2016

Dynamic Provisioning of Fault Tolerant Optical Networks for Data Centers

Ruchisree Das
University of Windsor

Dynamic Provisioning of Fault Tolerant Optical Networks for Data Centers

by

Ruchisree Das

A Thesis

Submitted to the Faculty of Graduate Studies

through the School of Computer Science

in Partial Fulfillment of the Requirements for

The Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

2016

Dynamic Provisioning of Fault Tolerant Optical Networks for Data Centers

by

Ruchisree Das

APPROVED BY:

_____

Dr. Jagdish Pathak, External Reader
Odette School of Business

_____

Dr. Dan Wu, Internal Reader
School of Computer Science

_____

Dr. Subir Bandyopadhyay, Advisor
School of Computer Science

_____

Dr. Arunita Jaekel, Advisor
School of Computer Science

Sep 7, 2016

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Survivability of files in data centers, when a disaster occurs, is becoming a major challenge in designing cloud-based services. When such a disaster occurs, a specific geographical area is affected and components of communication networks (e.g., nodes and fibers) within the affected area become faulty, leading to the failure of one or more on-going communication. To handle such a situation, a robust communication protocol is needed, so that provisions can be made to allocate an alternative fault-free path, when a disaster disrupts the path used for data communication before the disaster occurs. In this work we have presented a new approach to this problem, in the case of dynamic Route and Wavelength Assignment (RWA) in WDM networks. In our approach, a communication request can be handled only if it is possible to set up i) a primary lightpath that minimizes the number of disasters that may affect the lightpath and ii) (for each disaster that disrupts the primary lightpath), a backup lightpath that avoids the disaster. We have proposed, implemented and studied an efficient heuristic to solve this problem.

DEDICATION

Dedicated to my in-laws Rathindra and Supriya Roy Chowdhury, my parents Prabir and Reba Das and my husband Rajib Roy Chowdhury

ACKNOWLEDGEMENT

This thesis owes its experience to the help, support and inspiration of several people. Firstly, I would like to express my sincere appreciation and gratitude to Dr. Subir Bandyopadhyay and Dr. Arunita Jaekel for their guidance during my research. Their support and inspiring suggestions have been precious for the development of this thesis content.

I would also like to thank my thesis committee members Dr. Jagdish Pathak and Dr. Dan Wu for their valuable comments and suggestions for writing this thesis. In addition, entire computer science faculty members, graduate secretary and technical support staff deserves a special thanks for all the support they provided throughout my graduation.

A special thanks goes to Ms. Saja Al Mamoori, she has been the fundamental support throughout my work.

Finally, my deepest gratitude goes to my family for their unflagging love and unconditional support throughout my life and my studies. You made me live the most unique, magic and carefree childhood that has made me who I am now!

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY iii
ABSTRACT iv
DEDICATION v
ACKNOWLEDGEMENT vi
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF ABBREVATIONS/SYMBOLS xii

1 INTRODUCTION 1
1.1 Overview 1
1.2 Wavelength Division Multiplexing Networks 2
1.3 Faults in Optical Network 3
1.4 Problem Statement 4
1.5 Scope of the thesis 5
1.6 Organization of thesis 6
2 REVIEW 7
2.1 Optical Networks 7
2.2 Data Transmission 9
2.3 Wavelength Division Multiplexing 9
2.4 Physical Layer Impairments (PLI) 11
2.5 Optical Reach 12
2.6 Types of Network 13
2.6.1 Transparent Network 14
2.6.2 Opaque Network 14
2.6.3 Translucent network 14
2.7 Lightpath 15
2.8 Physical and Logical Topology 16
2.9 Routing and Wavelength Assignment 18
2.10 Different Lightpath Allocation Schemes 19
2.10.1 Static Lightpath Demands 20
2.10.2 Dynamic Lightpath Demands 20
2.11 Network Failures due to Disasters 21

2.12 Fault Management in WDM Networks 22
2.13 Datacenter 24
2.14 Literature Review 26
2.15 Summary of Literature Review 28
3 DYNAMIC RWA FOR DATACENTER NETWORKS 29
3.1 Problem Definition 29
3.2 Proposed Approach 30
3.3 Network State 32
3.4 Concept of Virtual Node 33
3.5 Establishment of Primary Lightpath 34
3.6 Establishment of Backup Lightpath 36
3.7 NOTATIONS & ALGORITHMS 40
3.7.1 Disaster Aware Dynamic RWA Algorithm 41
3.7.2 Dynamic RWA algorithm to find Primary Path 44
3.7.3 Dynamic RWA to find Backup Path 46
4 EXPERIMENTATION AND RESULTS 49
4.1 Simulation setup 49
4.1.1 Network topology 49
4.1.2 Algorithm inputs 52
4.2 Performance evaluation of Phase I & II 52
5 CONCLUSIONS AND FUTURE WORK 65
5.1 Conclusion 65
5.2 Future Work 66
BIBLIOGRAPHY/REFERENCES 67
VITA AUCTORIS 72

LIST OF TABLES

Table 2.1: Literature Review Summary 28

Table 3.1: Table showing primary and backup lightpaths 38

Table 3.2: Table showing primary and backup lightpaths 39

Table 4.1: Comparison of Avg. BP in COST-239 network (8 channels/fiber) 54

Table 4.2: Comparison of Avg. BP for COST-239 network (16 channels/fiber) 56

Table 4.3: Comparison of Avg. BP for 14-node NSFNET network (8 channels/fiber) 58

Table 4.4: Comparison of Avg. BP for 14-node NSFNET network (16 channels/fiber) 59

Table 4.5: Comparison of Avg. BP for 20-Node ARPANET network (8 channels/fiber) 60

Table 4.6: Comparison of Avg. BP for 20-Node ARPANET network (16 channels per/fiber) 62

Table 4.7: Comparison of Avg. BP for 24-Node USANET network (8 channels/fiber) 63

Table 4.8: Comparison of Avg. BP for 24-Node USANET network (16 channels/fiber) 64

LIST OF FIGURES

Figure 2.1: Optical Cable 7

Figure 2.2: Layers in an optical fiber 8

Figure 2.3: Total Internal Reflection inside optical fiber 9

Figure 2.4: Wavelength Division Multiplexing System 10

Figure 2.5: Types of Optical Networks 14

Figure 2.6: Lightpath setup in a 5 node network 16

Figure 2.7: Physical Topology 17

Figure 2.8: Logical Topology 18

Figure 2.9: Different Lightpath Allocation Schemes 21

Figure 2.10: Fault Management Schemes 23

Figure 2.11: Datacenter 24

Figure 2.12: Nodes representing datacenters in a network 25

Figure 3.1: Concept of Virtual Node 33

Figure 3.2: Establishment of Primary Lightpath using RWA 35

Figure 3.3: Establishment of backup lightpath to handle disaster on a datacenter node 38

Figure 3.4: Establishment of backup lightpath to handle disaster on an intermediate node 39

