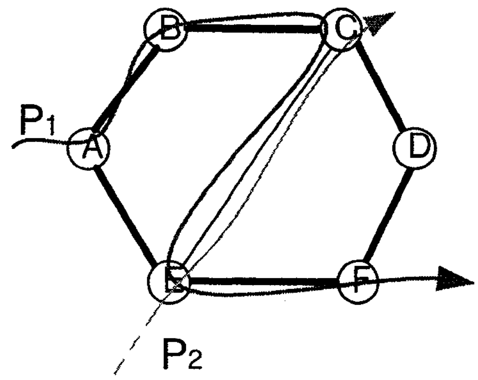


Figure 3.5: An example for having non-zero quadratic terms in the equation for computing $S^p_{\overline{r}}$

$\alpha^p_k \cdot N^p_{r,k}$. For determining the contribution of unavailability in categories 2 and 6, $S^p_{\overline{r}}$ is multiplied by O^p_r and S^p_r. Due to the quadratic terms in $S^p_{\overline{r}}$, these multiplications result in terms with degree 3 which eventually make the problem extremely hard to solve.

Meanwhile, it is important to observe that, the quadratic term in $S^p_{\overline{r}}$ refers to very rare cases that hardly occur in practical problems. In order to have non-zero quantities in the second summation of equation (3.27), there should be a path (r) that traverses a straddling span (k) which is not protected by (p), but the span (k) is also traversed by another service path ($r\prime$) which is protected by cycle (p). Figure 3.5 illustrates an example of such highly improbable cases which the straddling span $C - E$ traversed by P_1 and P_2 is only protected by p-cycle for P_2. Therefore, at this time, the second term in equation $S^p_{\overline{r}}$ can be neglected without the anxiety of losing the exactness.

3.4.3 Linearizing the Non-linear Constraints

The exact model for availability-aware design based on p-cycle has been formulated as an ILP model. As can be seem from the previous section, expression (3.23) exhibits some non linearities due to the multiplication of the expressions (3.24) to

(3.27) which yields non-linear constraints (3.21) and that renders the model difficult to solve. For example, to find the unavailability for Category-1, we write:

$$O^p_r \cdot O^p_{\bar{r}} = (\sum_{k \in r; \delta^p_k = 1} N^p_{r,k}) \cdot (l^p - \sum_{k \in r; \delta^p_k = 1} N^p_{r,k}) =$$

$$l^p \cdot \sum_{k \in r; \delta^p_k = 1} N_{r,k} - \sum_{k,k' \in r; \delta^p_k = \delta^p_{k'} = 1} N^p_{r,k} \cdot N^p_{r,k'} \qquad r \in R, p \in P \tag{3.28}$$

Indeed, the quadratic terms of $N^p_{r,k} \cdot N^p_{r,k'}$ make the expression nonlinear. To deal with this issue, we define a new variable:

$$N^p_{r,k} \cdot N^p_{r,k'} = Z^p_{r,kk'}, \qquad k, k' \in r; \delta^p_k = \delta^p_{k'} = 1; r \in R, p \in P \tag{3.29}$$

Accordingly, we replace the new variable $Z^p_{r,kk'}$ for finding the number of dual failures in categories of 1, 3, 5, and 7. This variable replacement results in the following new constraints:

$$Z^p_{r,kk'} \leq N^p_{r,k} \qquad k, k' \in r; \delta^p_k = \delta^p_{k'} = 1; r \in R, p \in P \tag{3.30}$$

$$Z^p_{r,kk'} \leq N^p_{r,k'} \qquad k, k' \in r; \delta^p_k = \delta^p_{k'} = 1; r \in R, p \in P \tag{3.31}$$

$$Z^p_{r,kk'} \geq N^p_{r,k} + N^p_{r,k'} - 1 \qquad k, k' \in r; \delta^p_k = \delta^p_{k'} = 1; r \in R, p \in P \tag{3.32}$$

Similarly, for finding the number of dual failures in the second category, we obtain the multiplication of the sets O^p_r and $S^p_{\bar{r}}$ as follows :

$$O_r^p \cdot S_{\bar{r}}^p = (\sum_{k\in r;\delta_k^p=1} N_{r,k}^p) \cdot \sum_{k\in r';\delta_k^p=2} \alpha_k^p =$$

$$\sum_{k\in r;\delta_k^p=1} \sum_{k'\in r' \delta_{k'}^p=2} N_{r,k}^p \cdot \alpha_{k'}^p \qquad r \in R, r' \in R \setminus \{r\}, p \in P \tag{3.33}$$

Similar to before, the multiplication of the two variables (i.e., $N_{r,k}^p$ and α_k^p) make our model nonlinear; hence we replace this multiplication with new variables as follows:

$$N_{r,k}^p \cdot \alpha_{k'}^p = Y_{r,kk'}^p \qquad k \in r, k' \in r'; \delta_k^p = 1, \delta_{k'}^p = 2; r \in R, r' \in R \setminus \{r\}, p \in P \tag{3.34}$$

The following are the resulting constraints in categories of 2, 4, and 6:

$$Y_{r,kk'}^p \leq N_{r,k}^p \qquad k \in r, k' \in r'; \delta_k^p = 1, \delta_{k'}^p = 2; r \in R, r' \in R \setminus \{r\}, p \in P \tag{3.35}$$

$$Y_{r,kk'}^p \leq \alpha_{k'}^p, \qquad k \in r, k' \in r'; \delta_k^p = 1, \delta_{k'}^p = 2; r \in R, r' \in R \setminus \{r\}, p \in P \tag{3.36}$$

$$Y_{r,kk'}^p \geq N_{r,k}^p + \alpha_{k'}^p - 1, \qquad k \in r, k' \in r'; \delta_k^p = 1, \delta_{k'}^p = 2; r \in R, r' \in R \setminus \{r\}, p \in P \tag{3.37}$$

Upon replacing the new variables (i.e., $Z_{r,kk'}^p$ and $Y_{r,kk'}^p$) and adding the new constraints, our model becomes an exact ILP model for availability-aware design in p-cycle based networks.

3.5 Practical Difficulties and Corresponding Alternatives

3.5.1 Promising Cycle Pre-selection

Our new model for availability-aware design based on p-cycle is an exact method only when we consider all possible candidate p-cycles in large-scale networks. In our model, candidate p-cycles are found using a depth-first search; however, considering all possible p-cycles makes the model quite large with a huge number of variables. Accordingly, only a subset of candidate p-cycles is considered. Therefore, the global optimal solution can not be guaranteed and we have to use some criteria for pre-selecting the most promising eligible cycles. Hence, our on-going work is using Column generation approach introduced in to produce cycles dynamically.

In [27], two different metrics of the p-cycles efficiency for cycle pre-selection are suggested: A Priori p-cycle Efficiency AE(r) and Demand-weighted p-cycle Efficiency EW(r). The efficiency measure AE(r) counts the number of protected links, divided by the cost of the p-cycle. In EW(r) the offered protection capacity is weighted with the working capacity which needs to be protected for each link. Hence this measure assumes that the demands are already routed. The optimal p-cycle according to the AE(r) measure is the p-cycle with the lowest average cost for link protection; hence they are compatible for our model and all candidate p-cycles are pre-selected using AE metrics.

3.5.2 Complexity Issues

In the design of survivable large-scale networks, the running time of the model is an important factor; hence, we have to find an estimation of the running time of CPLEX for solving the model. The total number of variables and constraints based on the total number of candidate cycles and demands in a network are the most important factors for estimating the running time. Table 3.2 shows the total number of variables used by the model for the sample network COST239 European network [46], with 254 pre-selected cycles and 174 demands. The demands are randomly assign to pairs of nodes ranging 0 to 5. Clearly, the number of variables is very large which makes the problem unsolvable in reasonable amount of time. The next section presents a method for reducing the complexity of the model.

Table 3.2: Total number of variables in exact model for COST239 network

Variable	Total number
S_k	26
N^p	254
$N^p_{r,k}$	46951
U_r	176
α^p_k	1658
$Z^p_{r,kk'}$	89691
$Y^p_{r,kk'}$	417630
Total Number of 0-1 variables	556386

3.5.3 Introducing Lazy Constraints and Lazy Variables

Clearly, the total number of constraints in the model is one of the most important factors affecting the running time of the model. In our design we find out the number of constraints based on (3.11), (3.12), (3.17), (3.21), (3.30), (3.31), (3.32), (3.35), (3.36), and (3.37). Constraints (3.30), (3.31), (3.32), (3.35), (3.36), and (3.37) which results from the linearization are very costly since their number is very large. In order to overcome this difficulty, we introduce them only as needed following the principle of the "Lazy Constraints" feature of CPLEX. In other words, our model is solved iteratively, starting with no or a very small number of linearization constraints (selected randomly), and adding some linearization constraints which are violated until reaching a feasible solution which satisfies all the linearization constraints.

Similarly, we introduce the variables ($Y^p_{r,kk'}$) as the "Lazy Variables", since the number of these variables is very large (see Table 3.2). We select a small number of them randomly and the rest will be set to zero. As mentioned before, our model is solved iteratively, and each time, we add some of the violated variables which are not equal to zero, until we obtain a feasible solution.

3.5.4 Round-Robin Method

Another heuristic for decreasing the size of ILP problem is Round-Robin method. In this heuristic, the set of cycles is partitioned into small subsets with a predetermined size. Then, the problem is solved by only considering the cycles in one subset. Afterwards, the distinct cycles which are utilized in the solution will be kept in the subset and new cycles will be added until the subset reaches the predetermined size. This procedure iterates until all cycles in the original set are considered. Then, it restarts from the beginning of the cycle list and continues until there is no improvement in the solution.

One major advantage of using Round-Robin method is that the number of variables will be decreased dramatically. This advantage can be seen in Table 3.3 (with subset size 35) when compared to number of variables without Round-Robin method which is shown in Table 3.2.

Table 3.3: Total number of variables in exact model for COST239 network

Variable	Total number
S_k	26
N^p	35
$N^p_{r,k}$	1496
U_r	176
α^p_k	138
$Z^p_{r,kk'}$	2516
$Y^p_{r,kk'}$	7397
Total Number of 0-1 variables	11784

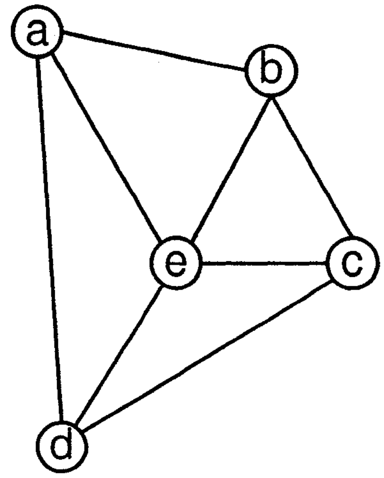


Figure 3.6: Sample Network

3.6 Experimental Results

The formulation provided in the previous section was tested with the sample network (as shown in Figure 3.6), containing 5 nodes and 8 spans. Without loss of generality, we assume that each node can perform full wavelength conversion. We also assume that each span has enough capacity to support the protection capacity required by the optimal solution. The solution for the Integer Linear Programming (ILP) was obtained by implementing the model in C++ using "Concert" API and solving using the solver CPLEX 11.0.1. A separate C program is used to find the input parameters for the C++ model.

The number of lightpaths in either direction between a source and destination pair are equal. In addition, the demands are assigned to a certain number of lightpaths ranging 0 to 5. The demands are routed using a Dijkstra shortest path algorithm to find eligible working routes. Lightpaths between a pair of nodes are routed individually; hence two lightpaths between the same pair of nodes can possibly have different working routes and each lightpath demand is associated with a working route. In addition, the preprocessing program finds candidate p-cycles using a depth-first search

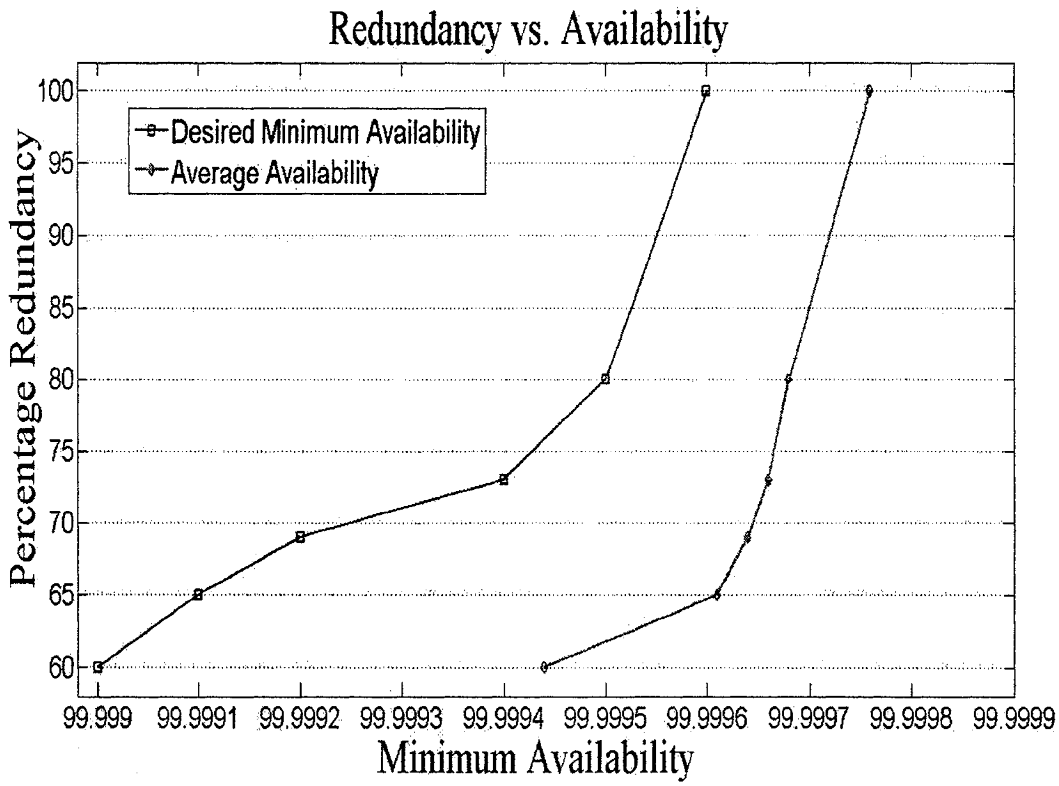


Figure 3.7: Redundancy Versus Minimum Availability

and pre-selected by hop count lengths for our model. All working routes and candidate p-cycles are provided as inputs to the CPLEX model.

In Figure 3.7, we show the redundancy obtained for certain level of availability for the sample network. Here, the ratio of the sum of spare capacity in each span to the sum of the working capacity in each span is defined as redundancy [2]. For computational simplicity, we assume the unavailability of each span to be equal. In our case, we assumed that the unavailability of each span is equal to 10^{-3}. The same set of demands is used to find the redundancy for each availability requirement.