Figure 3.5: Overview of disaster-aware RWA algorithm 42

Figure 3.6: Overview of function find_primary_LP 44

Figure 3.7: Overview of function find_backup_LP 46

Figure 4.1: COST-239 network (11 - node topology) 50

Figure 4.2: NSFNET network (14 - node topology) 50

Figure 4.3: ARPANET network (20 - node topology) 51

Figure 4.4: ARPANET network (24 - node topology) 51

LIST OF ABBREVATIONS/SYMBOLS

RWA – Routing and Wavelength Assignment

WDM – Wavelength Division Multiplexing

LP – Lightpath

OEO – Optical-Electrical-Optical

DC – Data Center

DCN – Data Center Networks

CDN - Content Distribution Network

WMD – Weapons of mass destruction

SPP – Shared Path Protection

DPP – Dedicated Path Protection

SLD – Static Lightpath Demands

DLD – Dynamic Lightpath Demands

PLI – Physical Layer Impairments

QoT – Quality of Transmission

OSNR – Optical Signal to Noise Ratio

BER – Bit Error Rate

MUX - Multiplexer

DEMUX - Demultiplexer

OXC – Optical Cross Connect

COST-239 – European network topology

NSFNET – National Science Network

ARPANET – Advanced Research Projects Agency Network

USANET – USA standard network topology

DWDM - Dense Wavelength Division Multiplexing

BP – Blocking Probability

ILP - Integer Linear Program

SRG – Shared Risk Group

IDE – Integrated Development Environment

Chapter 1

INTRODUCTION

1.1 Overview

The rapid growth of broadband communications has led to many new web applications such as online interactive maps, social networks, cloud computing and CDN (Content Distribution Network) services. Most of these applications are provided by Data Center Networks (DCNs) [1-2]. Recent studies show that DCN based applications are reshaping the network landscape, by pushing the traditional hierarchical and connectivity-oriented internet towards a more meshed and service-oriented infrastructure, offering applications the promise of scalability, availability and responsiveness at very low costs [3]. To meet the requirements set by the rising volume of traffic in a datacenter network, optical networks are ideally suited, given their high-bandwidth and low-latency characteristics. But, the risk of large-scale failures in DCN is increasing rapidly and multiple failures generally occur due to natural disasters like earthquakes, hurricanes, tsunamis or deliberate attacks like weapons of mass destruction (WMD).

Survivability against disasters, both natural and man-made attacks, is becoming a major challenge in communication networks. These events indicate that it is crucial to study and develop robust communication schemes to handle requests for communication in the case of a disaster. In this thesis we have proposed a heuristic based approach to the problem of designing robust Data Center Networks (DCN).

1.2 Wavelength Division Multiplexing Networks

Wavelength Division Multiplexing (WDM) in optical networks has made it possible to design large communication networks with high throughput [3]. WDM technology has been improving steadily in recent years, with existing systems capable of providing huge amounts of bandwidth on a single fiber link. WDM systems are currently being deployed in long-distance telecommunication networks on a point-to-point basis, with the optical signals being converted back to electronic signals at each node. These opto-electronic switching and processing costs at the nodes can be potentially be very high, leading to severe performance bottlenecks and limiting the delivery of optical link bandwidth to the end users [3-4]. Hence to avoid such bottlenecks we consider the concept of transparent lightpaths in an optical network. Transparent lightpath is a point-to-point communication path that optically connects a transmitter at a source node to a receiver at a destination node with no optical-electronic conversion at any intermediate node in the route from the source to the destination of the communication.

Lightpath: Signals travel through optical fibers in the form of light. These signals are used to communicate from a source node *S* to a destination node *D*, using an optical signal propagating from *S* to *D* through the fiber network [4].This optical communication path is referred to as a *lightpath* and is characterized by a physical path from *S* to *D* and a carrier wavelength (or channel) λ, on each link (optical fiber) of the path.

Several lightpaths can be transmitted on a single fiber, provided they are transmitted using different carrier wavelengths. One of the challenges involved in designing wavelength-routed networks is to develop efficient algorithms for establishing lightpaths in the optical network [5]. The algorithms must be able to select routes and assign wavelengths to connections in a manner which efficiently utilizes network resources (channels/wavelengths).

Routing and Wavelength Assignment (RWA): The problem of finding a route for a lightpath and assigning a wavelength to the lightpath is often referred to as *Routing and Wavelength Assignment* problem (RWA) [8]. The objective of the problem is to route lightpaths and assign wavelengths in a manner which minimizes the amount of network resources that are consumed, while ensuring that no two lightpaths share the same carrier wavelength on the same fiber link. Furthermore, in the absence of wavelength conversion devices, the RWA problem operates under the constraint that a lightpath must occupy the same carrier wavelength on each link in its route. This restriction is known as the *wavelength-continuity* constraint. The optimal formulation of the RWA problem is known to be NP-complete; therefore, heuristic solutions are often employed [5-6].

1.3 Faults in Optical Network

Fault management has become critical in managing survivability of high speed networks. The impact of failure is aggravated by extremely high volumes of traffic carried on WDM networks. In a WDM network, the failure of a network element may cause failure of several optical channels leading to large data loss which can interrupt communication services. Signal integrity may be compromised by a catastrophic event (fiber cut), or a component failure (laser, optical switch). The most prevalent form of failures in communication networks the is accidental disruption of buried telecommunication cables [10].

In this thesis, we address the problem of optical network survivability against disasters, which can cause severe service disruption due to cascading and correlated multiple node/link failures. Large-scale disasters affect specific geographic areas; as a result, a set of co-located nodes and links may go down simultaneously. Due to various faults there can channel failure, link failure or node failure [29].

1.4 Problem Statement

The problem is to handle a communication request from a user node *t* for a file f_1, where this file is kept at some data center node from the list of data centers *m.* For each such communication, we have to find out a primary path (in the fault free case) and backup paths (in the case of disasters) for transmitting the file f_1 to the user node *t*. The definition of disaster $d \epsilon D$ is such that if a disaster *d* occurs at a particular node of the primary path, then the edges to and from that node path are longer available to carry out any communication.

While finding out the primary path and backup path for a particular communication, we must ensure that the following conditions are satisfied:

- The primary lightpath is from some node j_1 (from the list of data centers m) to any user node requesting the file f_1 located at node j_1.
- The backup path to handle disaster, say d_1 can be from any other data center node, say j_2, other than j_1, to the user node *t* requesting the file f_1.
- The backup path to handle a disaster can be from the same data center node j_1, but through some other intermediate node, if some intermediate nodes of the primary path are disrupted by disaster$d$.
- Every lightpath has to satisfy the wavelength continuity constraint, which means that on all edges along a single path from source to destination, the same carrier wavelength will be used.

When a new request for communication is received for file f_1 from user node *t*, the network has a number of existing communications for the same file f_1 from different user nodes in the network. Every time a primary or backup path is established to handle a communication request, we have all the information like the edges traversed and the channels (wavelengths) used

on those edges for both the primary path (in fault-free case) and the backup paths (to handle disaster $d \epsilon D$) to establish a communication request, stored in our updated *edge-channel database*. After a communication is successfully handled, the *edge-channel database* is updated with primary path and backup path's information.

Our objective is to establish the new communication request using a minimum cost of resources, where the cost is the minimum total number of new channels used. We have explained the problem and our proposed solution to solve it more elaborately in chapter 3.

1.5 Scope of the thesis

This thesis considers disaster-aware RWA with SPP (Shared Path Protection) for Dynamic Lightpath Demands (DLD) in WDM networks. We propose a scheme to handle requests for communication that takes into account the possibility of disasters that may affect multiple edges, nodes and data centers. Our objective was to minimize the usage of network resources (links or channels on links) on optical fibers when our network receives a new communication request. We tried to minimize the cost associated with a new channel every time a backup path is established, by trying to share channels, if possible, between backup paths of an existing communication and a new communication request. Fiber channels for backup lightpaths are shared resources if they are used to handle different disasters giving a new avenue for significant resource savings. Our algorithm uses a shortest path algorithm to find a primary path (in fault free case) and backup paths (for disasters that affect the primary path). Our simulation results indicate that the blocking probability increases with more disasters and decreases as we increase the number of channels per fiber.

1.6 Organization of thesis

The rest of this thesis is organized as follows: Chapter 2 provides a review of some of the concepts and terminologies that are related to this work and seeks to provide more details of the areas related to this research. It also includes a review of some of the closely related work of other researchers. In Chapter 3, we define the problem and present the proposed algorithm. A heuristic solution for dynamic lightpath allocation is presented. In Chapter 4, we present a summary of the results of the experiments carried out. In Chapter 5, we conclude the thesis by proposing some future work.