Clearly, these results indicate that the availability increases at cost of additional spare capacity requirement. As shown in Figure 3.7, always the average availability of service paths are more than the minimum required availability; hence most of the demands have their availabilities spread. For instance, if the desired level of availability is 99.9991, the average availability of all demands is 99.9996. Indeed, the

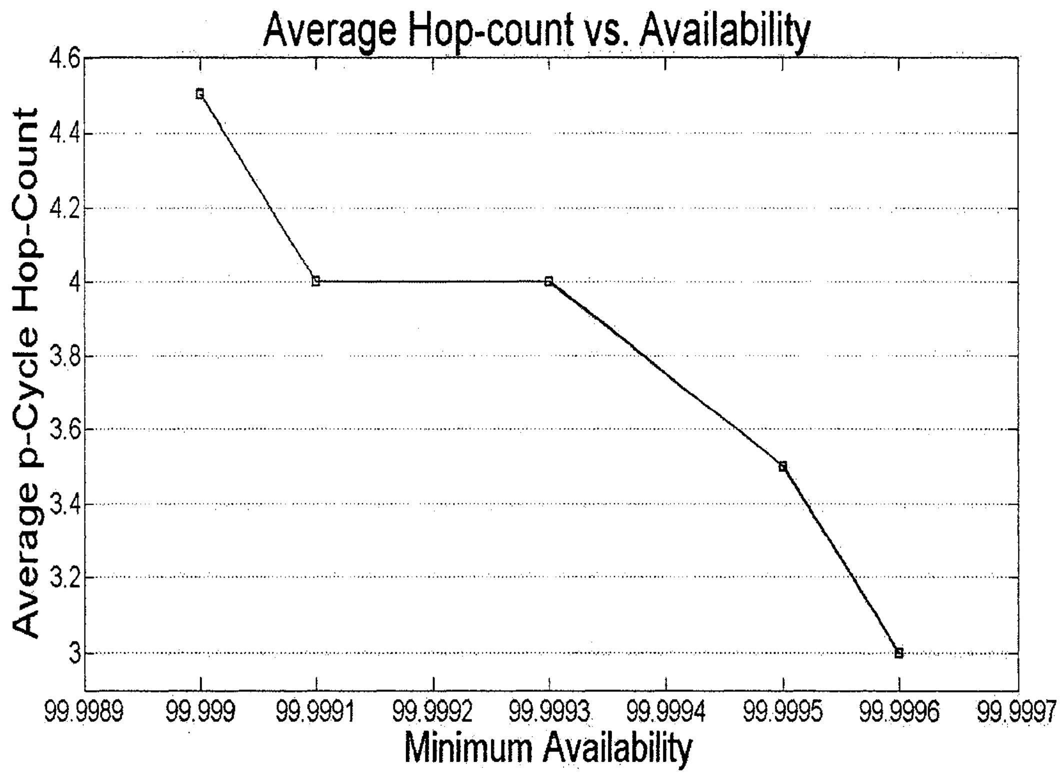


Figure 3.8: Average p-cycle hop-count versus availability for all working paths

availability-aware design results in over provisioning protection capacity in order to achieve the desired level of availability. That is, although all demands will enjoy the level of availability required, the solution is not optimal. As explained in previous section, since all existing cycles can not be considered in the model, our future work is to utilize exact methods, such as column generation, in order to provide the most promising cycles dynamically.

Next, we tried to obtain some insights on the impact of the length of p-cycle (in terms of hop-counts) on the service availability. Here, we are interested in analyzing the average cycle length, since longer cycles tend to be more efficient from spare capacity redundancy point of view. We show in Figure 3.8 the average hop-count of p-cycles for different availability requirements. Obviously, the average cycle length decreases as the level of availability is increased. Accordingly, shorter p-cycles tend to limit the sharing of spare capacity more than longer p-cycles; hence, higher resource

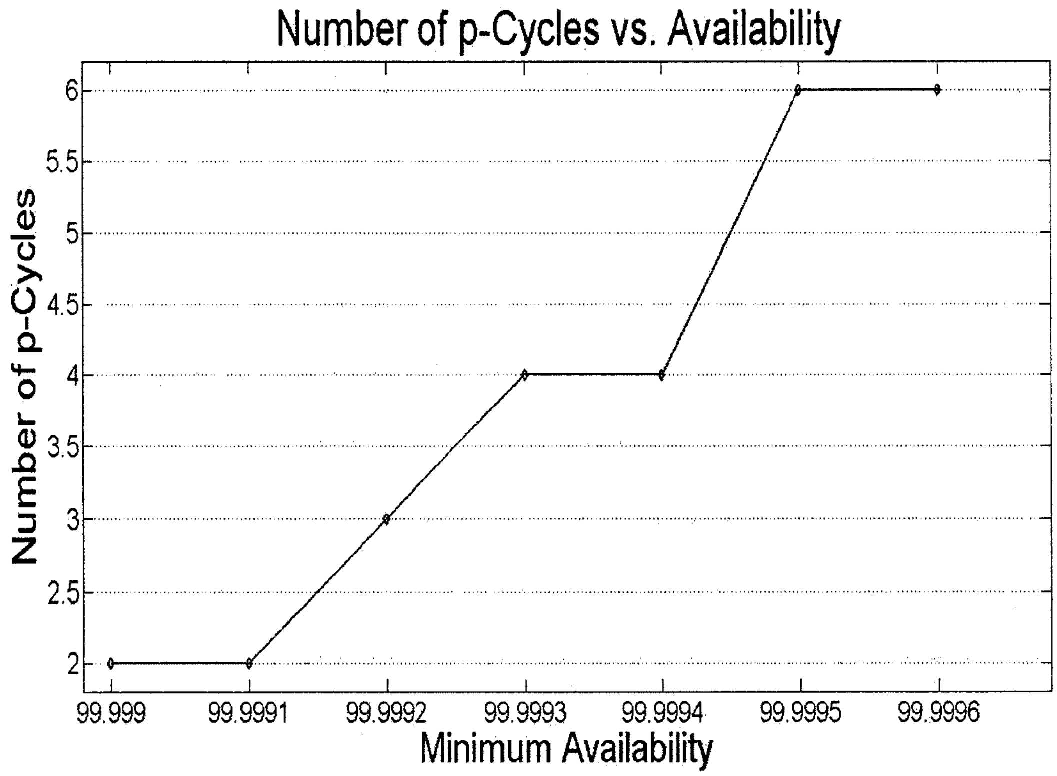


Figure 3.9: Number of p-cycles utilized by model versus availability

redundancy is obtained.

Finally we show the number of p-cycles utilized by the model to protect all the demands against single failure, with achieving certain level of availability; the results are shown in Figure 3.9. It can be seen that the higher the required availability in our design, the more is the number of p-cycles used by model. For example, if the required availability is 99.999, the number of p-cycles utilized by the model is 2, as opposed to 6 when the required availability is 99.9996. Clearly, utilizing more number of p-cycles explains the higher redundancy obtained by our model for achieving the higher level of availability.

3.7 Conclusion

In this chapter, we revisited the problem of availability analysis in p-cycle based networks and presented an exact model for availability-aware provisioning by highlighting some flaws in prior work. The new model provided a technique for allocating p-cycles to restore single link failures such that the unavailability of all the demands in the network is bounded by an upper limit. We then provided some heuristic to restrict the number of variables and constraints in an ILP formulation in order to solve our model in a reasonable amount of time. Our results showed that by adding limitation on the availability, the model will utilize more number of p-cycles with smaller hop-counts. We have to mention that our conclusions are drawn based on the sample network used for the experiments. In the future, larger networks will be considered in our study.

Chapter 4

Availability-Aware Design in Mesh Networks with Failure-Independent Path- Protecting (FIPP) p-Cycles

4.1 Introduction

As we already mentioned early, survivability of high-capacity optical wavelength-division multiplexing (WDM) mesh networks has been a topic of great interests for many years now. A (component) failure in this type of networks can lead to severe service disruption. Hence, various types of survivability mechanisms have been developed against network element failures. As mentioned before, these mechanisms can be classified into two main large categories; namely, protection and restoration [14], [2]. The main objective of both survivability mechanisms is to recover and replace all the affected service paths after the occurrence of a single network failure. However, using either protection or restoration methods to make a network fully restorable for single failures is not a guarantee that the availability of the network in the occurrence of the second failure will be 100%. Hence, availability-aware network design has emerged as an important topic for achieving some level of robustness against multiple failures.

Currently, a variety of protection and restoration methods exist for optical transport networks [14]. These include shared-back-up path protection (SBPP), pre-configured protection cycles, and recently failure-independent path-protection (FIPP), among others as mentioned in Chapter 2. These methods have different spare capacity requirement and restoration speed. FIPP p-cycles improve p-cycles by adding the property of providing end-to-end failure independent path switching against a network component failure while retaining other advantages of p-cycles. FIPP protection method has shown to be more efficient than span p-cycle protection method; this is attributed to the fact that more demands can be protected by a single FIPP p-cycle which yields a better resource redundancy. In p-cycle based design, it has been shown that the cycle length plays an important role in determining the unavailability of the portion of the working path which is protected by the cycle [8], [11]. In [1], the authors showed that allocating shorter service paths to longer p-cycles yields better resource efficiency, since longer cycles tend to have more straddling links, and better service availability.

In this work, we try to determine whether the FIPP protection method maintains its advantages, when our design is based on limiting the end-to-end service unavailability. We observe that, similar to span p-cycle protection method, the length of the FIPP p-cycle plays an important role in determining the availability of the service working path. That is, higher network availability is obtained when the FIPP p-cycle contains fewer hops. Alternatively, as the length of the FIPP p-cycle becomes shorter, the redundancy increases for a network with high availability. In addition, the number of demands protected by the same FIPP p-cycle affects both the efficiency of the FIIP protection method as well as the availability of service working path. We notice that the higher the required availability, the less efficient FIPP p-cycle becomes. This is because the design will limit the number of demands sharing the same FIPP p-cycle. Accordingly, we note that both high availability and better resource efficiency are conflicting objectives. In addition, when availability is not a design issue, we notice that the FIPP protection method yields lower average service availability than span p-cycle protection method.

4.1.1 Outline

In Section 4.2 we show, with the help of examples, a computational analysis of the unavailability of the working path which is protected by a single FIPP p-cycle. In this section, we try to find out the unavailability for different combination of dual failures which results in a service outage. Enumerating the unavailability of dual failures in FIPP p-cycle based networks is not as simple as that of p-cycle based networks. Hence, in this section, after explaining each specific case we try to find out the upper and lower bound for unavailability of dual failures. In Section 4.3 we use the analysis developed in the previous section to derive an Integer Linear Program (ILP) model. This model optimizes the allocation of FIPP p-cycles for pre-routed working paths and allocates FIPP p-cycles such that the unavailability of each working path is less than a desired upper limit. This desired upper limit and other parameters such as the set of eligible cycles, spans and working paths are given as input parameters to the ILP model. The model presented is based on the analysis of availability of network during the effects of dual-failures. In Section 4.4, the results and analysis of ILP model and comparison of availability in span p-cycle and FIPP p-cycle are provided. Finally Section 4.5 summarizes our findings.

4.2 Unavailability Analysis in FIPP p-Cycle-Based Networks

The most common aim in designing for survivability in any network is to obtain restorability against all single failures with an objective to minimize the required spare capacity. This makes the dual failures as the main factor which can contribute to the unavailability of the service. Since the probability of occurrence of triple failures is significantly less than the probability of occurrence of dual failures, it is reasonable to neglect the probability of occurrence of triple and more failures in networks. Moreover, in networks where pre-connected protection schemes have been used to achieve restorability against single failures, the recovery speed is expected to be very fast. In such networks the unavailability of service during the recovery time is also insignificant. Numerical examples in [8] have shown that the effect of dual span-failures are in fact more important in determining the expected service path unavailability and considering only dual failures is sufficient to obtain a good estimate on the availability of the service.

In previous works, analyzing the unavailability of a set of spans protected by a span p-cycle is done by using the concept of protection domain [11], [1]. In [1], a protection domain is defined as the set of spans which are protected by the same span p-cycle; however, in [11], if a span on a path was protected by a p-cycle as an on-cycle span and another span on the same path was protected as a straddling span, then these two spans were counted as two different domains.

In this work, we modify the definition of protection domain since any working path could be protected by only one FIPP p-cycle. Accordingly, we define a protection domain as the set of mutually disjoint working paths which are protected by the same FIPP p-cycle. In other words, each FIPP p-cycle can protect a set of disjoint working paths. These working paths and their respective spans are all in one protection domain. Here, according to this definition, we are modeling our network as a set of mutually disjoint working paths with a set of FIPP p-cycles. Please note that according to our objective, which is protecting only against span failures, the mutual disjointness is properly applied only to spans.

4.2.1 Unavailability Analysis of a Working Path

We will analyze the different categories of all dual span-failures that can result in an outage for a given working path in a FIPP p-cycle protection network. We find and categorize all possible combinations of dual failures such that all these combinations are independent from each other and a dual failure can only belong to one of these categories. This analysis assumes that each span has the same physical unavailability (U). In this thesis, the extension of unavailability expression which is developed to consider span specific unavailability should be straightforward. We have derived six different sets of spans in a FIPP p-cycle by using the following notations:

O_r^p: The set of on-cycle spans in FIPP (p) that are on the working path (r).

$O_{\bar{r}}^p$: The set of spans in FIPP (p) that are on-cycle but are not on the working path (r).

S_r^p: The set of spans that are on the working path (r) which is totally straddling the FIPP (p).

$S_{\bar{r}}^p$: The set of spans traversed by those paths which straddle the FIPP (p) and are not part of the concerned path.

OS_r^p: The set of spans that are on the working path (r) which is partially on the FIPP (p). These spans belong only to the straddling part of path (r).

$OS_{\bar{r}}^p$: Set of spans whose paths are partially on the FIPP (p) but they are neither on the concerned path (r) nor on the FIPP cycle (p).

We can categorize all dual failure scenarios that may cause any outage within a FIPP p-cycle into one of the following categories:

Category-1: Dual failure scenario in which one of the spans belongs to O_r^p and the other span belongs to $O_{\bar{r}}^p$.

Category-2: Dual failure scenario in which one of the spans belongs to O_r^p and the other span belongs to $S_{\bar{r}}^p$.

Category-3: Dual failure scenario in which one of the spans belongs to O_r^p and the other span belongs to $OS_{\bar{r}}^p$.

Category-4: Dual failure scenario in which one of the spans belongs to S_r^p and the other span belongs to $O_{\bar{r}}^p$.

Category-5: Dual failure scenario in which one of the spans belongs to S_r^p and the other span belongs to $S_{\bar{r}}^p$.

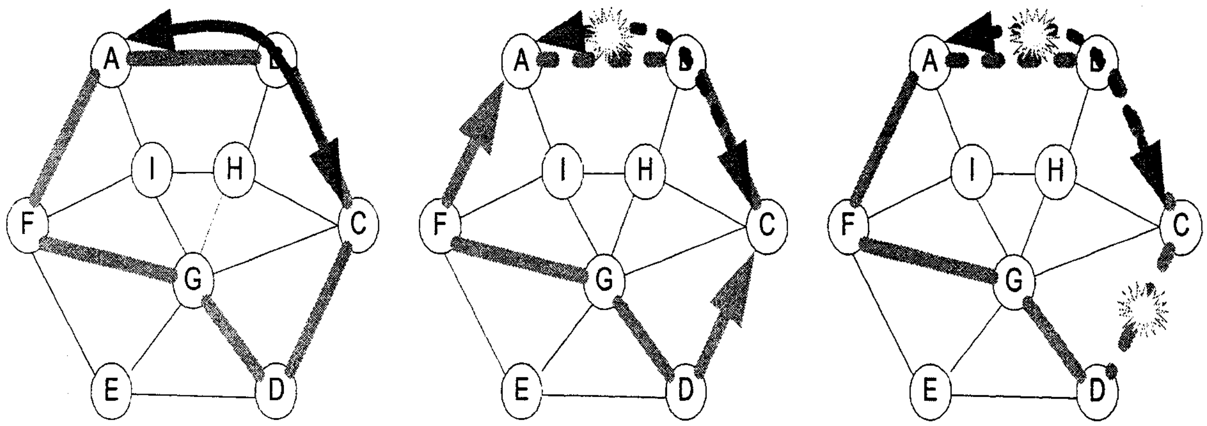
Category-6: Dual failure scenario in which one of the spans belongs to S_r^p and the other span belongs to $OS_{\bar{r}}^p$.

Category-7: Dual failure scenario in which one of the spans belongs to OS_r^p and the other span belongs to $O_{\bar{r}}^p$.

Category-8: Dual failure scenario in which one of the spans belongs to OS_r^p and the other span belongs to $S_{\bar{r}}^p$.

Category-9: Dual failure scenario in which one of the spans belongs to OS_r^p and the other span belongs to $OS_{\bar{r}}^p$.

We will be referring to these categories throughout the rest of this chapter. As we explained in the FIPP protection section, end nodes for protection of a partially on-cycle path can switch to the default or secondary direction of the FIPP p-cycle. Hence, depending on place of failure occurrence, two different paths of protection may exist for one partially on-cycle path. Having two different protection paths results in some special cases, where in some of them we can restore the second failure with the rest of the cycle and in some other cases we may not use the rest of the cycle to protect the second failure (as will be shown shortly). Accordingly, we will find the upper and lower bounds of unavailability for this kind of paths, rather than exact



(a) No failure (b) After first failure (c) After second failure

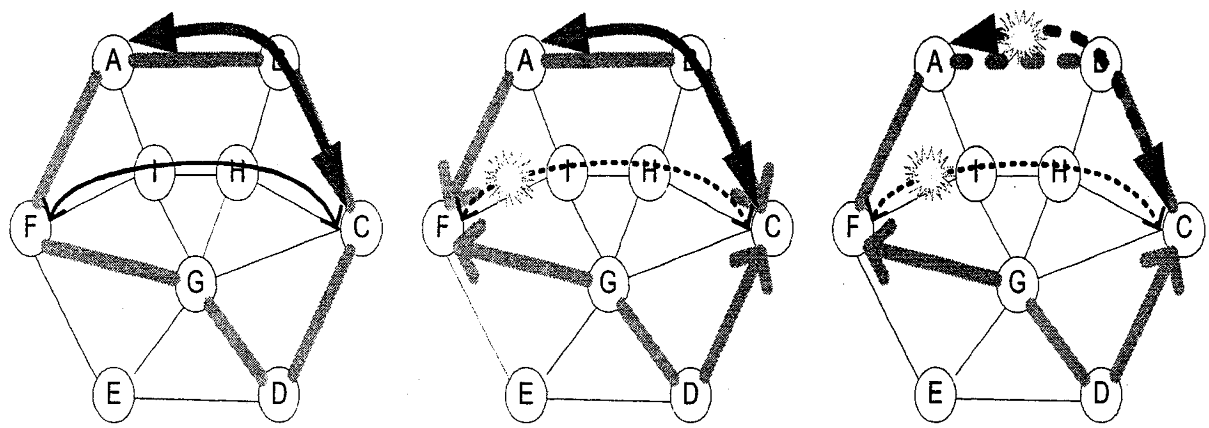
Figure 4.1: Dual failure leading an outage for a totally on-cycle working path ($A - B - C$, path of concern), Category-1

values. Next, we will analyze the probability of having an outage in each of the nine scenarios with the help of few examples.

4.2.2 Example 1

In the first example, we analyze the first three categories which are the combinations of the set O_r^p with three other sets($O_{\bar{r}}^p$, $S_{\bar{r}}^p$ and $OS_{\bar{r}}^p$). Figure 4.1 shows a part of the network in which a working path (path of concern $(A - B - C)$) is totally on-cycle. This on-cycle working path is protected with FIPP p-cycle $(A–B–C–D–G–F–A)$. Our objective is to determine the probability that the traffic entering at source node A suffers an outage before it reaches the destination node C. In other words, we try to derive the unavailability of this working path.

Category-1 : In Figure 4.1, we show the case where the first failure occurs on one of the spans $(A - B)$ or $(B - C)$ on the working path itself followed by the second failure on one of the on-cycle spans that are not on the working path (i.e., one span among the spans $(C - D)$, $(D - G)$, $(G - F)$, and $(F - A)$). Scenarios like this are certain to cause an outage, because the FIPP p-cycle only offers one protection option for an on-cycle working path. Obviously, the case where the two failures occur



(a) No failure (b) After first failure (c) After second failure

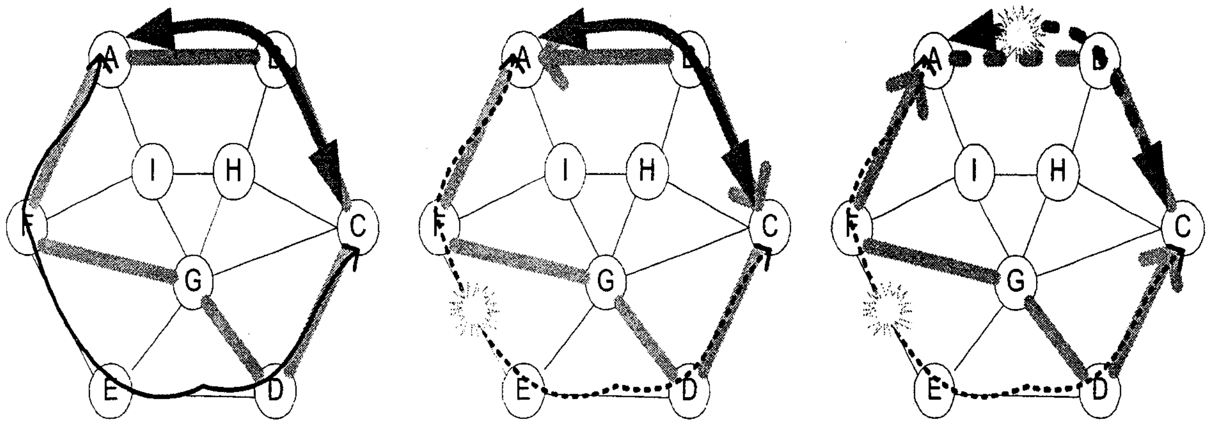
Figure 4.2: Dual failure leading an outage for a totally on-cycle working path ($A - B - C$, path of concern), Category-2

in the reverse order is also guaranteed to cause an outage for the working path. The number of possible combinations of dual failures is $|O^p_r|.|O^p_{\overline{r}}|$. Thus, we denote the exact unavailability due to a dual failure in this category as (where U is the physical unavailability of any span in the network):

$$U_{Category-1} = |O^p_r|.|O^p_{\overline{r}}|.U^2 \tag{4.1}$$

Category-2 : The second category of dual failures that may result in an outage for an on-cycle working path in this example is Category-2 in which one of the spans belongs to O^p_r (i.e., either of the spans $(A - B)$ or $(B - C)$) and the other spans belong to $S^p_{\overline{r}}$ (i.e., any of the spans $(C - H)$, $(H - I)$ and $(I - F)$) where another demand is routed. In this category, if the first failure occurs on a straddling path and assume that the FIPP p-cycle is fully loaded [1], the straddling path would be restored by using both parts of the FIPP p-cycle. Thus the second failure on any span in the set O^p_r will result in a service outage for the working path as shown in Figure 4.2. Unlike the previous category, the order in which failures occur is important. If

[1]Fully loaded FIPP p-cycle is one which provides restoration to tow units of working capacity in all straddling working paths and one unit of working capacity to all totally and partially on-cycle working paths



(a) No failure (b) After first failure (c) After second failure

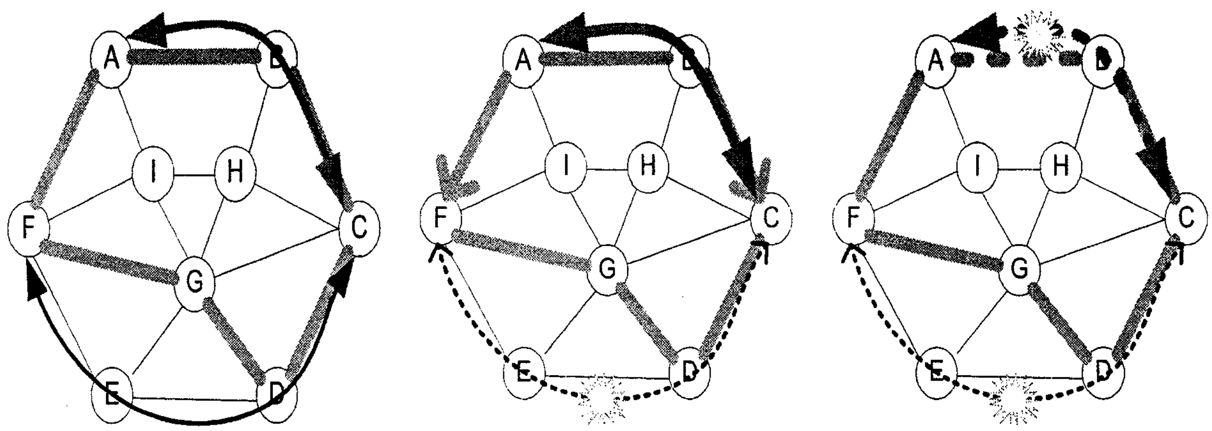
Figure 4.3: Special dual failure without any outage for a totally on-cycle working path ($A - B - C$, path of concern), Category-3

the first failure occurs on a span belonging to the set O_r^p , then the working path would be switched to the default back-up path by the end nodes. The second failure occurring on a straddling working path will not affect the restoration path of the first failure, thus there would be no service outage. We formally denote the unavailability due to dual failures in this category by:

$$U_{Category-2} = \frac{1}{2}|O_r^p|.|S_{\bar{r}}^p|.U^2 \tag{4.2}$$

Note that, here, the factor 0.5 serves the probability that a straddling working path fails first.

Category-3 : In this example, the third category of dual failures that may result in an outage for an on-cycle working path is considered. In this case, one of the spans belongs to O_r^p (i.e., either of the spans $(A - B)$ or $(B - C)$) and the other spans belong to $OS_{\bar{r}}^p$ (i.e., $(D - E)$, or $(E - F)$) where we are considering the second demand $(C - D - E - F - A)$. As we explained before, since we have a partially on-cycle path in this category, we try to find out the upper and lower bounds for unavailability of this category. If the first failure occurs on one of the on-cycle spans in the working path (i.e, $A - B - C$ shown in Figure 4.3), the FIPP p-cycle uses



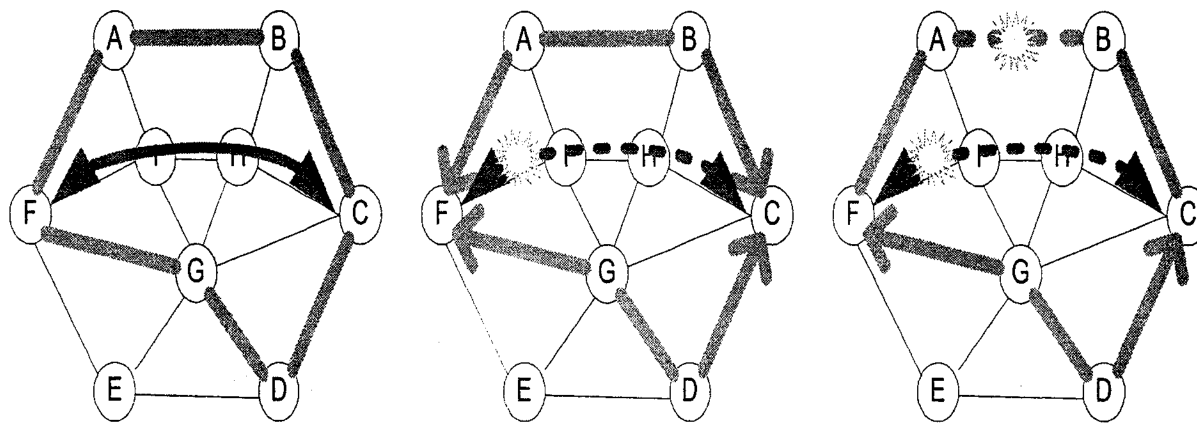
(a) No failure (b) After first failure (c) After second failure

Figure 4.4: Dual failure leading an outage for a totally on-cycle working path ($A - B - C$, path of concern), Category-3

the rest of the cycle to restore the on-cycle working path; hence the second failure on the partially on-cycle path in the set OS_r^p will not affect the restoration path ($C-D-G-F-A$), and there would be no service outage (for the path of concern). It should be mentioned that the set OS_r^p consists of spans which are belonging to the straddling part of the path. The rest of the spans in the path are on-cycle spans, and they are considered in the first category (i.e., Category-1).

In this category, the order in which the failures occur is important. If the first failure occurs on a span belonging to OS_r^p, then the path would be switched to the default back-up path ($A - B - C$) by the end nodes. The second failure occurring on the on-cycle working path will affect the back-up path of the first failure, but the end nodes can restore the on-cycle working path by switching to the rest of the cycle ($A - F - G - D - C$). Figure 4.3 presents an example of this kind of dual failure. In this case, the demand can be restored by using the FIPP p-cycle, hence there would be no service outage and accordingly the lower bound for the unavailability contribution by this type of dual failure is:

$$U_{Category-3-lower} = 0 \tag{4.3}$$



(a) No failure (b) After first failure (c) After second failure

Figure 4.5: Dual failure leading an outage for a totally straddling working path ($F - I - H - C$, path of concern), Category-4

Alternatively, there are some cases where the end nodes could not restore a failure on the concerned working path. Here, if the first failure occurs on the partially on-cycle path ($C - D - E - F$), Figure 4.4, the second failure on the on-cycle path ($A - B - C$) *may* cause a loss on the backup path ($F - A - B - C$) restoring the first failure. Since this failed backup path and the backup path for the path of concern shares span ($A - F$), then the concerned service path cannot be restored. Note, however, if the end nodes ($F - C$) can release the resources (after the second failure) on backup path $F - A - B - C$, then the concerned path may not suffer any outage and end nodes A, C can recover the demand using the rest of the FIPP. This, however, is not suitable as it incurs additional delays and requires signaling. In this special case, if the dual failure occurs in the reverse order, clearly there would be no service outage. Obviously, for the upper bound, the number of possible combinations of dual failures in this category is $(|O_r^p|.|OS_{\bar{r}}^p|)/2$. Hence, the upper bound unavailability is:

$$U_{Category-3-Upper} = \frac{1}{2}|O_r^p|.|OS_{\bar{r}}^p|.U^2 \qquad (4.4)$$

The upper bound of unavailability of the working path ($A - B - C$) can therefore be calculated as the sum of the unavailability obtained in equations 4.1, 4.2 and 4.4:

$$U_{total-upper} = U_{Category-1} + U_{Category-2} + U_{Category-3-Upper} \tag{4.5}$$

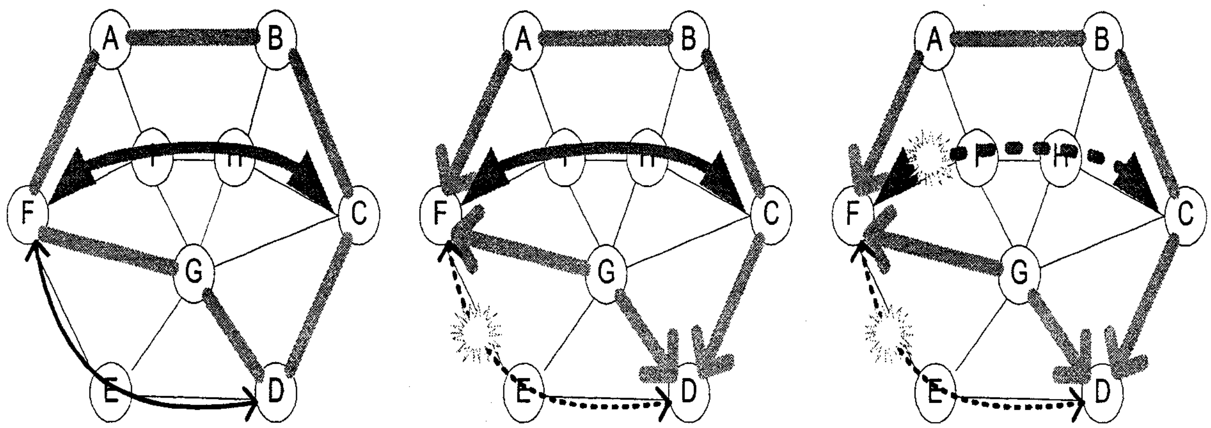
Similarly, we can derive the lower bound for unavailability of the working path as follow:

$$U_{total-lower} = U_{Category-1} + U_{Category-2} + U_{Category-3-Lower} \tag{4.6}$$

In the working path $(A-B-C)$ for all categories $1, 2$ or 3 in this example, if the failures occur in both spans $A-B$ and $B-C$, it can be shown that there would be no outage. When the first failure occurs in either $A-B$ or $B-C$, end nodes A and C will switch to default back up path and after that there would no traffic on the former working path. Hence, the second failure on a span belonging to the former working path does not affect the back-up path.