Chapter 2

REVIEW

2.1 Optical Networks

An optical network connects computers or any other devices which can generate or store data in electronic form using optical fibers. An optical network consists of number of nodes which are interconnected to each other to carry out communication across the network. Data is transmitted between sender and receiver node in the form of light pulses (optical signal) through optical fibers. Optical fibers are long, thin strands of pure glass having a diameter of a human hair [15]. They are generally arranged in bundles, known as optical cables and are used to transmit signals over long distances. The figure 2.1 shows the optical cable with the bundle of several optical fibers. [Image taken from http://fibresale.com.au]

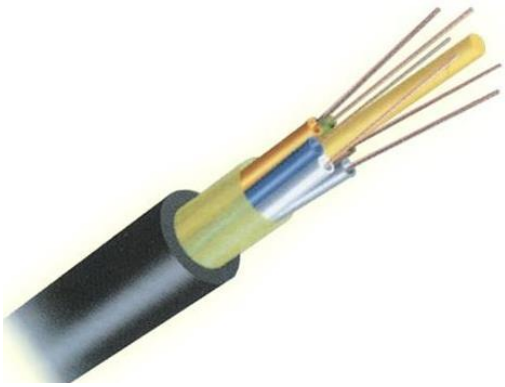


Figure 2.1: Optical Cable

Fiber optic data transmission systems send signals through optical fibers by converting electronic signals into light [16].

An optical fiber is made up of three layers: *core*, *cladding* and *buffer*. *Cylindrical core* is the innermost layer and is made up of a very high quality glass (silica) or plastic [14-16]. *Cladding* is the outer material surrounding the core and it is also made of glass. The third layer i.e. *buffer* is known as the outermost layer of an optical fiber and is made up of plastic such as nylon or acrylic. A buffer protects both the core and cladding from any kind of physical damage. An optical signal travels through the core in the form of light pulses and bounces into the cladding which again reflects the light back to the core. This phenomenon is known as *total internal reflection* and it results in lower light signal attenuation and less energy loss [16]. Fig. 2.2 shows a typical optical fiber and its three layers.

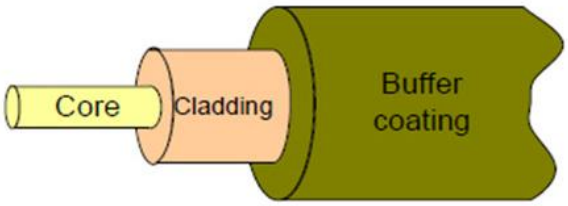
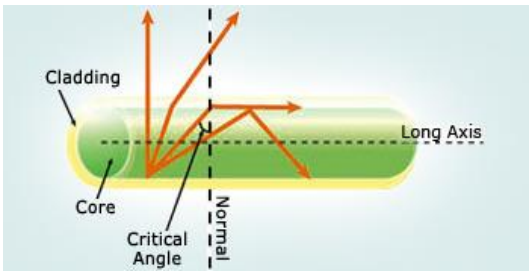
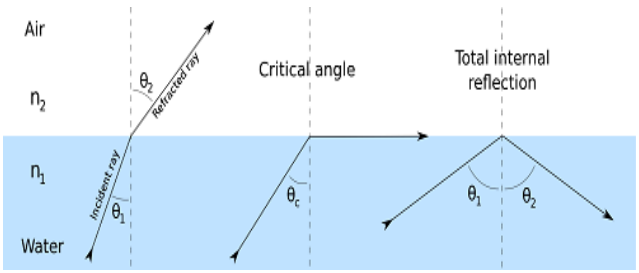


Figure 2.2: Layers in an optical fiber

Fig. 2.4 shows the cross-section a typical optical fiber. [Image taken from http:www.ustudy.in]





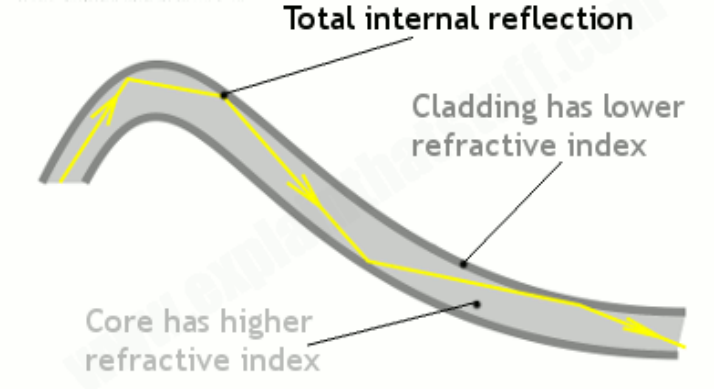


Figure 2.3: Total Internal Reflection inside optical fiber

2.2 Data Transmission

In an optical network, data communication is achieved by the use of transmitters at source at receivers at destination of the lightpath [15]. The main component of transmitter is a laser diode that is used to provide a beam of light, and the main component of a receiver is a photodetector which is used to detect a beam of light. Modulation is a procedure of converting data in electronic form to encode an optical signal [3]. The heart of the receiver is photodetector or photodiode which converts an optical signal into an electrical form at the destination at some particular carrier wavelength. It restores or extracts the data into the original form i.e. is electrical form. However, there may be some distortion of signal due to the presence of *physical layer impairments*. Every channel owns a corresponding transmitter and receiver pair.

2.3 Wavelength Division Multiplexing

The enormous bandwidth requirements faced by today's communication networks have stimulated the massive deployment of optical backbone networks. Wavelength-Division Multiplexing (WDM) has emerged as the most popular technology for optical networks, due to

its flexibility and robustness. In a WDM network, an end-to-end connection is established through a wavelength/channel, known as *lightpath* [8].

The technology which combines multiple optical signals in a single optical fiber by using different carrier wavelengths of laser light is known as *Wavelength Division Multiplexing* (WDM) [12]. As optical network supports huge bandwidth, WDM network splits this huge bandwidth into a number of smaller bandwidth optical channels. It allows multiple data stream to be transferred along the same fiber at the same time [7-12] by using different carrier wavelengths.

WDM networks can transfer data on multiple channels by using a single fiber. In WDM networks light from different laser sources, each with a different wavelength is combined into single beam with the help of a multiplexer. At the receiving end a Demultiplexer (DEMUX) is placed that separates the wavelengths from the beam into independent optical signals. Generally the transmitter consists of a laser and a modulator. The light source generates an optical carrier signal at either fixed or tunable wavelength. The receiver consists of a photodiode detector which converts an optical signal into an electrical signal [12].

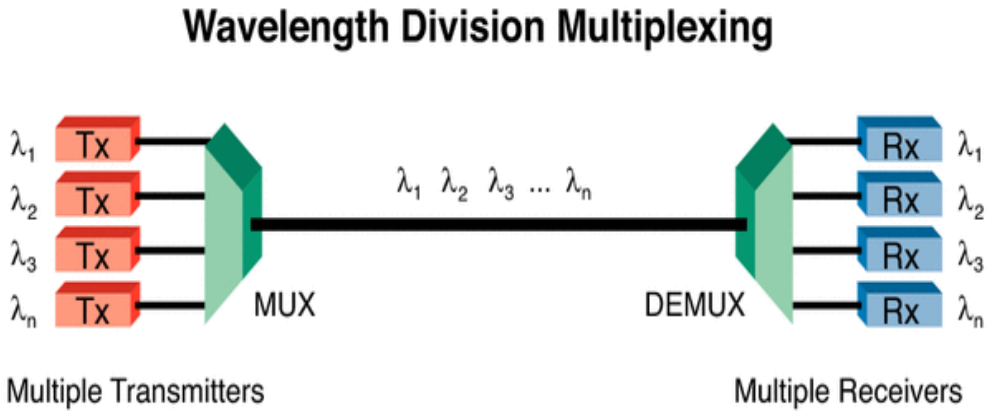


Figure 2.4: Wavelength Division Multiplexing System

Fig. 2.5 shows a WDM system with *n* channels (wavelengths). The sender end has *n* transmitters, each tuned to a different wavelength from λ_1 to λ_n. At the input end, multiplexer combines the signals together into one composite signal to be transmitted through the fiber. The input data to be communicated is converted from electrical to optical form. Similarly, at the receiver side there are *n* receivers, each tuned to a different wavelength from $\lambda 1$ to λn. So, the signal is de-multiplexed (separated) and tuned into the corresponding wavelength and converted from optical to electrical to retrieve the original signal at the receiver end. An advantage of high speed data transmission using optical networks is that a single fiber can accommodate hundreds of channels.

2.4 Physical Layer Impairments (PLI)

In optical networks, as signal propagates through a fiber it undergoes degradations. These degradations are caused by physical layer impairments (PLI) in the fiber [7]. In other words, the distance a signal can travel along an optical fiber is limited, which is known as *optical reach*. As the distance increases, the effects of PLI are more pronounced and the signal degrades more and more. After a certain distance (optical reach), the signal becomes so distorted that it cannot be used further for any communication [8].

Whenever an optical signal is propagated over a network, physical layer impairments (PLIs) in the components and links of the optical network lead to degradation of a signal. It is important to understand that PLI are not necessarily defects, but characteristics of the fiber. As travelling distance increases, the effects of the PLI become more significant. At some point, the quality of the signal falls below the acceptable level, which leads to error in transmission. The overall effects of PLI determine the feasibility of an optical path [12].

To verify the feasibility of a lightpath, quality of transmission (QoT) is measured. QoT is evaluated at the destination node of a lightpath by computing the Q-factor, which is directly linked to the bit error rate (BER) and optical signal-to-noise ratio (OSNR) [17]. The bit error rate is defined as the rate at which errors occur in a transmission system [18]. This can be directly translated into the number of errors that occur in a string of a stated number of bits [17]. It is the error that is generated in the transmission system. For each lightpath, one verifies whether the signal quality at destination is acceptable or not with respect to the admissible BER threshold. In some cases the BER is so high that we cannot adequately recover from the errors and we say that the QoT is unacceptable and that the lightpath is not feasible for data communication [19].

Blocking Probability (BP): The blocking probability is defined as the ratio of the blocked request (lightpaths cannot be provisioned, either due to PLI or lack of network resources) to the total number of requests for communication.

2.5 Optical Reach

Optical reach [18] (also called as transmission reach [24]) is the maximum distance an optical signal can travel before signal regeneration is needed. Longer an optical reach is, lesser is the amount of regeneration needed. Many factors affect the optical reach; for example, the type of amplification, launched power of the signal, and modulation format of the signal [25].

An optical signal suffers from physical layer impairments such as noise, dispersion and nonlinear effects that are induced by long-haul and ultra-long-haul optical equipment [22]. As a result of this deterioration a high bit error rate is induced at destination of a lightpath. Due to PLI the QoT of signal is also deteriorated, as the distance increases. Generally, in WDM networks as channel gap decreases or number of channels increases, the signal quality becomes very poor due to high degree of interference. Due to all of these factors it is very important to minimize the

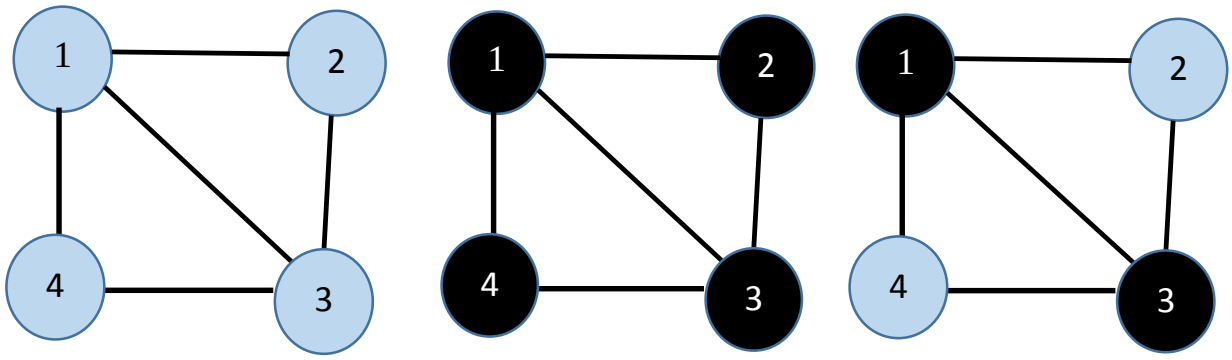
impact of PLIs. This problem can be tackled, using a well-known technique called maximum distance constraint [25].

The impact of linear impairments can be measured analytically by measuring the value of ratio of optical signal to noise (OSNR) for a given bit error rate. Nevertheless, there is a more intuitive way in which QoT of a signal is calculated based on the distance travelled by a signal in a fiber. It can be observed that the quality of signal falls as the distance increases, and the impact on QoT becomes more visible [21]. As soon as it reaches certain distance, the configuration of signal drops resulting in unacceptable value of bit error rate, making the signal useless for communication further than that distance. This progression is known as the maximum distance constraint, which is defined as the maximum distance a signal can traverse that, can enable a proper communication.

On the other hand, as we know that non-linear impairments/constraints are dynamic in nature, and hence difficult to identify and calculate. In addition, the intensity and quantity of an interference and disruption may fluctuate substantially. To reduce the impact of various impairments/constraints a bound is put on the maximum transmission distance, which is known as optical reach. An *optical reach* is thus defined as the maximum distance an optical signal can travel before the signal quality degrades to a level that necessitates regeneration [25].

2.6 Types of Network

An optical network can be classified into three categories, based on the presence and usage of regenerators. Figure 2.6 shows the three different types of networks, namely (a) *Transparent Network*, (b) *Opaque Network* and (c) *Translucent Network.*



(a) (b) (c)

Figure 2.5: Types of Optical Networks

2.6.1 Transparent Network

In *transparent optical* networks there is no Optical-Electrical-Optical (OEO) conversion at intermediate nodes, and the signal always stays in an optical domain between the source and the destination. In other words, there is no regeneration at any node as shown in Figure 2.6 (a). One important constraint for transparent lightpath is the total length of the lightpath should not exceed the optical reach.

2.6.2 Opaque Network

In *opaque networks* every node performs signal regeneration as shown figure 2.6 (b). A regenerator is placed at each node so a signal undergoes O-E-O conversion and re-amplified, reshaped and retimed at every intermediate node along its path. An advantage of opaque networks is that it minimizes blocking probability and hence increases the network's performance. On the other hand, the cost of performing regeneration at every node is very high.

2.6.3 Translucent network

Translucent network architectures have been proposed as a compromise between opaque and all-optical (transparent) networks [18]. In this type of network regenerators are placed

sparsely and strategically to uphold the adequate signal quality from source node to destination. In translucent networks a signal can travel as long as the signal quality does not falls below a threshold value. This approach also eliminates much of the electronic processing required in opaque networks, and allows a signal to remain in the optical domain for much of its path [18].

2.7 Lightpath

In an optical fiber signal travels in the form of light from a source to a destination node through the optical fiber network. A lightpath in a WDM network is a unidirectional optical connection between a source node and a destination node, which carries data in the form of encoded optical signals and may span multiple fiber links and use one or multiple wavelengths [4]. Different lightpaths in WDM networks can use the same wavelength, as long as they do not share any common links. Due to the WDM technology, multiple lightpaths can be established on the same fiber using different carrier wavelengths. A lightpath is established based on two criteria:

- Finding a route for the lightpath.
- Assigning a certain unique wavelength to the lightpath along that route.

If two lightpaths share the same fiber, then they must be assigned two different wavelengths on that fiber. In our work we assume the wavelength continuity constraint, which requires that the same wavelength be maintained along the entire lightpath [42]. Lightpaths are clear optical paths between two edge nodes. Figure 2.6 below shows an example of 5 lightpaths established over a 5-node physical fiber network.