4.2.3 Example 2

Figure 4.5 shows part of a network in which a working path crosses a FIPP p-cycle $(A-B-C-D-G-F-A)$ with three spans $((C-H)$, $(H-I)$ and $(I-F))$ which are totally straddling the FIPP p-cycle. In this example, we analyze the second three categories (4, 5, and 6) which are the combinations of the set S^p_r with three other sets $(O^p_r, S^p_r$ and $OS^p_r)$. As in the previous example, we try to find the unavailability of this working path $(C-H-I-F)$. In other words, for each of these three categories we find the number of combinations of dual failures that can result in an outage.

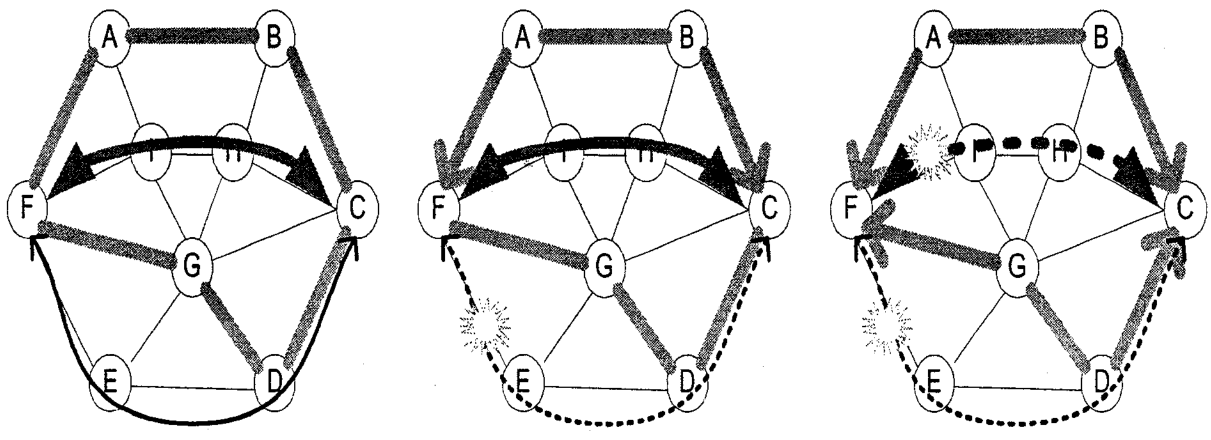


(a) No failure (b) After first failure (c) After second failure

Figure 4.6: Dual failure leading an outage for a totally straddling working path ($F-I-H-C$, path of concern), Category-5

Category-4 : In Figure 4.5 we show the case where one span among the straddling working spans fails and one span among the on-cycle spans not crossed by the working path fails. This means that one span belongs to set S_r^p and the next one belongs to the set $O_{\bar{r}}^p$. There is 50% chance that the on-cycle span fails first which is then followed a second failure occurring on the concerned straddling working path. This order of dual failures will certainly result in service outage along the working path. On the other hand, if the order is reversed, i.e., the first failure occurs on a span along the straddling working path and the second failure occurs on one of the on cycle spans. the second failure on the on-cycle span will affect *only one* of the protection routes for the straddling path of concern (either lower or upper part of the FIPP). Given this (un-ordered) combination of failures, the probability of an outage for concerned path is therefore 75%. The number of combination of dual failures in this category is $|S_r^p|.|O_{\bar{r}}^p|$. We formally denote the unavailability due to a dual failure in this category as:

$$U_{Category-4} = \frac{3}{4}|S_r^p|.|O_{\bar{r}}^p|.U^2 \tag{4.7}$$



(a) No failure (b) After first failure (c) After second failure

Figure 4.7: Special dual failure in which network can protect only one unit of capacity in totally straddling working path ($F-I-H-C$, path of concern), Category-6

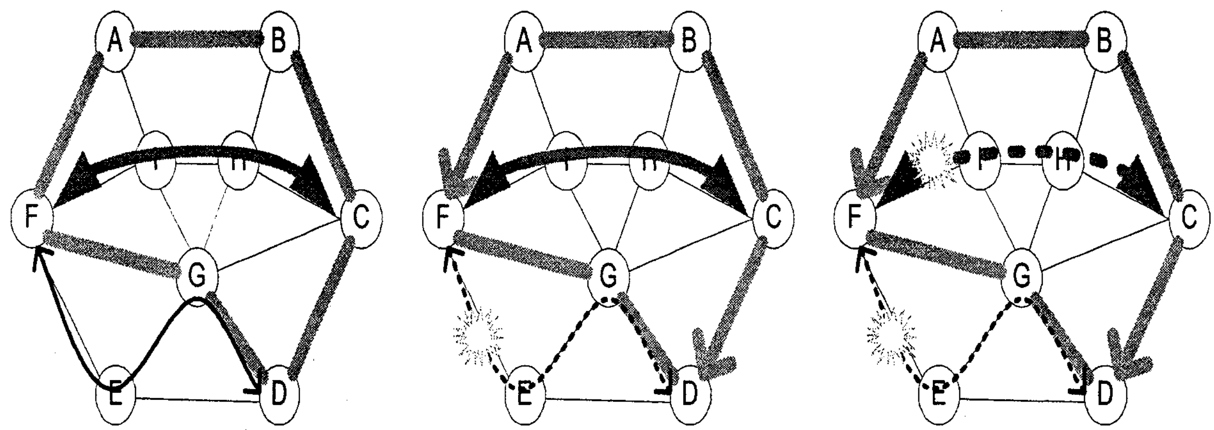
Category-5 : Here we consider the straddling working path (path of concern) $C-H-I-F$ and analyze its availability. In this category, one span failure belongs to the set S^p_r(i.e., $(C-H)$, $(H-I)$ or $(I-F)$) and the other one belongs to the set $S^p_{\bar{r}}$(i.e., $(D-E)$ or $(E-F)$) where we show another straddling path (Figure 4.6). With our assumption in this thesis that the FIPP p-cycle is fully loaded, the occurrence of the failure on this totally straddling path $(D-E-F)$ will activate both halves of the FIPP p-cycle; therefore, a second failure on the working path ($F-I-H-C$, path of concern) will definitely lead to an outage and that is shown in Figure 4.6. In this category, the order of failures is important. If the first failure occurs on a span that belongs to the concerned straddling working path, the end nodes utilize both halves of the FIPP p-cycle ($F-A-B-C$ and $C-D-G-F$), and the second failure on another totally straddling path does not affect the back-up path. Hence there would be no service outage and the unavailability contribution due to failures in this category to be:

$$U_{Category-5} = \frac{1}{2}|S^p_r|.|S^p_{\bar{r}}|.U^2 \qquad (4.8)$$

Category-6 : Here, we are dealing with a dual failure that may result in an outage for a straddling working path in which one of the spans belongs to the set S_r^p and the other one belongs to set $OS_{\bar{r}}^p$ (i.e., where we have a partially straddling path). The $OS_{\bar{r}}^p$ consists of spans belonging to part of paths which straddle the FIPP p-cycle. Before calculating the unavailability of this category, we have to mention that by having a partially on-cycle path we can not find the exact value for unavailability of this category. Hence, as we explained before, we try to find out the upper and lower bound for unavailability of this category. Here, there is a 50% chance that the first failure occurs on the totally straddling working path (path of concern, $C-H-I-F$). A second failure occurring on the partially on-cycle path ($C-D-E-F$) will not affect the restoration paths ($F-A-B-C$ and $C-D-G-F$) of the working path which failed first, hence there would be no outage. However, the two failures can occur in the reverse order. If the first failure occurs on one of the spans of the set $OS_{\bar{r}}^{\bar{p}}$ (i.e, $(D-E)$ and $(E-F)$), a second failure on the totally straddling working path can be restored by using half of the FIPP p-cycle ($C-D-G-F$); however if we assume that the FIPP p-cycle is fully loaded, only one unit capacity of working path can be restored. Figure 4.7 shows an example of this case. The lower bound of the unavailability due to dual failure in this case is:

$$U_{Category-6-Lower} = \frac{1}{4}|S_r^p|.|OS_{\bar{r}}^p|.U^2 \tag{4.9}$$

For the upper bound of unavailability, if the first failure occurs on the partially on-cycle path ($D-G-E-F$), then a second failure in totally straddling working path may lead to a service disruption (as shown in Figure 4.8) and hence there is a 50% chance that some failure in this category will result in an outage. We can formally denote the upper bound of unavailability due to dual failure in this category by considering the above worst case:



(a) No failure (b) After first failure (c) After second failure

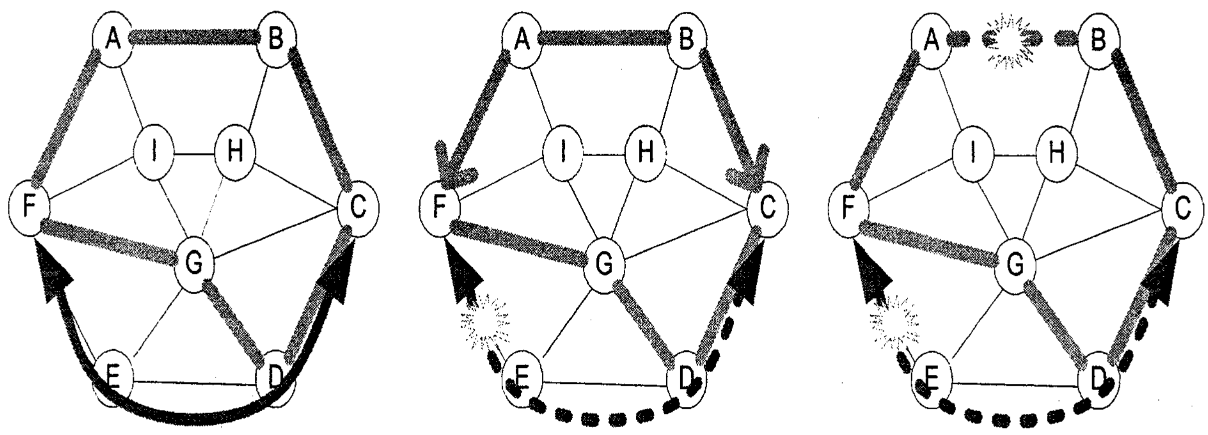
Figure 4.8: Dual failure leading an outage for a totally straddling working path ($F-I-H-C$, path of concern), Category-6

$$U_{Category-6-Upper} = \frac{1}{2}|S^p_r|.|OS^p_{\bar{r}}|.U^2 \tag{4.10}$$

The upper bound of unavailability of the totally straddling working path ($C-H-I-F$) can therefore be calculated as the sum of the unavailability obtained in equations 4.7, 4.8 and 4.10. In the same way, we can derive the lower bound for unavailability of the totally straddling working path by sum of the equations 4.7, 4.8 and 4.9.

4.2.4 Example 3

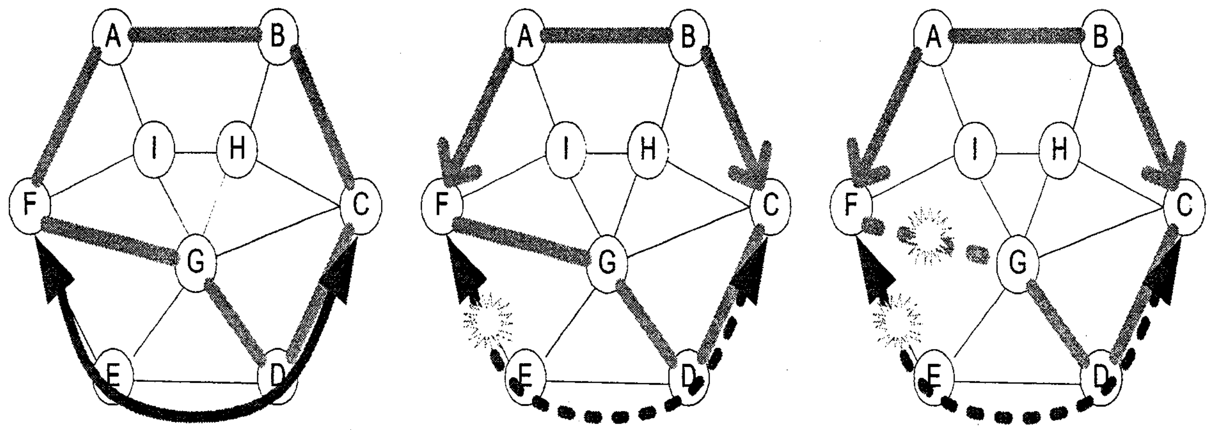
Finally, we take an example of dual failure in which one of the span failure belongs to OS^p_r and the other span failure belongs to one of these sets $O^p_{\bar{r}}, S^p_{\bar{r}}$ and $OS^p_{\bar{r}}$. Figure 4.9 shows a working path traversing a FIPP p-cycle through on-cycle ($C-D$) span and straddling spans (i.e, $D-E$ and $E-F$). In this example, we consider the last three categories ($7, 8,$ and 9) for dual failure. In these three categories, as we explained before, by having a partially on-cycle path we try to find out the upper and lower bound of unavailability instead of exact value.



(a) No failure (b) After first failure (c) After second failure

Figure 4.9: Dual failure leading an outage for a partially on-cycle working path ($C-D-E-F$, path of concern), Category-7

Category-7 : Figure 4.9 and Figure 4.10 show the case where one spans belonging to OS_r^p fails (the path of interest is $C-D-E-F$) and the other span belonging to $O_{\bar{r}}^p$ fails. We consider OS_r^p as a set of spans belonging to the straddling part of the working path (i.e, $(D-E)$, $(E-F)$). The probability of the span failure in the on-cycle part of the working path was calculated in the first category (i.e., category-1). If the first failure occurs on a span in OS_r^p, then the demand of concern will be restored to its protection path $(F-A-B-C)$. When the second failure occurs (on-cycle span fails now), the second failure may in some cases (e.g., failure of on cycle span $(A-B)$) disrupt the protection of the first demand and cause service outage and in some cases (e.g., failure of span $(F-G)$) it may not and the demand will not be affected. Hence, we derive both upper and lower bounds ($|OS_r^p|.|O_{\bar{r}}^p|.U^2$, and 0) for unavailability of this order. Alternatively, if the order of failures is reversed, where the first failure occurs on an on cycle span and the second failure on a span belonging to OS_r^p, then there will be a complete outage. For example, if span $(A-B)$ fails, then any demand routed through $(A-B)$ will be restored through the rest of FIPP and hence the failure of a span on the path of concern will be guaranteed to cause an outage. The unavailability obtained for this order is $|OS_r^p|.|O_{\bar{r}}^p|.U^2$. Finally



(a) No failure (b) After first failure (c) After second failure

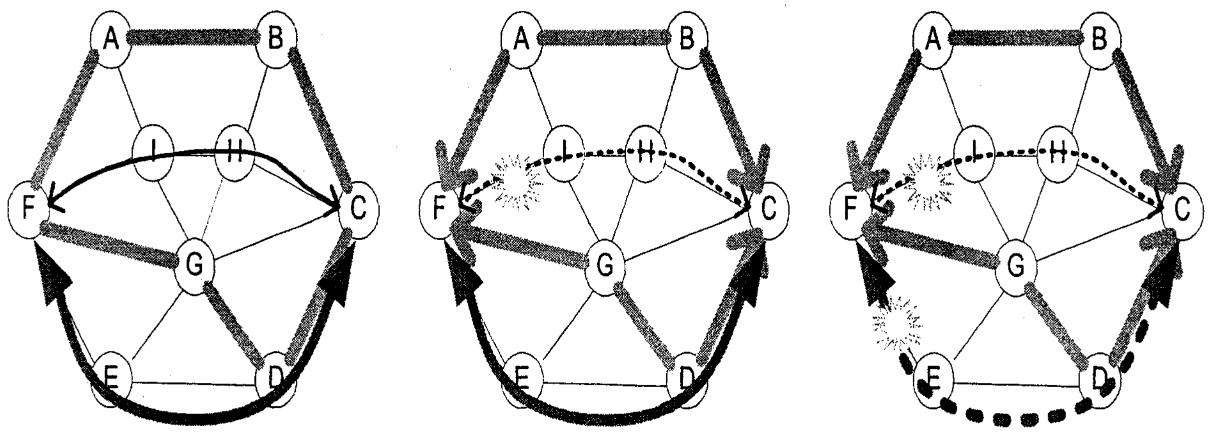
Figure 4.10: Special dual failure without any outage for a partially on-cycle working path ($C - D - E - F$, path of concern), Category-7

since each order are equally likely to occur, then we can derive the lower and upper bounds for this category as follows:

In this category, irrespective of the order in which the failures occur, the working path outage is guaranteed (considering as the worst case). Hence the number of the combinations of such failures is clearly $|OS_{\overline{r}}^{p}|.|O_{\overline{r}}^{p}|$ and the upper bound of the unavailability as a result of dual failure, in this category is given by:

$$U_{Category-7-Upper} = |OS_{\overline{r}}^{p}|.|O_{\overline{r}}^{p}|.U^2 \tag{4.11}$$

Note that in this category, the place of the on-cycle failure is important. We can have some special cases where the second failure belonging to the set $O_{\overline{r}}^{p}$ may not affect the back-up path of the first failure, for instances, as shown in Figure 4.10. Here, the second failure on one of the span $D-G$ or $G-F$ does not affect the back-up path ($F - A - B - C$) of the working path , hence there would no service outage. Obviously, the case where the two failures occur in the reverse order is guaranteed to cause an outage for the working path. As we explained, only half of the cases in this category result in an outage for the network and the lower bound of unavailability due to dual failure becomes:



(a) No failure (b) After first failure (c) After second failure

Figure 4.11: Dual failure leading an outage for a partially on-cycle working path ($C-D-E-F$, path of concern), Category-8

$$U_{Category-7-Lower} = \frac{1}{2}|OS_r^p|.|O_{\bar{r}}^p|.U^2 \tag{4.12}$$

Category-8 : The second category of dual failures in this example which may result in an outage for a partially on-cycle working path is category 8 in which one of the spans belongs to OS_r^p and the other spans belongs to $S_r^{\bar{p}}$. Here, the first failure occurring on one of the spans of a totally straddling path (i.e, $(C-H)$, $(H-I)$ or $(I-F)$ Figure 4.11) would result in using the spare capacity in both halves of the FIPP p-cycle to protect the traffic on the totally straddling path (i.e, $F-A-B-C$ and $C-D-G-F$). Hence the second failure on the straddling spans of a partially on-cycle working path, the path of concern, (i.e, $(D-E)$ or $(E-F)$) will result in an outage as the FIPP p-cycle would already be pre-occupied (as shown in Figure 4.11). If the failures happen in reverse order, the failed straddling spans of the partially on-cycle working path ($D-E$ or $E-F$) would be protected by the FIPP p-cycle and the second failure on the straddling path will not affect the back-up path $(F-A-B-C)$, hence there would be no service outage (for the path of interest). It should be noted that the working capacity in both directions of the straddling path is stored by the FIPP p-cycle (here we are assuming fully loaded FIPP p-cycle).

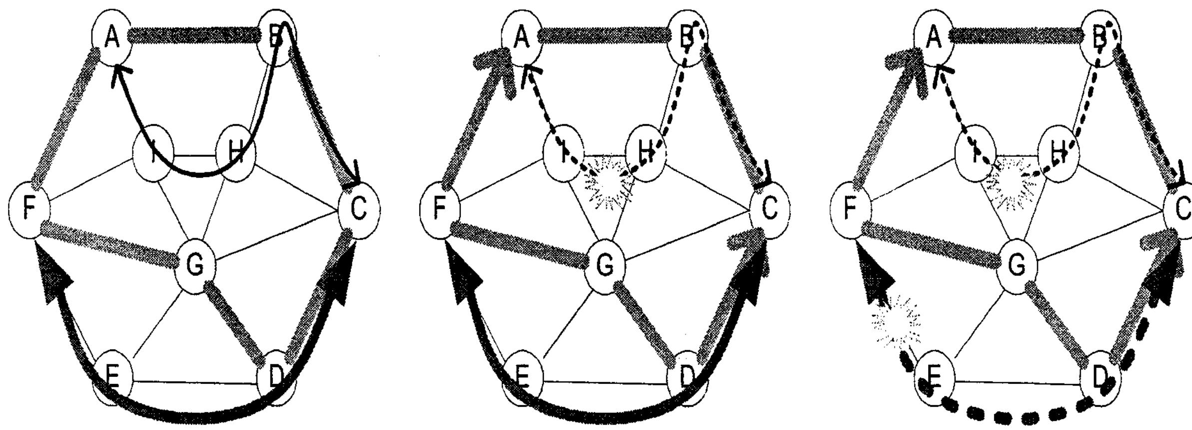


Figure 4.12: Dual failure leading an outage for a partially on-cycle working path ($C-D-E-F$, path of concern), Category-9

Obviously, the possible number of such dual dual-failure combinations is $|OS_r^p|.|S_{\bar{r}}^p|$, which leads the unavailability contribution due to failures in this category to be:

$$U_{Category-8} = \frac{1}{2}|OS_r^p|.|S_{\bar{r}}^p|.U^2 \tag{4.13}$$

Category-9 : In the last category of this example, we are dealing with combinations of dual failure, that may result in an outage in which both paths are partially on-cycle (Figure 4.12, 4.13). This case is the most common category among the other categories. Figure 4.12 shows this case, in which a span belonging to the straddling part of the working path (i.e, $(D-E)$ or $(E-F)$) fails and another span belonging to the straddling part of another partially on-cycle path (i.e, $(B-H)$, $(H-I)$ or $(I-A)$) fails. Here, there is 50% chance for the first failure to occur on a partially on-cycle path $(A-I-H-B-C)$ which is not our concerned working path. Since the FIPP p-cycle is occupied by the first failure (i.e, back-up path is $A-F-G-D-C$) , the second failure on the partially on-cycle working path $(C-D-E-F)$ is guaranteed to cause a service outage along the working path. If the two failures occur in the reverse order, there would be no service outage. Hence we could formally denote the upper bound of the unavailability due to a dual failure in this category as:

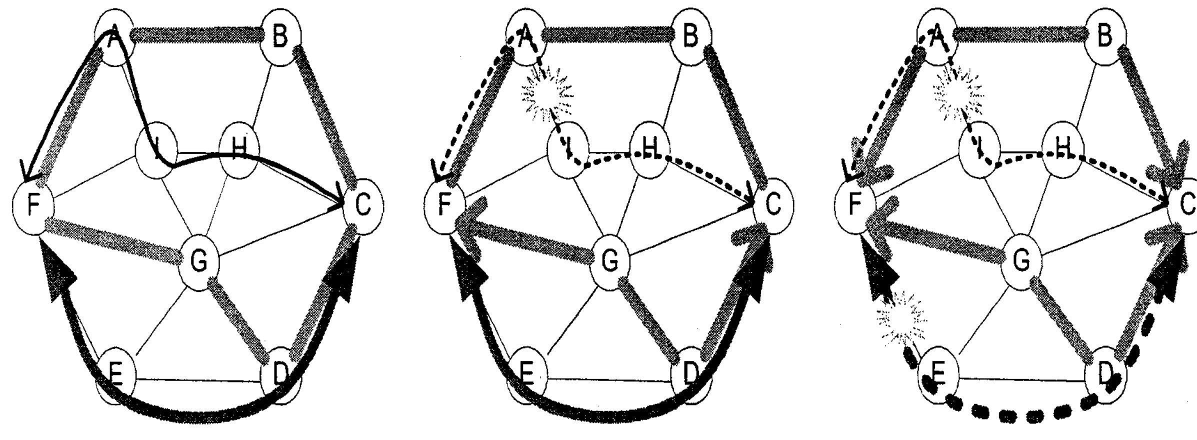


Figure 4.13: Special dual failure without any outage for a partially on-cycle working path ($C - D - E - F$, path of concern), Category-9

$$U_{Category-9-Upper} = \frac{1}{2}|OS_r^p|.|OS_{\bar{r}}^p|.U^2 \quad (4.14)$$

Alternatively, we can have some specific combinations of dual failure, irrespective of the order in which the failures occur, where there would be no service outage. Figure 4.13 presents an example of this kind of dual failure, If the first failure occurs on the straddling part of the path (($C - H$), ($H - I$) and ($I - A$)) and the default protection switching preplan is $A - G - D - C$, then the FIPP p-cycle could provide a protection path for the second failure on the partially on-cycle working path ($F - A - B - C$). Hence, the lower bound of the unavailability due to dual failure, in this category is:

$$U_{Category-9-Lower} = 0 \quad (4.15)$$

Without much elaboration it can be stated that the upper and lower bound for total unavailability of a working path can be given by the sum of the contribution towards the upper and lower bound of unavailability by each category. Formally, we have:

$$UU_{working-path} = |O^p_r|.|O^p_{\overline{r}}|.U^2 + \frac{1}{2}|O^p_r|.|S^p_{\overline{r}}|.U^2 + \frac{1}{2}|O^p_r|.|OS^p_{\overline{r}}|.U^2 + \frac{3}{4}|S^p_r|.|O^p_{\overline{r}}|.U^2 + \frac{1}{2}|S^p_r|.|S^p_{\overline{r}}|.U^2 + \frac{1}{2}|S^p_r|.|OS^p_{\overline{r}}|.U^2 + |OS^p_r|.|O^p_{\overline{r}}|.U^2 + \frac{1}{2}|OS^p_r|.|S^p_{\overline{r}}|.U^2 + \frac{1}{2}|OS^p_r|.|OS^p_{\overline{r}}|.U^2 \tag{4.16}$$

$$LU_{working-path} = |O^p_r.O^p_{\overline{r}}|.U^2 + \frac{1}{2}|O^p_r.S^p_{\overline{r}}|.U^2 + \frac{3}{4}|S^p_r.O^p_{\overline{r}}|.U^2 + \frac{1}{2}|S^p_r|.|S^p_{\overline{r}}|.U^2 + \frac{1}{4}|S^p_r|.|OS^p_{\overline{r}}|.U^2 + \frac{1}{2}|OS^p_r|.|O^p_{\overline{r}}|.U^2 + \frac{1}{2}|OS^p_r|.|S^p_{\overline{r}}|.U^2 \tag{4.17}$$

4.3 Enhanced Availability

In this section, we propose a formulation for capacity placement of demands to enhance availability. This formulation optimizes the allocation of spare capacity in order to find the minimal cost capacity placement that allows us not only to guarantee that every working path is protected against single span failure but also to ensure that the unavailability of all working paths is less than the desired upper bound. The routing of demands and finding working paths are done ahead of the placement of FIPP p-cycle for protection of them. The routing of demand could be done using any load balancing routing algorithm or any other suitable algorithms and we used K-shortest path algorithm to find all working paths. All the working paths are provided as inputs for the Integer Linear Program (ILP). Thus, the optimization is a non-joint optimization problem. In the simulation, we restrict our consideration to bi-directional FIPP p-cycle; also we assume networks with full wavelength conversion or similar characteristic.

4.3.1 FIPP p-Cycle Design

Allocating the eligible FIPP p-cycle to sets of mutually disjoint paths is defined as the FIPP p-cycle network design problem. The FIPP p-cycle network design could be done in two different approaches [10]. The first approach can be stated as: "Try to find out sets of paths which are all mutually disjoint. Then, form a FIPP p-cycle by routing the end nodes of these paths." Finding all possible sets of disjoint paths is the main problem in this approach. If we consider that our network has n nodes, $n(n-1)/2$ will be the number of end nodes pairs. We have to find out all possible combination of 2-route sets, 3-route sets and so on. It is an intractable problem, of $O(n^2!)$, just to find our all possible combination of paths which are all mutually disjoint and this is only if each demand between end nodes uses a single working path [21]. The second approach can be stated as: "Choose one of eligible FIPP p-cycle, find out sets of paths whose end nodes are on the FIPP p-cycle but they are mutually disjoint." Since this approach can be applied to our methods for finding eligible sets of FIPP p-cycles, it is practical for our availability problem. Hence in the rest of the chapter, we will apply the second approach.

4.3.2 Notations

The following notations are used for sets and parameters:

- **Input Parameters**

  S Set of spans.

  P Set of simple cycles eligible for allocation.

  R Set of working paths.

C_k Cost of a span k.

π^p_k On-cycle relation between a cycle (p) and a span (k). π^p_k is equal to '1' if cycle (p) crosses span (k), otherwise it is '0'.

δ^p_r Protection relation between an eligible cycle (p) and a working path (r). δ^p_r is equal to '1' if the working path (r)'s end nodes are in the cycle (p) and at least one of the spans in the working path (r) is on-cycle (p). δ^p_r is equal to '2' if the working path (r)'s end nodes are in the cycle (p) and the working path (r) straddles cycle (p) and '0' if the working path is not protected by the cycle (p).

ϕ^r_k Equal to '1' if span (k) is part of the working path (r) , otherwise it is '0'.

$\xi_{x,y}$ Equal to '1' if the working path (x) and the working path (y) are not disjoint, otherwise it is '0'.

MU Maximum Unavailability of any working path after the allocation of FIPP p-cycles.

- **Output Parameters**

S_k Number of spare unit placed on span k.

N^p_r Equal to '1' if FIPP (p) is allocated to protect the working path (r), '0' otherwise. This is an intermediate parameter to find the rest of parameters.

N^p Number of protection units per FIPP cycle (p).

O^p_r Number of on-cycle spans in FIPP (p) which are also on the working path (r).

$O^p_{\bar{r}}$ Number of on-cycle spans in FIPP (p) which are not on the working path (r).

S_r^p Number of spans that are on the working path (r) which is totally straddling the FIPP (p).

$S_{\bar{r}}^p$ Number of spans traversed by those paths which straddle the FIPP (p) and are not part of the concerned path (r).

OS_r^p Number of spans that are on the working path (r) which is partially on-cycle. These spans are not on-cycle.

$OS_{\bar{r}}^p$ Number of spans whose paths are partially on the FIPP (p) but they are neither on the working path (r) nor on the FIPP (p).

U_r^p Total end-to-end unavailability of working path (r) in the FIPP (p).

Note that in the previous section O_r^p, $O_{\bar{r}}^p$, S_r^p, $S_{\bar{r}}^p$, OS_r^p, and $OS_{\bar{r}}^p$ were defined as sets of spans not number of spans.