The lightpaths in the following example are:

Lightpath 1: node 1 → node 5

Lightpath 2: node 1 → node 2

Lightpath 3: node 5 → node 4

Lightpath 4: node 5 → node 3

Lightpath 5: node 2 → node 4

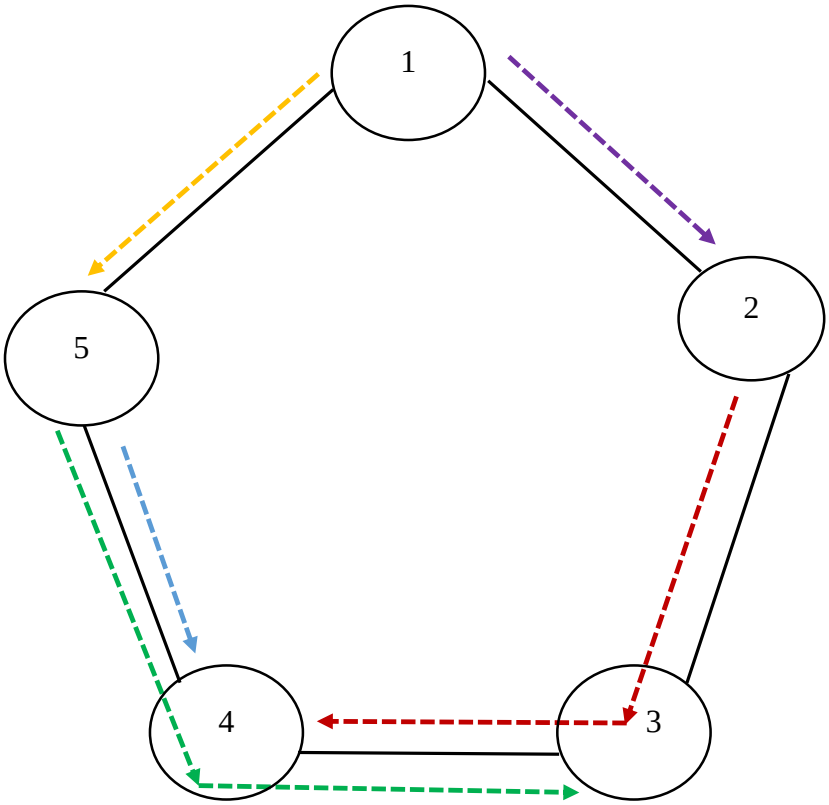


Figure 2.6: Lightpath setup in a 5 node network

2.8 Physical and Logical Topology

In fig 2.7, we have shown a 5 node network, which represents a map like structure of the network components being used in a network. This structure shows the relationship among different network components in which the circles represent the nodes of the network and solid

lines with arrows represent the actual unidirectional fiber that acts the link between the two end nodes. The physical topology of the network is represented by graph *G [N, E]*, where *N* is the set of the nodes in the network and *E* is the edges of the network. Each edge of the network (i.e. the bidirectional link between the nodes) is constructed using two unidirectional links.

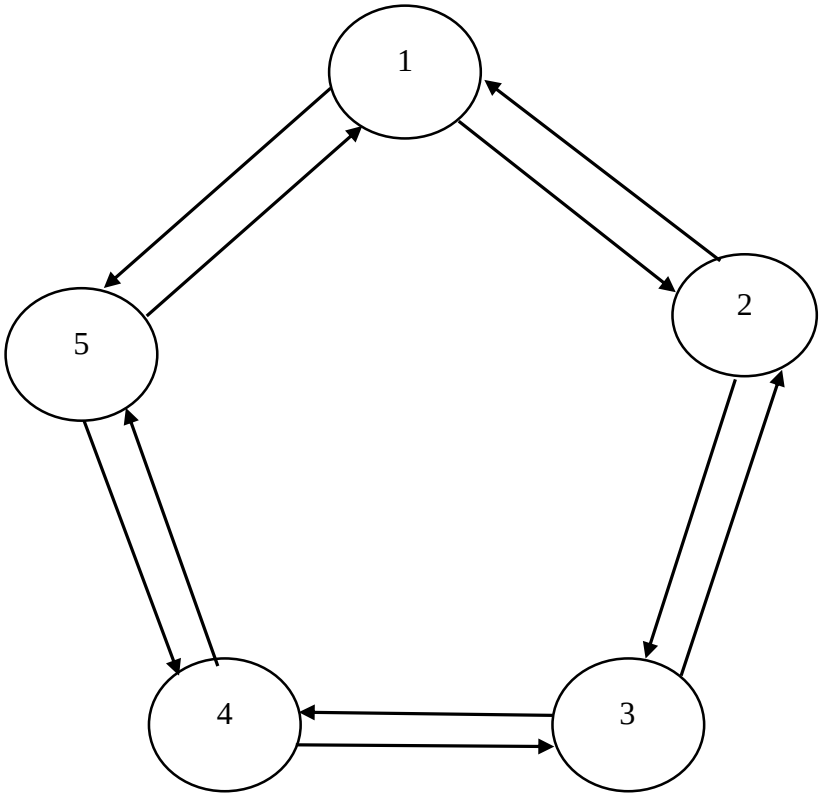


Figure 2.7: Physical Topology

Logical topology is the strategy that is followed for establishing communication between a source and a destination node in a network. In this instance, the lightpaths are viewed as the edges in the network connecting to the nodes in the physical topology. These nodes are same as the ones in the physical topology, but the lightpath between the sources and the destination nodes are the logical edges, and this representation of a network is known as a logical or virtual

topology [24]. Figure 2.8 represents the logical topology that is established over the physical topology in Figure 2.7, and lightpath setup in Figure 2.6.

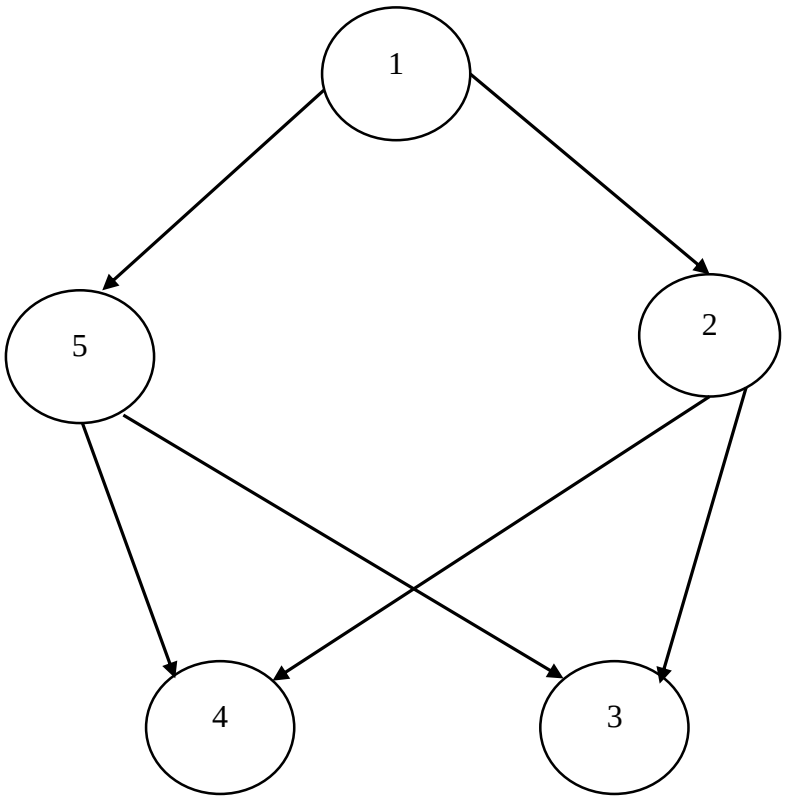


Figure 2.8: Logical Topology

2.9 Routing and Wavelength Assignment (RWA)

In WDM optical networks, the data is communicated from one node to another through lightpaths. To establish a lightpath for a connection request, a physical route from the source node to the destination is first determined and then an available wavelength on each link of the route is assigned. This problem of routing and assigning a wavelength to establish a lightpath is known as the Routing and Wavelength Assignment (RWA) problem [5].

The main objective of the RWA problem is to establish as many lightpaths as possible, keeping resource limitations in mind, which minimizes the network operation cost and increases the network performance [5]. As can be seen in Figure 2.7 a lightpath is implemented by selecting a path of physical links between the source and destination edge nodes, and reserving a particular wavelength on each of these links for the lightpath.

Routing and wavelength assignment are subject to the following two constraints:

- **Wavelength continuity constraint:** a lightpath must use the same wavelength on all the links along its path from source to destination node.
- **Distinct wavelength constraint:** All lightpaths using the same link must be allocated distinct wavelengths.

2.10 Different Lightpath Allocation Schemes

Network traffic or lightpath demands can be broadly divided into two categories: static lightpath demands and dynamic lightpath demands. The major difference between them is the lifetime of these requests. In static lightpath allocation all the requests are known in advance. This is also referred to as permanent (or semi-permanent) lightpath allocation; because once the request is set up that lightpath is expected to continue for a relatively long time - weeks, months or years. After some time, if the traffic pattern changes, a new set of lightpaths can be established. The RWA corresponding to a set of static lightpath requests is typically computed offline.

On the other hand in dynamic lightpath allocation the requests are not known in advance they are handled as and when they occur. These requests have a specified lifetime i.e., a start time when the lightpath is set up and an end time when the lightpath is taken down [27], which is typically of much shorter duration compared to static lightpaths. These requests are taken

down when communication is over, and the resources allocated to the lightpath can be reused. The dynamic lightpath allocation is further divided into scheduled lightpath demands and ad-hoc lightpath demands. The requests for which the start time and end time are known in advance (and are often periodic) are called scheduled lightpath demands (SLD). The demands, for which we neither know the start time nor the duration of such request in advance, are known as ad-hoc lightpath demands (ALD) [28]. In this thesis we focus on dynamic lightpath demands. Figure 2.7 shows the various lightpath allocation schemes.

2.10.1 Static Lightpath Demands (SLD)

In static allocation all lightpaths are planned in advance so that either a specific lightpath is preassigned for each possible source destination pair or the entire set of lightpath requests is known beforehand and channel assignments are made for the request as a whole [39].

2.10.2 Dynamic Lightpath Demands (DLD)

In dynamic allocation, Lightpaths are created on demand and are taken down when the communication is over, and the WDM channels used for this communication are reclaimed for future use in some other communication [40], [41].

In such a scheme, all existing lightpaths have to be considered when creating a new lightpath to sup-port a new communication request. A dynamic scheme does not guarantee that communication tom a source to a destination will always be possible. If the conditions for establishing a lightpath are not satisfied, the communication will be blocked and may possibly be attempted again after some delay with the expectation that the conditions for establishing a lightpath are now satisfied. Clearly, in realistic situations, it is necessary to ensure that the probability that a communication is blocked is very low.

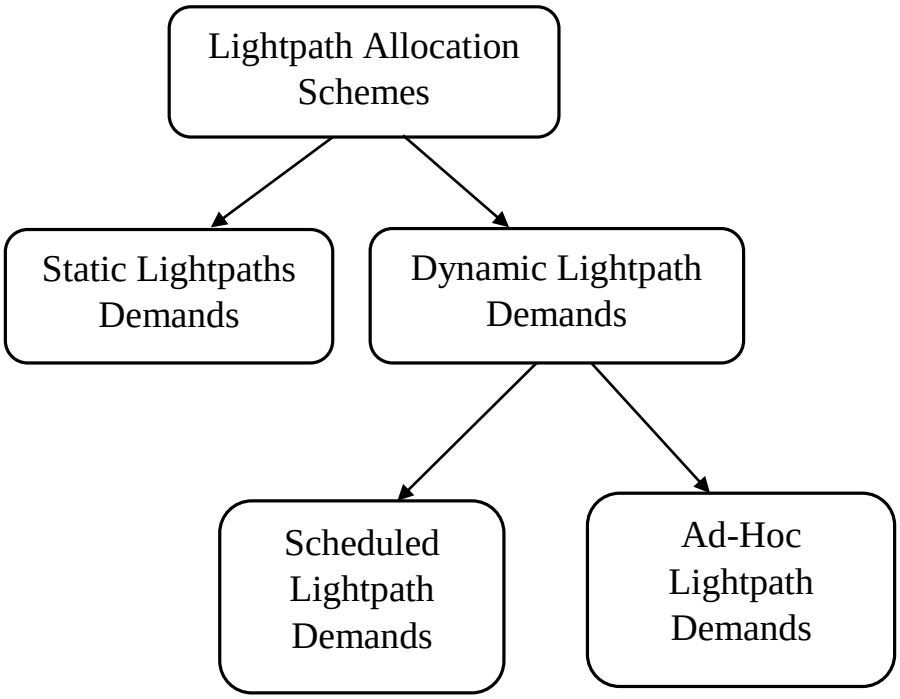


Figure 2.9: Different Lightpath Allocation Schemes

2.11 Network Failures due to Disasters

Failures: A network component may fail due to various reasons, but the most common type of failure is a link failure. In a link failure, the communication between two adjacent nodes gets disrupted due to a fault in the fiber connecting the two nodes. Survivability, the ability of a network to withstand and recover from failures, is one of the most important requirements of networks [9]. In order to design a survivable optical network, it is important to lay out the possible failures under which the network must be survivable. The basic types of network failures generally considered are *link* and *node* failures. Cable cuts that cause link failures are common in optical networks. Man-made errors and uncontrollable natural phenomena (e.g., floods and earthquakes) cause equipment, and therefore node, failures. Besides node and link failures, which are common failure situations in any communication network,

channel failure is also possible in WDM optical networks. A channel failure is usually caused by the failure of transmitting and/or receiving equipment operating on that channel (wavelength) [10].

In case of disasters, multiple nodes/links in the network may be disrupted and the same time. There has been significant research focus on both single link and node failure events [28]. Recently multi-failure studies have been conducted [29], [30]. However, protection against correlated node-link failures caused by a single *dominant disaster* is a topic that needs serious attention. In our work, we have focused on such failures while trying to minimize the network cost in terms of channel/wavelength usage.

2.12 Fault Management in WDM Networks

Fault management in optical network is done through reserving backup resources in advance called *protection* or discovering spare backup resources in an online manner called *restoration* [43]. Protection of paths against failures can be achieved by providing a backup path to the same destination, such that this backup path should be link-disjoint to the primary path. The classical protection schemes do not provide protection against dominant failure scenario (disaster) especially when it covers a region which affects both the primary and backup paths. Protection schemes include *dedicated path protection (DPP), shared path protection (SPP), dedicated link protection* and *shared link protection* while restoration schemes include *path restoration* and *link restoration* [31-43]. An organization of these schemes is shown in Figure 2.8.