4.3.3 Formulation to Limit the Working Paths Unavailability

The objective is to minimize the total cost of spare capacity:

$$\min \sum_{\forall k \in S} C_k S_k \tag{4.18}$$

Constraint (4.19) given below ensures that at least one FIPP p-cycle will be assigned to protect the working path. This constraint finds the set of FIPP p-cycles that have to be placed to ensure that every working path (r) will be protected against single failures on every span along the working path. N_r^p contains information that a FIPP p-cycle is allocated for a certain working path or not.

$$\sum_{p \in P; \delta_r^p > 0} N_r^p \geq 1, r \in R. \tag{4.19}$$

By the intermediate output (N_r^p) , we can determine the variables of interest: O_r^p, $O_{\overline{r}}^p$, S_r^p, $S_{\overline{r}}^p$, OS_r^p, and $OS_{\overline{r}}^p$. Equations (4.20) to (4.25) calculate them.

$$O_r^p = \sum_{k \in S} N_r^p \phi_k^r \pi_k^p, r \in R; p \in P \tag{4.20}$$

$$O_{\overline{r}}^p = \sum_{k \in S} N_r^p (1 - \phi_k^r) \pi_k^p, r \in R; p \in P \tag{4.21}$$

$$S_r^p = \sum_{k \in S; \delta_k^p = 2} N_r^p \phi_k^r, r \in R; p \in P \tag{4.22}$$

$$S_{\overline{r}}^p = \sum_{k \in S; \delta_k^p = 2} N_r^p (1 - \phi_k^r), r \in R; p \in P \tag{4.23}$$

$$OS_r^p = \sum_{k \in S} N_r^p \phi_k^r (1 - \pi_k^p), r \in R; p \in P \tag{4.24}$$

$$OS_{\overline{r}}^p = \sum_{k \in S} N_r^p (1 - \phi_k^r)(1 - \pi_k^p), r \in R; p \in P \tag{4.25}$$

Now we can compute the upper and lower bound for U_r^p, which is the total end-to-end unavailability of the working path (r) protected by FIPP p-cycle (p). This could be done by using the output of equations (4.20) to (4.25) and using equation (4.16) for upper bound and equation (4.17) for lower bound as follow:

$$\begin{aligned} UU_r^p = \{ & O_r^p . O_{\overline{r}}^p + \frac{1}{2} O_r^p . S_{\overline{r}}^p + \frac{1}{2} O_r^p . OS_{\overline{r}}^p + \\ & \frac{3}{4} S_r^p . O_{\overline{r}}^p + \frac{1}{2} S_r^p . S_{\overline{r}}^p + \frac{1}{2} S_r^p OS_{\overline{r}}^p \\ & + OS_r^p . O_{\overline{r}}^p + \frac{1}{2} OS_r^p . S_{\overline{r}}^p + \\ & \frac{1}{2} OS_r^p . OS_{\overline{r}}^p \} . U^2, \\ & r \in R; p \in P \end{aligned} \tag{4.26}$$

$$
\begin{aligned}
LU_r^p = \{O_r^p.O_{\overline{r}}^p + \frac{1}{2}O_r^p.S_{\overline{r}}^p + \frac{3}{4}S_r^p.O_{\overline{r}}^p + \\
\frac{1}{2}S_r^p.S_{\overline{r}}^p + \frac{1}{4}S_r^p.OS_{\overline{r}}^p + \frac{1}{2}OS_r^p.O_{\overline{r}}^p + \\
\frac{1}{2}OS_r^p.S_{\overline{r}}^p\}.U^2, \\
r \in R; p \in P
\end{aligned}
\tag{4.27}
$$

Next, we can compute the lower and upper bounds for U_r^p (LU_r^p and UU_r^p) using (4.16), (4.17) and making use of the the output of (4.20)−(4.25) and we constrain the unavailability of the working path to a desired upper limit which is our main objective. The upper limit is an input parameter to our ILP. ILP will allocate FIPP p-cycles such that this constraint is satisfied. That is, for a lower value of the constraint, FIPP p-cycle with more hop-count will be allocated and for a more relaxed value of the constraint the optimizer will tend to allocated smaller FIPP p-cycles. Obviously if the desired value for unavailability is too low, a solution may not exist.

$$
LU_r^p, UU_r^p \leq MU, r \in R; p \in P \tag{4.28}
$$

Constraint (4.29) ensures the FIPP specification which any FIPP p-cycle only protects a set of mutually disjoint working paths. If working paths x and y are not disjoint ($\xi_{x,y} = 1$), then only one of them can be protected by FIPP p-cycle p. It means that only one or none of N_x^p or N_y^P can be 1 at the same time.

$$
\xi_{x,y} + N_x^p + N_y^P \leq 2, (x, y) \in R^2 | x \neq y; p \in P \tag{4.29}
$$

Equation (4.30) given below finds the number of unit protection capacity allocated on the FIPP p-cycle (p). The right hand side of the equation is the number of instances of FIPP p-cycle (p) required by any span (k) so as to protect the working path which traverses k. The right hand side of the equation is not necessarily an integer value. For instance if a FIPP p-cycle protects a certain working path as a totally straddling working path for 5 unit of working capacity, then the number of instances of the FIPP p-cycles required by that working path would be calculated as 2.5, as each instances of FIPP p-cycle can provide 2 unit of protection of totally straddling working path. However N^p is an integer value, thus it is more than or equal to right hand side of the equation.

$$N^p \geq \sum_{r \in R; \delta_r^p = 2} \frac{1}{2} N_r^p \phi_k^r + \sum_{r \in R; \delta_r^p = 1} N_r^p \phi_k^r, k \in S \tag{4.30}$$

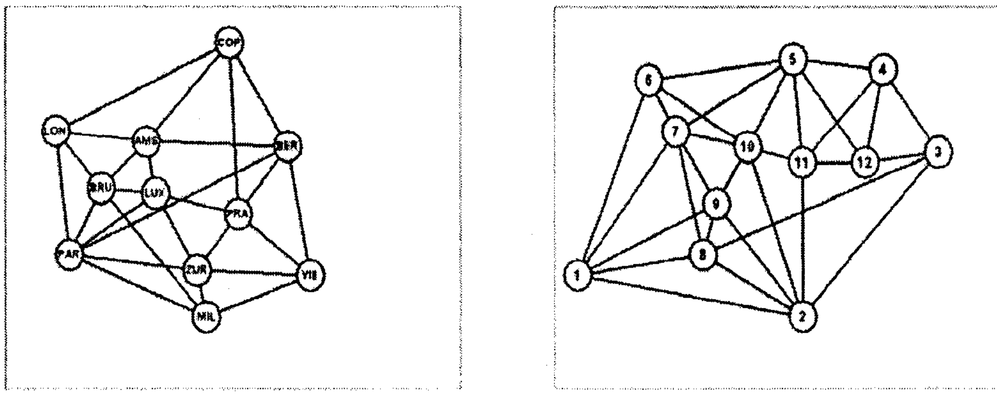
The total spare capacity in each span is given by equation (4.31) and this spare capacity is optimized as per objective (4.18).

$$S_k = \sum_{p \in P} N^p \pi_k^p, k \in S; p \in P \tag{4.31}$$

Note that the more nodes in the network, the more candidate cycles we have especially in dense network. Finding an optimal set of cycles using the ILP formulation presented previously can be shown to be an NP-hard problem [47] though limiting the number of candidate FIPP p-cycle can reduce the computation time by compromising the optimality.

We have to note that expressions (4.26) and (4.27) do not make the model non-linear as a result of the multiplication of expressions (4.20) to (4.25). For instance to find unavailability for Category-1, we have:

$$O_r^p = \sum_{k \in S} N_r^p \phi_k^r \pi_k^p = N_r^p \cdot \sum_{k \in S} \phi_k^r \pi_k^p, \forall r \in R; p \in P \tag{4.32}$$



(a) COST239 network (b) 12n30s network

Figure 4.14: Test Networks

$$O^p_{\bar{r}} = \sum_{k \in S} N^p_r (1 - \phi^r_k)\pi^p_k = N^p_r \cdot \sum_{k \in S}(1 - \phi^r_k)\pi^p_k, \forall r \in R; p \in P \tag{4.33}$$

$$\begin{aligned} O^p_r \cdot O^p_{\bar{r}} = (N^p_r \cdot \sum_{k \in S} \phi^r_k \pi^p_k) \cdot (N^p_r \cdot \sum_{k \in S}(1 - \phi^r_k)\pi^p_k) = \\ (N^p_r \cdot N^p_r) \cdot (\sum_{k \in S} \phi^r_k \cdot \sum_{k \in S}(1 - \phi^r_k)\pi^p_k) \qquad \forall r \in R; p \in P \end{aligned} \tag{4.34}$$

Since variable N^p_r is a binary variable, we can replace the quadratic term of $N^p_r \cdot N^p_r$ by:

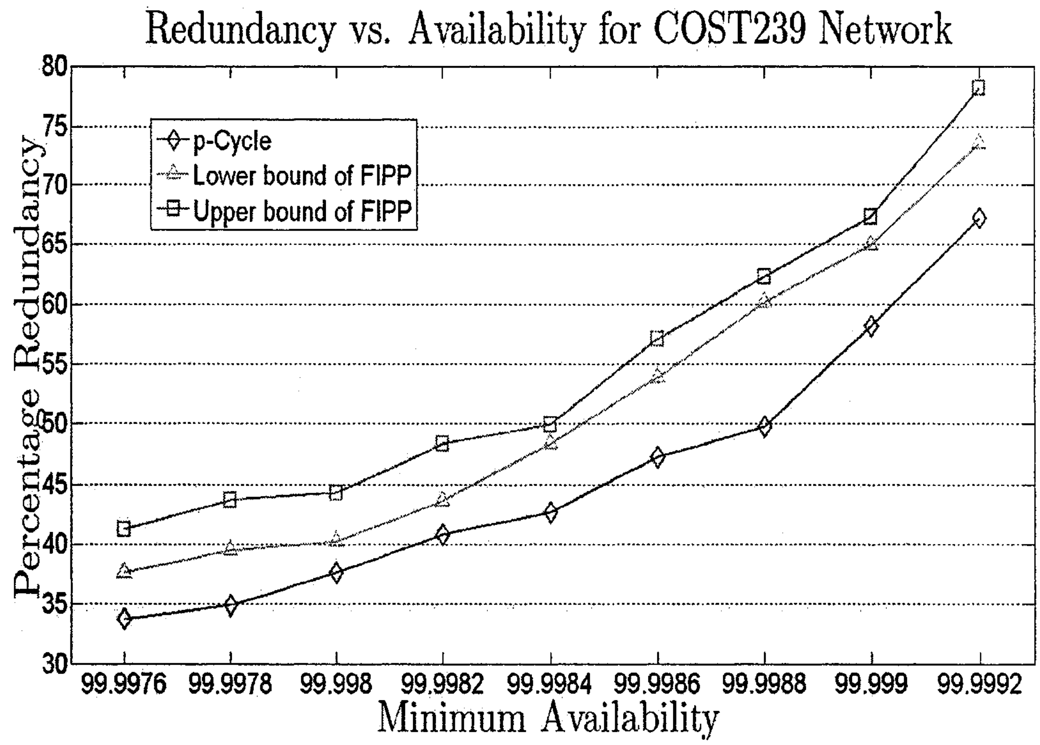
$$N^p_r \cdot N^p_r = N^p_r \qquad \forall r \in R; p \in P \tag{4.35}$$

4.4 Numerical Results

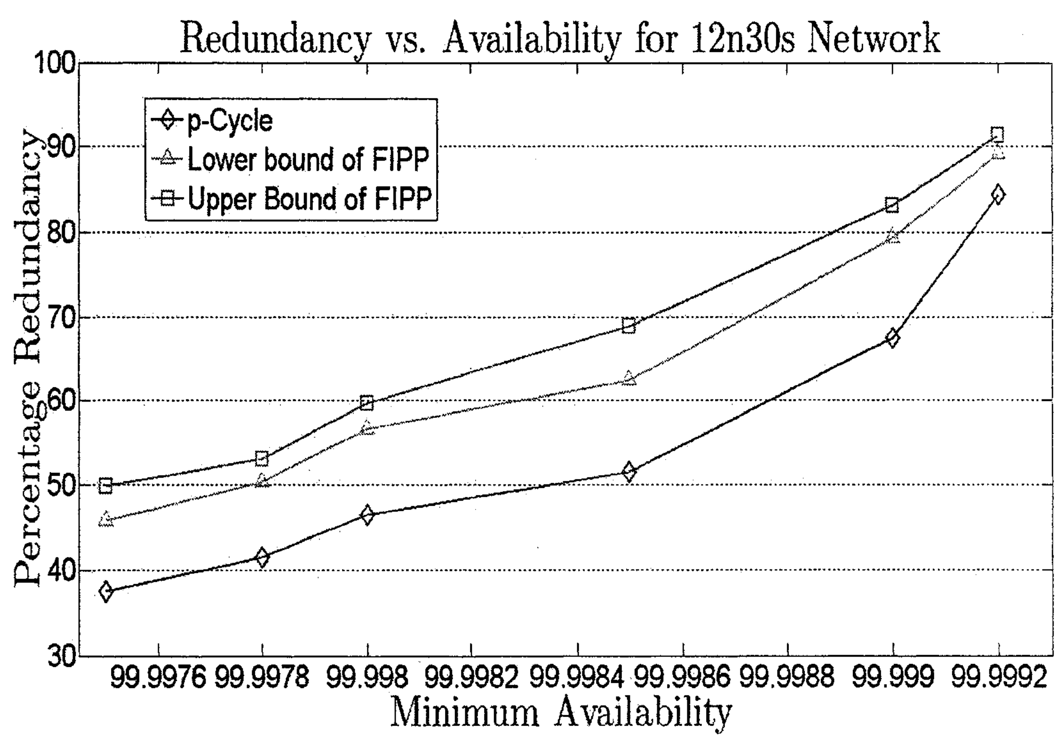
The formulation provided in the previous section was tested with the $COST239$ and the $12n30s$ networks. These networks are shown in Figure 4.14. We assume that each span has enough capacity to support the protection capacity required by the optimal solution; we also assume that each node can perform full wavelength conversion. The formulation is implemented as an AMPL model and solved by using the solver *CPLEX*9.1.3 [48]. A separate Java program is used to find the input parameters for the AMPL model. In our implementation, the demands between any two pair of nodes are symmetric; it means that the number of lightpaths in either direction between a source and destination pair are equal. The demands are randomly assign to a certain number of light paths ranging from 0 to 10. The demands are routed using a Dijkistra shortest path algorithm to find eligible working routes. Lightpaths between pair of nodes are routed individually; hence two lightpaths between the same pair of nodes can possibly have different working routes. The preprocessing Java program finds candidates FIPP p-cycles using a depth-first search and pre-selected by hop count lengths. All working routes and candidates FIPP p-cycle are provided as inputs to the AMPL model.

We show in Figure 4.15 the redundancy obtained for certain level of availability for both COST239 and 12n30s networks and compare the results obtained for upper and lower bounds of redundancies in FIPP p-cycle protecting method with the same results for span p-cycle [1]. Here, the ratio of the sum of the spare capacity in each span to sum of the working capacity in each span is defined as the redundancy. As mentioned in the previous section, we assume the unavailability of each span to be equal to 10^{-3} [1]. The same set of demands is used to find the redundancy for each availability requirement.

Obviously, these results indicate that in order to achieve the same service availability the redundancy for FIPP p-cycle is larger than span p-cycle; however, similar to span p-cycle, the availability increases at the cost of additional spare capacity requirement. As shown in Figure 4.15, constraining the lower bound for the unavailability of FIPP p-cycle yields more redundancy than span p-cycle. In order to better



(a)



(b)

Figure 4.15: Comparison between FIPP p-cycle and span p-cycle [1] according to redundancy versus minimum availability for (a) COST239 network, (b) 12n30s network

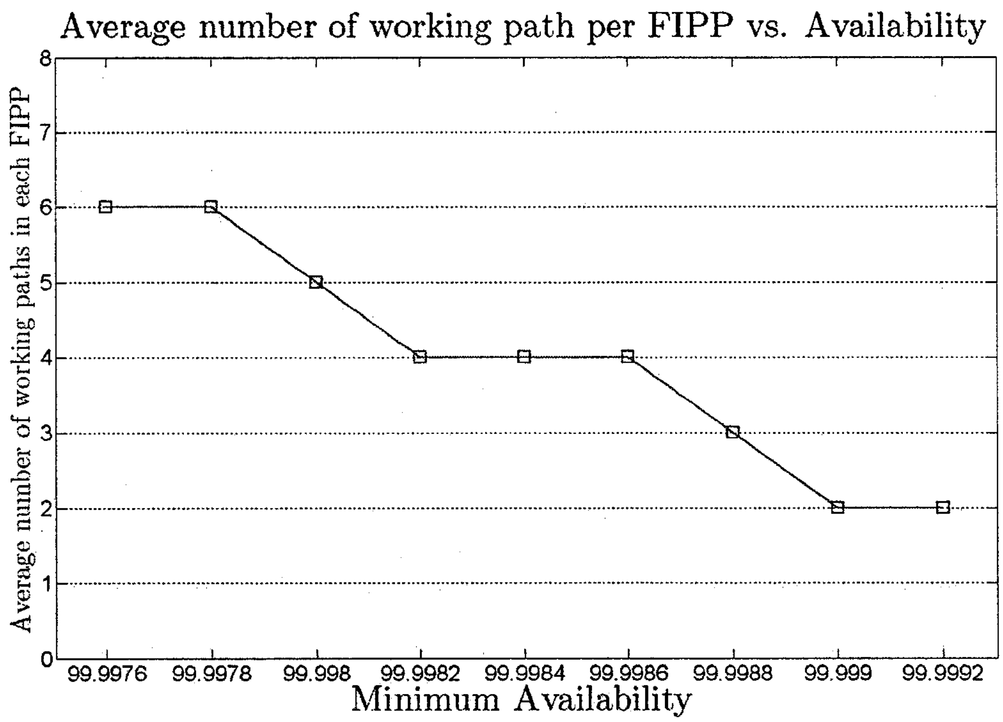


Figure 4.16: Average number of mutual disjoint working paths protected by same FIPP p-cycle versus minimum availability for COST239 network

understand these findings[2], we tried to find out the redundancy and availability of both methods without having any limitation on availability. We provided the same set of demands as an input for AMPL and solved this new problem (without constraint (4.28)) and then found out the minimum, maximum and average availabilities for both FIPP p-cycle and span p-cycle (the results are shown in Table 4.1). Our results show that the average availability of FIPP p-cycle-based network is 99.9963 (i.e., an equivalent of 20.49 min/year unavailability), however the average availability of span p-cycle-based network is 99.9987 (i.e an equivalent of 6.83 min/year unavailability). This implies that with the FIPP p-cycle method the probability of the system failure is around *three times more* than span p-cycle method in one year. Alternatively, these results show that a span p-cycle has more redundancy (i.e, less efficiency) than FIPP p-cycle (similar to [10]) but the average availability for span

[2]The authors of [39] initially showed that FIPP has much better efficiency than p-cycle methods with respect to spare capacity redundancy.

p-cycle is more than that of the FIPP p-cycle method.

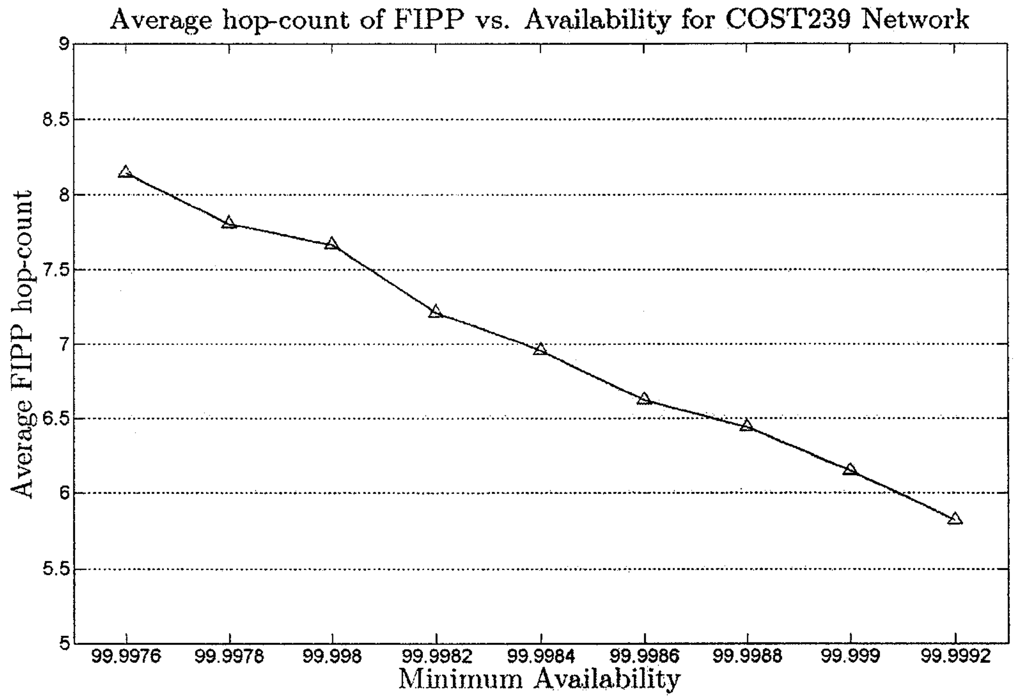
Table 4.1: Comparison between FIPP p-cycle and p-cycle [1] for COST239 network without having any limitation on availability

	FIPP p-cycle	p-cycle
Percentage Redundancy	19.24	27.22
Minimum Availability	99.9945	99.9942
Maximum Availability	99.9986	99.9996
Average Availability	99.9963	99.9987
Average Down-Time	20.49 min/year	6.83 min/year

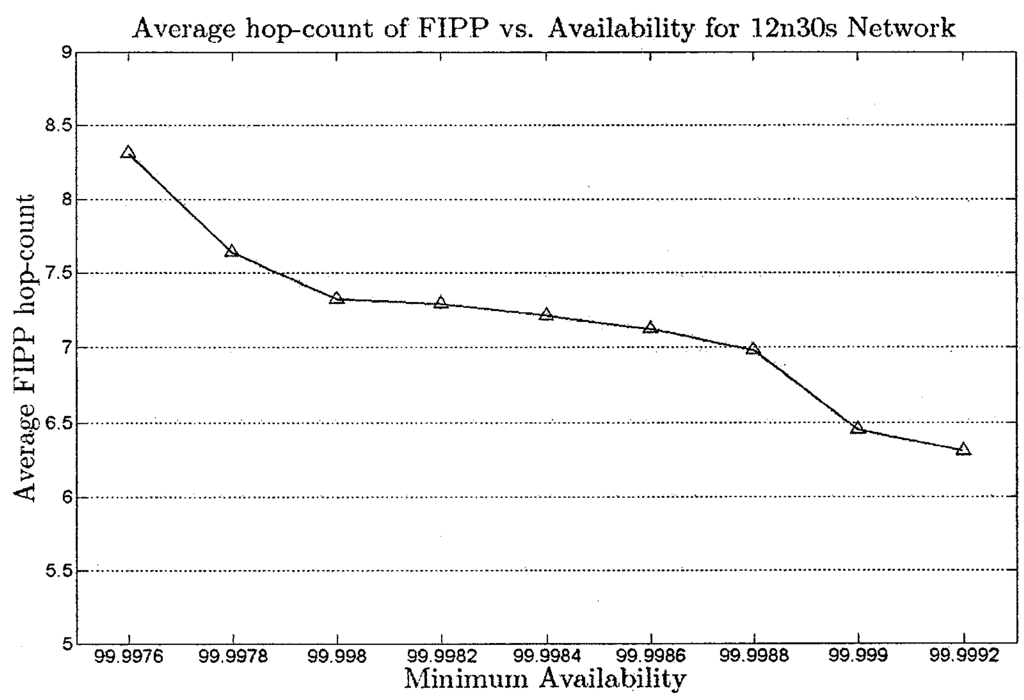
Clearly, when our design is based on limiting the availability, in order for FIPP p-cycle method to achieve the same level of availability as span p-cycle method, it will sacrifice its resource efficiency by creating more redundancies. We observed that by adding constraints on the availability, the FIPP p-cycle method will be forced to (1) protect fewer demands and (2) select FIPP cycles with smaller hop count. Unfortunately, both of these properties are the main reason why FIPP was originally more advantageous than the span p-cycle method. Figure 4.16 shows the average number of demands protected by the same FIPP p-cycle; obviously, the higher the required availability in our design, the smaller is the number of the demands a FIPP p-cycle can protect. For instance, if the required availability is 99.999, the number of demands sharing a FIPP is only 2, as opposed to 6 when the required availability is 99.9976.

Next, we tried to obtain some insights on the impact of the length of FIPP p-cycle (in terms of hop counts) on the service availability. We show in Figure 4.17 the average hop count for a FIPP p-cycle for different availability requirements. Clearly, the higher the availability, the shorter the FIPP becomes; shorter FIPP cycles tend to limit the sharing of spare capacity more than longer FIPP cycles and hence the higher the resource redundancy (as shown in Figure 4.18).

To further get additional insights on the higher redundancy obtained using FIPP p-cycles in our design method, we measure the individual availabilities obtained from



(a)



(b)

Figure 4.17: Average FIPP p-cycle hop count for different working path length for (a) COST239 network, (b) 12n30s network

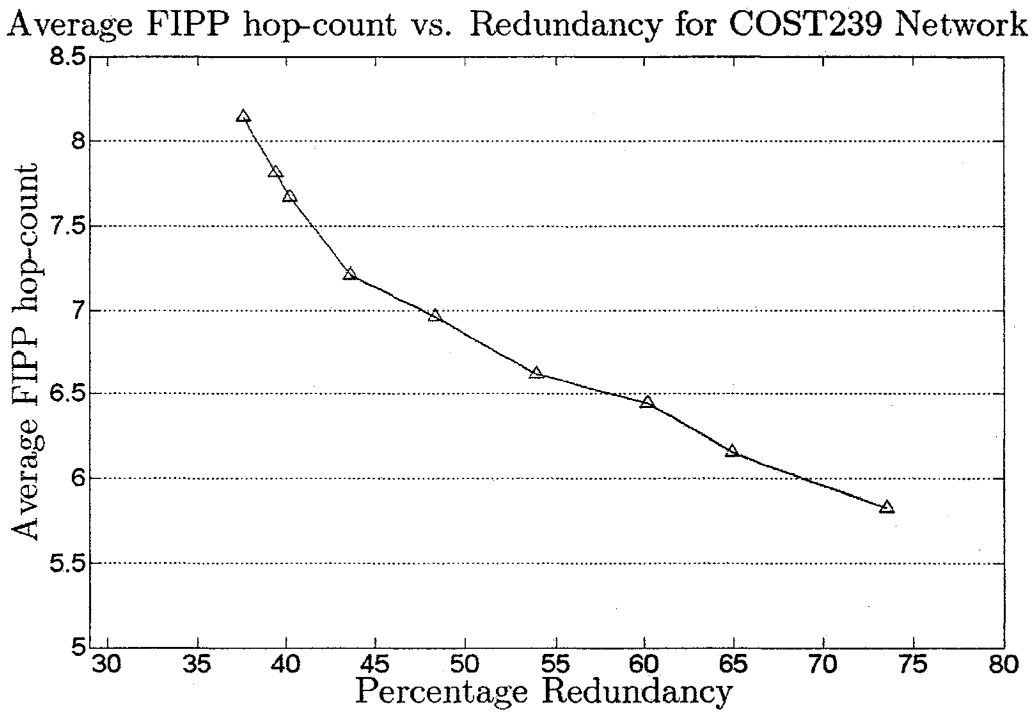
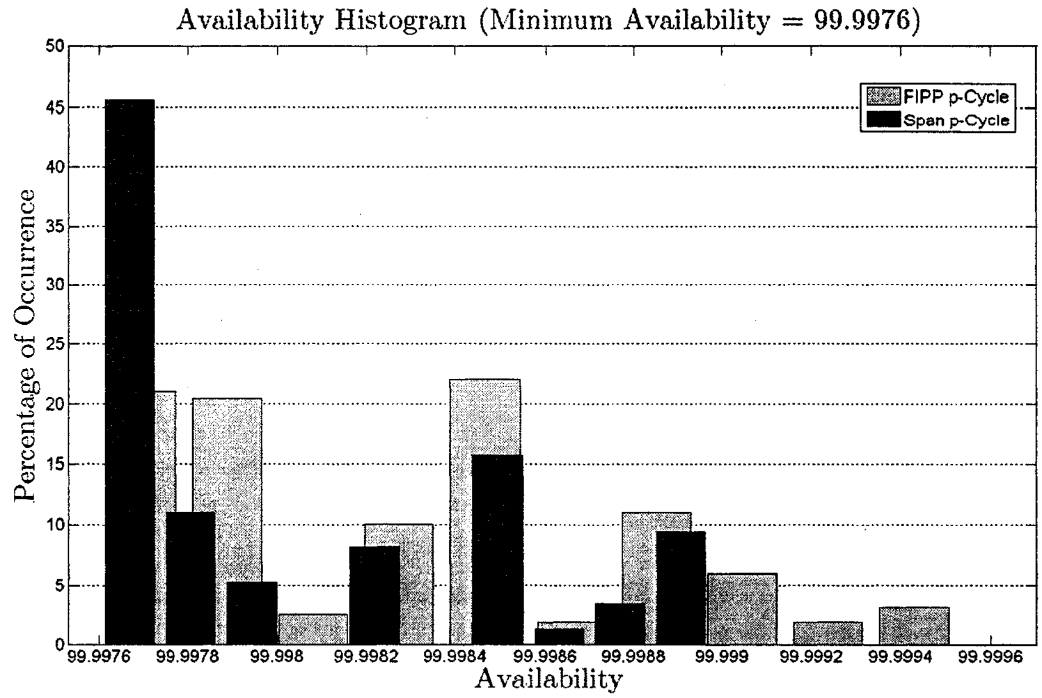
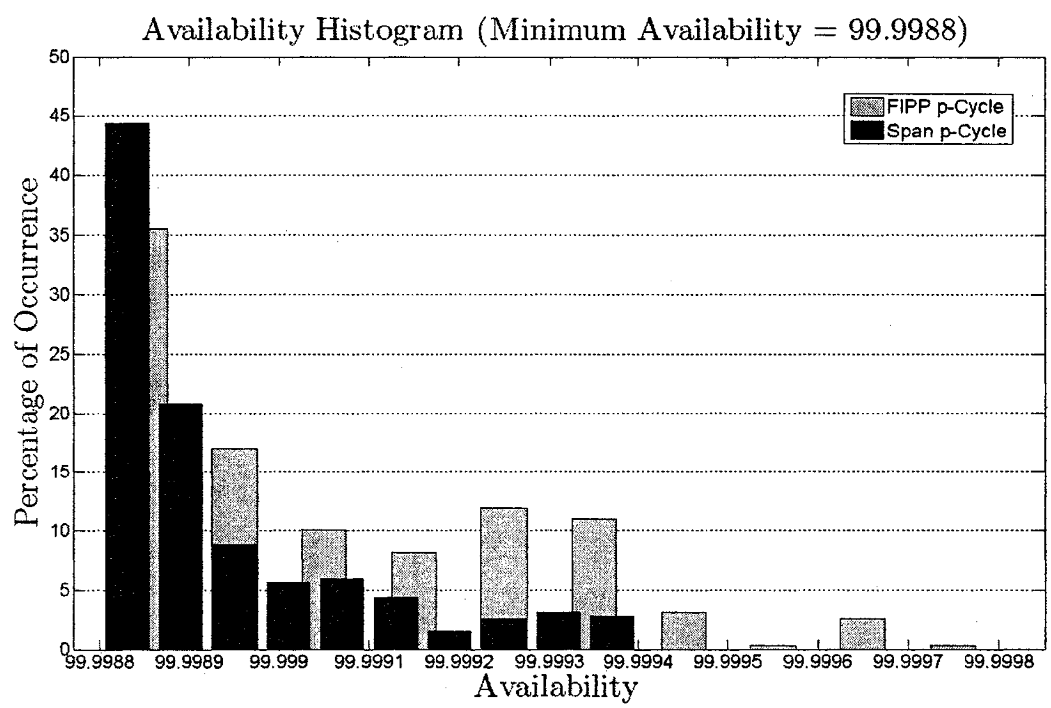


Figure 4.18: Average FIPP p-cycle hop-count (length) versus redundancy for COST239 Network.

the solution for all protected demands, using both span p-cycle and FIPP p-cycle. Figure 4.19 shows the distribution of the availability of the protected demands for different minimum availability requirement in our design (a-d). Indeed, the figure shows that while most of the demands protected with span p-cycles have their availabilities close to the desired minimum required availability, demands protected using FIPP p-cycles tend to have their availabilities more spread (above the minimum required availability); that is, more demands in the latter scheme have their availabilities higher than in the former scheme. For instance, if the desired level of availability is 99.9976, only 21.06% of working paths have the exact level of availability in FIPP p-cycle as opposed to 45.59% for span p-cycle method and 78.94% of the demands have their availabilities higher than 99.9976 as opposed to 54.41% of demands protected by span p-cycles. Indeed this higher availability requires more spare capacity, which explains the higher redundancy of FIPP p-cycle based design method.



(a) MA = 99.9976



(b) MA = 99.9988

Figure 4.19: Availability histogram for different level of availability in COST239 network for span p-cycle [1] and FIPP p-cycle with different values for Minimum availability (MA).