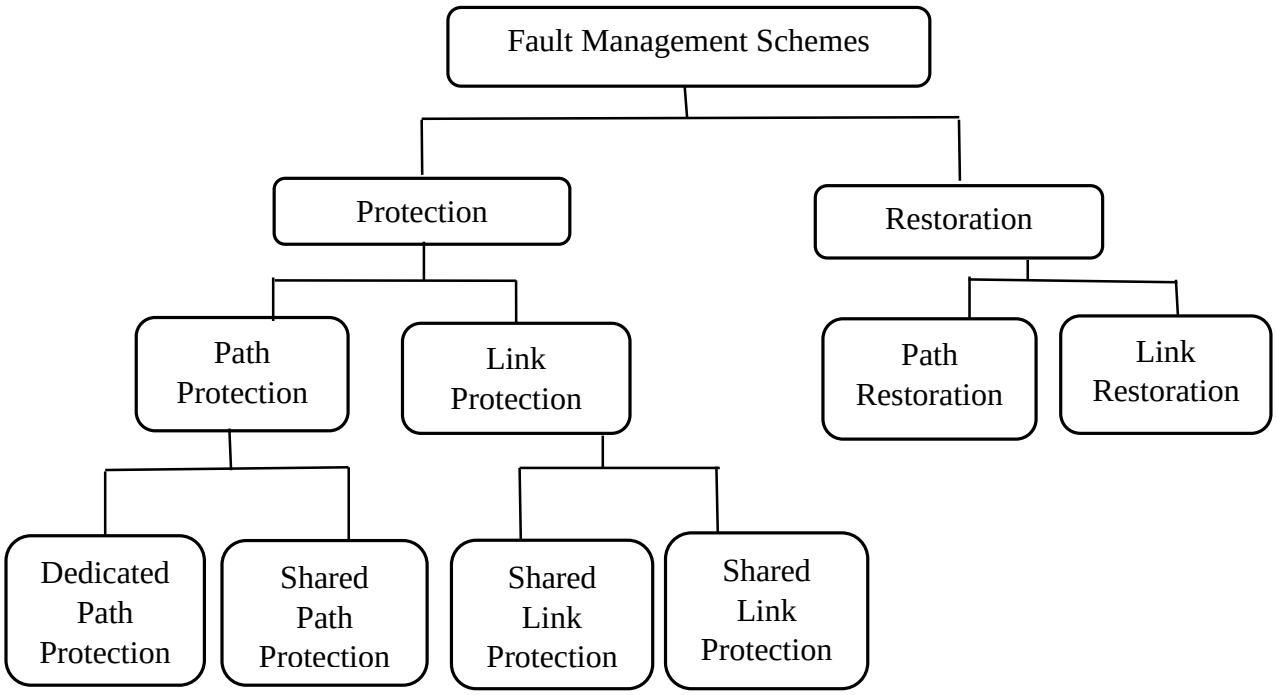


Figure 2.10: Fault Management Schemes

There are essentially two types of path protection [9-30] schemes: *dedicated path protection* and *shared path protection*. In *dedicated path protection*, at the time of call setup for each primary path, a link-disjoint backup path and wavelength are reserved which are reserved for/dedicated to that call. The backup wavelength reserved on the links of the backup path are dedicated to that call only, and are not shared with other backup paths. In *shared-path protection*, at the time of call setup for a primary path, a link-disjoint backup path and wavelength are also reserved. However, the backup wavelength reserved on the links of the backup path may be shared with other backup paths. As a result, backup channels are multiplexed among different failure scenarios (which are not expected to occur simultaneously), and therefore shared-path protection is more capacity efficient when compared with dedicated-path protection. In our work we are considering shared path protection so that channels can be shared among different backup paths.

2.13 Datacenter

A Datacenter (DC) can be visualized as a server that is used for storage, computing resources and distribution of large amount of data. The resources and data in a datacenter are served to customers through a network of datacenters, which is referred to as the cloud [14]. With the rise in demand for cloud services, huge amounts of digital content are being created and shared all the time over the network. Typically the content and services of data centers are replicated over multiple data centers, so that a user request can be served by any datacenter that hosts the required content. This replication strategy also solves the problem of data availability in the event of a disaster (earthquake, tsunami etc.), which may lead to failure of network components (failure of a node or optical fiber in the network). The concept of data replication was initially studied in [11] with the objective of reducing network cost and latency. In [13] the importance of a certain replica is based on the popularity of data.

Figure 2.11: Datacenter

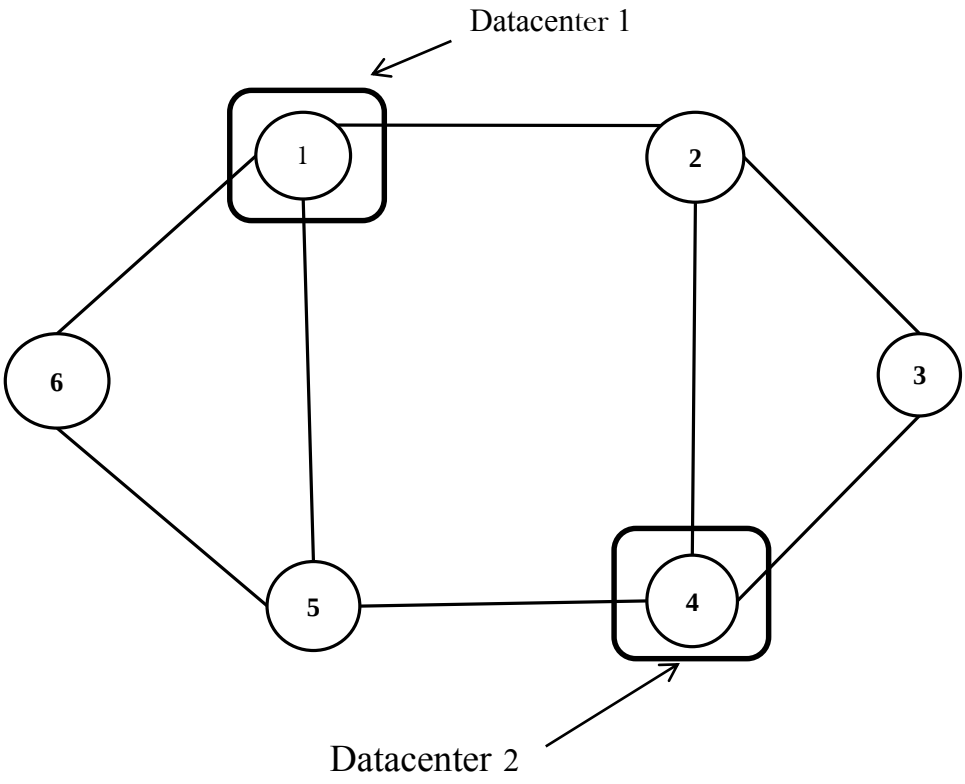


Figure 2.12: Nodes representing datacenters in a network

Survivability of datacenters against disasters, both natural and man-made attacks is becoming a major challenge in designing cloud-based services, hence making cloud network design an important problem. In our work we consider that the communication requests are not static in nature and they arise dynamically, hence we use the concept of *dynamic lightpath allocation* for establishing any communication request. In dynamic lightpath allocation using path protection when a communication request from source *S* to destination *D* is received, our objective is to search for resources to set up two lightpaths – a primary lightpath and a backup lightpath.

Cloud services delivered by datacenter networks yield new opportunities to provide protection against disasters. In such a network, heterogeneous contents and services are

replicated over multiple datacenters, so that a user request can be served by any datacenter that supports the specified content or service. This scheme, where the required content/service can be served from one of many potential datacenters, is known as *anycast* service [38]. Such services require infrastructures with high capacity, high availability, and robustness to serve the rising volume of traffic, and optical networks are well suited to meet these requirements [20].

Research has been initiated introducing backup datacenters following the *anycast* principle to reduce bandwidth consumption [38]. Content/data protection is a fundamental problem in datacenter networks, as the failure of a single datacenter should not cause the disappearance of a specific content/data from the whole network. According to recent studies, it is crucial that the network supporting such services is resilient to data loss or service disruptions, hence making cloud network design an important problem.

2.14 Literature Review

In this section we discuss in detail the papers that are directly related to our thesis.

In [33] the authors address the problem of path protection in datacenter networks that offer cloud services. They also address the problem of content placement, routing of path in the network. The authors have developed a new integrated Integer Linear Program (ILP) formulation to design an optical datacenter network. The disaster protection scheme proposed in this paper uses anycast service and provides more protection, but uses less capacity than dedicated single-link failure protection. The authors formulate their problem of assigning paths to high bandwidth connections, determining content replica placement and providing shared path protection against single disaster failure (i.e., multiple failures caused by a single disaster) for both paths and content. Since, the ILP does not scale well, the authors also propose a heuristic to derive a feasible solution for a request *(s,c)* from the LP relaxation of the original ILP.