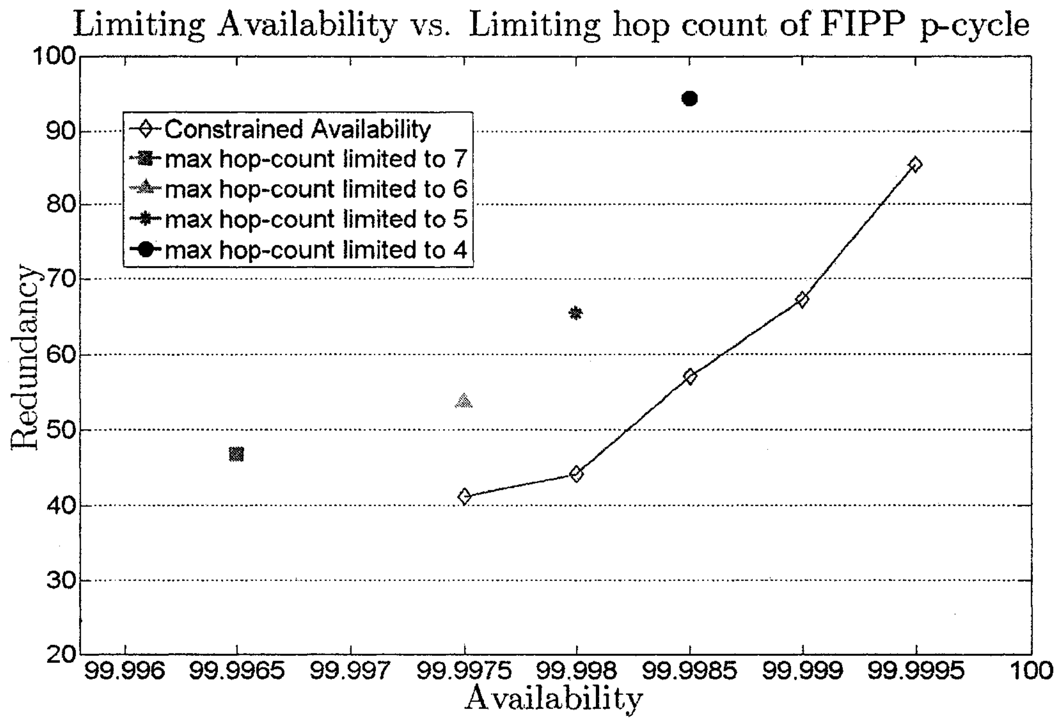


Figure 4.20: Directly limiting unavailability vs. limiting hop-count of candidate FIPP p-cycles for COST239 network.

Finally, we simulate our ILP without considering the desired upper bound limit on unavailability, given by equation (4.28), but we constrained the length of eligible FIPP p-cycle by hop-count. Since the eligible FIPP p-cycle are pre-selected based on the hop-count, we try to find out the working path with the minimum availability (i.e., worst-case). In Figure 4.20, we compare the upper-bound-availability of any working path in the COST239 network when the unavailability is limited directly by constraint (4.28) and the length of the FIPP p-cycle is not restricted, with the worst case availability of any working path when the FIPP p-cycle hop-count is limited. Obviously, directly limiting the unavailability with certain upper limit proves to be a better design option than restricting the length of the eligible FIPP p-cycles (based on the hop-count). As the figure shows, this method achieves a better spare capacity redundancy for the same availability or better availability for almost the same redundancy.

4.5 Conclusion

In this chapter we studied the relationship between the unavailability of a working path and the topology of the FIPP p-cycles, which is allocated for restoration of the working paths in mesh networks with FIPP p-cycle based protection method. We then presented an availability-aware design method for networks protected by FIPP p-cycles. Our results showed that the FIPP protection method requires more network capacity (8-13%) to obtain the same level of availability that basic p-cycle method achieves. We observe that by adding limitations on the availability, the FIPP method will be forced to protect fewer demands and also select FIPP cycles with smaller hop-count.

As a topic for future study, allocating the FIPP based on availability can be done by finding all possible combinations of mutual disjoint working paths, hence providing a heuristic for finding a subset of these combinations based on different levels of availability will be important.

Chapter 5

Conclusion and Future Directions

5.1 Conclusion

Survivability of high-capacity optical wavelength-division multiplexing (WDM) mesh networks has received much research attention for many years now. These networks are typically designed to survive single component failures. The method of pre-configured protection cycles (p-Cycles), proposed by W. Grover's research group [9], promises to achieve ring-like high speed protection with mesh-like high efficiency in use of spare capacity. In such networks, which are designed to withstand only single failures, service availability comes to depend on dual-failure (or more) considerations. Hence, availability-aware service provisioning emerged as a topic of great importance in the past few years.

In this thesis, we proposed an exact model for the availability-aware design for p-Cycle based network by overcoming some flaws in availability analysis of span p-Cycles in previous works. Our method provided a technique for allocating p-Cycles to restore single link failures such that the unavailability of all the demands in the network is bounded by an upper limit. We then provided some heuristic for restricting the number of variables and constraints in an Integer Linear Programming resultant formulation for our method in order to solve our problem in reasonable amount of time. Our results showed that by adding limitation on the availability, the model will utilize more number of p-Cycles with smaller hop-counts.

We also studied the relationship between the unavailability of a working path and the topology of the FIPP p-Cycle, which is allocated for restoration of the working paths in mesh networks with FIPP p-Cycle based protection method. We presented an availability-aware design method for networks protected by FIPP p-Cycles. Then, we compared FIPP protection technique with conventional p-Cycle method based on level availability of all working paths. Our investigation showed that the FIPP protection method requires more network capacity (8-13%) to obtain the same level of availability that basic p-Cycle method achieves.

5.2 Future Directions

The work presented in this thesis provided considerable benefits in availability enhancement for optical network based on p-Cycles. However, there are still several future directions that can provide additional benefits.

As a topic for future study, allocating the FIPP cycles based on availability can be done by finding all possible combinations of mutual disjoint working paths, hence providing a heuristic for finding a subset of these combinations based on different levels of availability will be important. In addition, in most of the available literature, for the design of optimal networks, only a subset of candidate p-Cycles is considered. Therefore, the global optimal solution can not be guaranteed. Using decomposition techniques for solving large ILP problems, particularly Column Generation Algorithm, large instances of network designs can be solvable.

Another future direction is to find the exact optimal solution of network designs by using CG. The objective can be the design of availability-aware span and FIPP

p-Cycle networks with joint and non-joint working and spare capacities. In each case the current models have to be rewritten in order to match the CG algorithm. The main problem with CG algorithm is the huge number of variables in pricing problem and then finding the ILP solution from non-integer solutions.

Another interesting problem which needs to be studied is the protection of multicast sessions by p-Cycles. Substantial studies have been carried out for the protection of node-pair demands, however the protection of multicast sessions has not been considered reasonably. Then, the availability analysis of joint optimization problem in multicast networks should be considered with due diligence.

Usually, it is assumed that the demands are bidirectional, whereas in practice the required bandwidth can be very different on the two sides of the connection, e.g. in upstream and downstream connections. Therefore, another future work would be to consider asymmetric traffic. In addition, further demands in a network can be classified into different priority levels, where each priority level would have a different upper bound for the end-to-end unavailability, thus providing different classes of protection and availability to various classes of services is another topic of future study.

Bibliography

[1] D. S. Mukherjee, C. Assi, and A. Agrawal, "An alternative approach for enhanced availability analysis and design methods in p-cycle-based networks," *IEEE Journal on Selected Areas in Communacation, vol. 24, no 12, December 2006.*

[2] W. D. Grover, *Mesh-Based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking.* Upper Saddle River, NJ:: Prentice-Hall PTR, 2003.

[3] O. Gerstel and R. Ramaswami, "Optical layer survivability: a service perspective," *IEEE Communication Magazine, vol. 3, no. 38, March 2000, pp. 104-113.*

[4] D. S. Mukherjee, "Optical network survivability using p-cycle," *M. Sc. Thesis, Concordia University, 2006.*

[5] B. Mukherjee, *Optical WDM Networks.* Springer, Optical networks series, 2006.

[6] R. Ramaswami and K. N. Sivarajan, *Optical Networks.* Morgan Kaufmann Publishers, 2002.

[7] "Defence research and development canada," *http://www.css.drdc-rddc.gc.ca/.*

[8] M. Clouqueur and W. D. Grover, "Availability analysis of span-restorable mesh networks," *IEEE J. Sel. Areas Commun., special issue on recent advances in fundamentals of network management, vol. 20, no. 4, pp. 810-821, May 2002.*

[9] W. D. Grover and D. Stamatelakis, "Cycle-oriented distributed preconfiguration: Ring-like speed with mesh-like capacity for self-planning network restoration," *in Proc. IEEE International Conference on Communications (ICC) 1998, pp. 537-543.*

[10] A. Kodian and W. Grover, "Failure-independent path-protecting p-cycles: Efficient and simple fully preconnected optical-path protection," *Journal of lightwave technology, vol. 23, no. 10, October 2005.*

[11] M. Clouqueur and W. D. Grover, "Availability analysis and enhanced availability design in p-cycle-based networks," *Photonic Network Communications, Springer Science, vol. 10,no. 1, pp. 55-71, 2005.*

[12] M. S. Kiaei, C. Assi, and B. Jaumard, "A survey on the p-cycle protection method for optical mesh networks," *IEEE Communications surveys and tutorials.*

[13] D. Zhou and S. Subramaniam, "Survivability in optical networks," *IEEE Network, Vol. 14, No. 6 (Nov/Dec 2000), pp. 16-23.*

[14] S. Ramamurthy and B. B. Mukherjee, "Survivable wdm mesh networks, part i: protection," *in Proc. IEEE Infocom 1999, pp. 744-751.*

[15] T. E. Stern and K. Bala, *Multiwavelength optical networks layered approach.* Addison wesley Longman, Inc., 1999.

[16] S. Ramamurthy and B. B. Mukherjee, "Survivable wdm mesh networks, part ii: Restoration," *Proc. ICC, 1999, pp. 2023-2030.*

[17] V. Anand and C. Qiao, "Static versus dynamic establishment paths in wdm networks part i," *Proceedings of IEEE ICC'00, New Orleans, LA, June 2000, vol. 1, pp. 198-204.*

[18] W. Grover and D. Stamatelakis, "Bridging the ring-mesh dichotomy with p-cycles," *in Proc. Second International Workshop on the Design of Reliable Communication Networks (DRCN 2000), pp. 92-104.*

[19] D. A. Schupke, C. G. Gruber, and A. Autenrieth, "Optimal configuration of p-cycles in wdm networks," *in Proc. IEEE International Conference on Communications (ICC) 2002, pp. 2761-2765.*

[20] D. Stamatelakis and W. D. Grover, "Rapid span or node restoration in ip networks using virtual protection cycles," *in Proc. Third Canadian Conference on Broadband Research (CCBR 9), pp. 33-44.*

[21] A. Kodian and W. Grover, "A disjoint route-sets approach to desing of path-protecting p-cycle networks," *Proc. Workshop on Design of Reliable Communacation Networks (DRCN 2005), Ischia (Naples), Italy, pp. 231-238, 16-19 October 2005.*

[22] D. Stamatelakis, "Theory and algorithms for preconfiguration of spare capacity in mesh restorable networks," *M. Sc. Thesis, University of Alberta, Spring 1997.*

[23] R. Iraschko, M. H. MacGregor, and W. D. Grover, "Optimal capacity placement for path restoration in mesh survivable networks," *Proc. ICC96, pp. 1568-1574, 1996.*

[24] M. H. MacGregor, W. D. Grover, and K. Ryhorchuk, "Optimal spare capacity preconfiguration for faster restoration of mesh networks," *Journal of Network and Systems Management, vol. 5, no. 2, Jun. 1997, pp. 159-171.*

[25] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research.* Mc Graw Hill, 2005.

[26] V. Chvatal, *Linear Programming.* W. H. Freeman and Company, 1983.

[27] W. D. Grover and J. Doucette, "Advances in optical network design with p-cycles: Joint optimization and pre-selection of candidate p-cycles," *Proc. IEEE LEOS Summer Topicals 2002, Mont Tremblant, Qubec, Canada, 15-17 Jul. 2002, pp. 49-50.*

[28] B. Jaumard, C. Rocha, D. Baloukov, and W. Grover, "A column generation approach for design of networks using path-protecting p-cycles," *International Workshop on the Design of Reliable Communication Netowrks, France, October 2007.*

[29] M. E. Lubbecke and J. Desrosiers, "Selected topics in column generation."

[30] G. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Oper. Res., 8:101-111, 1960.*

[31] D. A. Schupke, "The tradeoff between the number of deployed pcycles and the survivability to dual fiber duct failures," *in Proc. IEEE International Conference on Communications (ICC) 2003, pp. 1428-1432.*

[32] D. Schupke, "Multiple failure survivability in wdm networks with *p*-cycles," *in Proc. IEEE International Symposium on Circuits and Systems (ISCAS) 2003, pp. 866-869.*

[33] I. Rados, P. Tarulija, and T. Sunaric, "Availability model of bidirectional line swithed ring," *in Proc. Transparant Optical Networks 2001, pp.312-316.*

[34] W. D. Grover, "Availability analysis and aps systems," *A monograph distributed at DRCN 2007, La Rochelle, France, October 7-11, 2007.*

[35] J. Zhanga, K. Zhu, H. Zang, N. S. Matloff, and B. Mukherjee, "Availability-aware provisioning strategies for differentiated protection services in wavelength-convertible wdm mesh networks," *IEEE/ACM Trans. Netw. 15(5): 1177-1190 (2007).*

[36] D. A. Schupke, W. D. Grover, and M. Clouqueur, "Strategies for enhanced dual failure restorability with static or reconfigurable *p*-cycle networks," *in Proc. IEEE International Conference on Communications (ICC) 2004, pp. 1628-1633.*

[37] A. Ranjbar and C. Assi, "Availability-aware design in fipp p-cycles protected mesh networks," *International Conference on Optical Networking Design and Modeling (ONDM 2008).*

[38] Z. R. Zhang, W.-D. Zhong, and B. Mukherjee, "A heuristic method for design of survivable wdm networks with *p*-cycles," *IEEE Commun. Lett., vol. 8, no. 7, pp. 467-469, July 2004.*

[39] W. D. Grover and A. Kodian, "Failure-independent path protection with *p*-cycles: Efficient, fast and simple protection for transparent optical networks," *Proc. 7th International Conference on Transparent Optical Networks (ICTON 2005), Barcelona, Spain, 3-7 July, 2005, pp 363-369.*

[40] W. D. Grover, "High availability path design in ring-based optical networks," *IEEE/ACM Trans. Networking, vol. 7, no. 4, pp. 558-574, Aug. 1999.*

[41] M. Clouqueur and W. D. Grover, "Computational and design studies on the unavailability of mech-restorable networks," *in Proc. IEEE/ VDE Design of Reliable Communication Networks (DRCN), Munich, Germany, April 2000, pp. 181-186.*

[42] ——, "Quantitative comparison of end-to-end availability of service paths in ring and mesh-restorable networks," *Proceedings of the 19th Annual National*

Fiber Optics Engineers Conference (NFOEC), Orlando, September 2003, pp. 317-326.

[43] L. Song, J. Zhang, and B. Mukherjee, "Dynamic provisioning with availability guarantee for differentiated services in survivable mesh networks," *IEEE Journal on Selected Areas in Communications 25(S-3): 35-43(2007).*

[44] L. Zhou and W. Grover, "A theory for setting the safety margin on availability guarantees in an sla," *Proc. 5th International Workshop on Design of Reliable Communication Networks (DRCN 2005), Ischia (Naples), Italy, 16-19 October 2005.*

[45] W. Grover and A. Sack, "High availability survivable networks: When is reducing mttr better than adding protection capacity?" *Proc. DRCN 2007, La Rochelle, France, 7-10 October 2007.*

[46] P. B. et al., "Ultra high-capacity optical transmision networks: Final report of action cost239," *Faculty of Electrical Engineering and Computing, University of Zagreb, Tech. Rep., 1999.*

[47] K. Lo, D. Habibi, Q. V. Phung, and H. N. Nguyen, "Dynamic p-cycles selection in optical wdm mesh networks," *in Proc. 13th International Conference on Networks (ICON2005), pp. 844-849.*

[48] "Cplex/ampl:," *http://www.ilog.com/products/cplex/.*