In [34], the authors jointly optimize the problem of Datacenter Network (DCN) placement with service routing and protection to minimize the network cost, while ensuring fast protection of all services against link or server failure. The authors develop an ILP to achieve an optimal design to solve the problem. The authors consider a WDM optical backbone network with topology, where DC nodes are placed at a selected subset of network nodes. They assume that a sufficient number of wavelengths can be multiplexed onto each link for high-speed optical transmissions.

In [38] the authors proposed a disaster-aware protection scheme that adopts the principle of *manycasting* - establishing multiple paths to provision bandwidth for services distributed over multiple servers/datacenters, where datacenters are placed at a selected subset of network nodes and the files/contents are not fully replicated. The scheme provides degraded service (versus no service at all) and survivability in the case of multiple disasters using a probabilistic model for disasters.

In [35], the authors address the problem of fast and coordinated data backup in geographically distributed optical inter-DC networks in order to improve the data-transfer efficiency of the regular DC backup. In order to prevent data loss in an optical inter-datacenter (inter-DC) network, a cloud system usually leverages multiple datacenters (DCs) for obtaining sufficient data redundancy.

In [37], the authors considered) the problem of resource allocation before disasters and ii) re allocation of resources after a disaster. The study used a probabilistic model for failures and proposed a proactive (i.e., before disaster) disaster-aware provisioning schemes with the objective of minimizing the loss/penalty in the case of a disaster. The authors also investigated

a reactive (i.e., after a disaster occurs) scheme for re-provisioning the connections affected by the network component failures resulting from the disaster.

2.15 Summary of Literature Review

Table 2.1: Literature Review Summary

Reference	Type of communication request handled	Protection Scheme used	Solution Approach	Type of failure addressed
Habib et al. 2012 [33]	Static requests	Dedicated Path Protection	ILP	Single link failure
Xiao et al. 2013 [34]	Dynamic requests	Dedicated Path Protection	ILP and Heuristic	Single link failure
Mukherjee et al. 2014 [38]	Static requests	Shared Path Protection	ILP	Single link failure
Yao et al. 2015 [35]	Static requests	Shared Path Protection	ILP and Heuristic	Single link failure
Tornatore et al. 2015 [37]	Static requests	Dedicated Path Protection	ILP	Single link failure

Chapter 3

DYNAMIC RWA FOR DATACENTER NETWORKS

3.1 Problem Definition

This section presents the problem that we are trying to solve using a heuristic approach. The problem is to handle a request for communicating a file f_i to a destination node t requesting the file. There are replicated copies of the file f_i located at different *data centers*. The replication strategy is such that in the case of disaster $d \epsilon D$, for each file f_i, we want to make sure that at least one site of the file can communicate with any destination node *t*. Given a communication request (f_i, t) , we have to;

1) Select a site s_i for the file f_i when the network has no disasters.
2) Find a viable lightpath using Routing and Wavelength Assignment (RWA) from site s to destination node t requesting the file. This lightpath will be known as *Primary Lightpath* for communicating a file f_i to destination node t.
3) The channel used by a primary lightpath is exclusively allotted for that lightpath itself. For each fiber on the primary path, the channel cannot be used by any other communication request to establish their primary paths or backup paths.
4) For each disaster $d \epsilon D$, find out a viable lightpath to transfer the file f_i from site s_j containing f_i to destination node t, requesting the file f_i. In the case of a disaster d, affecting the primary lightpath, the lightpath obtained from step 2 will not be effective

any longer. Hence we want to find an alternate route and channel for every disaster $d \epsilon D$ affecting the primary path. These lightpaths will be known as *Backup Lightpaths*, since they will be used if and only if there happens to be disaster d affecting the primary lightpath.

5) The channel used by a backup lightpath is sharable with other communication requests' backup lightpath, if and only if the backup lightpaths handle different disasters $d \epsilon D$.

Our approach is different from previous approaches as the previous approaches address the problem of *static* lightpath allocation in datacenter networks, while in our work we address the problem of *dynamic* Routing and Wavelength Assignment (RWA).

3.2 Proposed Approach

The requested file f_i is located at some node, which we call as $datacenter$ node in our network. There are replicated copies of the file f_i located at more than one datacenter node. We have used an optimal replication strategy that has been designed and implemented by one of our team members, which gives us the information of how many copies of file f_i are there in the network and at which data centers store are the replicated copies of file f_i.

If we are successful in handling a request for communicating file f_i to a destination node t requesting the file, we should be able to get information about the *primary lightpath* which will be used to carry out the communication in a fault-free (*disaster free*) case and also the information about the *backup lightpaths* which will be used to carry out the communication in case of each disaster, $d \epsilon D$, where a disaster can occur either on the datacenter node itself or on any other intermediate nodes of the primary lightpath. The backup lightpath should be such that it avoids all the edges from the primary lightpath whichever has been affected by disaster $d \epsilon D$.

The result of the primary lightpath and the backup lightpath(s) will give us the information about which edges (optical fibers) have been used to establish the primary lightpath and the backup lightpath and which wavelengths (channels) have been used on those edges to establish the lightpaths. While determining the primary and backup lightpaths for a new communication request (f_i, t), for file f_i, to be transferred to node t, the following conditions must be satisfied:

- A lightpath is established to handle the fault-free case henceforth called a *primary lightpath* from some site s_i (a datacenter node containing the file f_i) to node t (destination node requesting the file f_i).
- A lightpath established to handle the scenario of any disaster, $d \in D$, where a disaster d disrupts edges of the primary lightpath, is from some node site s_j (where j may be different from k) to the same node t (destination node). Such a lightpath must avoid any of the edges disrupted by the disaster d $(d \in D)$ and is known as backup lightpath to handle disaster d, $d \in D$.
- Both the primary and backup lightpaths should satisfy the *wavelength continuity constraint.* Wavelength continuity constraint states that all the edges of a lightpath should have the same wavelength.
- Both the primary and backup lightpaths should satisfy the *wavelength clash constraint* with respect to the existing lightpaths on the same fibers. Wavelength clash constraint states that no two lightpaths, sharing the same edge/fiber can share a common wavelength, at the same time.

In our work the replication strategy is predefined. The replication strategy ensures that during any dominant disaster, at least one copy of each file from a set of files F is available

to the user at any node *t*. We consider that a channel *k* on the edge $e\ (i\ \rightarrow\ j)\ \epsilon\ E$ belongs to one of the following categories:

- *Category 1:* The channel on an edge has not been used by any previous communication. Such a channel can be used by a new communication request and the cost of such a channel in order to be used on a primary path or a backup path will be 1.
- *Category 2:* The channel is used by a primary path of some existing communication. Such a channel cannot be used by any new communication neither to establish a primary path nor a backup path.
- *Category 3:* The channel has been used by an existing communication to handle one or more disasters. We use the channel sharing strategy in this case, while designing our backup paths. If we want to design a backup path to handle a disaster d_{i} for a new communication, then we can use a channel that has been used previously by an existing communication to handle any other disaster d_{j}, where $d_i \neq d_j$. By this we mean that two different communications can share the same channel to handle different disasters d_{i} and d_{j} but they can never share a channel to handle the *same* disaster. The cost associated with such usage of channel will always be 0.

3.3 Network State

When we receive a request for transmitting a file f_i to a node t, requesting the file, the network already contains a number of ongoing communications, which were established earlier. Each ongoing communication has a primary lightpath and backup lightpath(s) established for them. The details of all the existing communications, like the path and the channel used by the primary lightpath (in fault-free case) and the backup lightpaths (to handle disaster d, $(d\ \epsilon\ D)$) are known to us.

Given such a network scenario, our objective is to minimize the cost of resources associated with the establishment of the primary lightpath and the backup lightpath(s) while handling a new request for transmitting a file f_i to a destination at node t. Resources used to handle a request are measured in terms of the total number of channels (*k*) used by both primary lightpath (in fault-free case) and backup lightpaths (if any) on all edges of the communication request. The cost of a primary path is counted as 1 on each edge *e* that is used by the primary path, while the cost of backup path for disaster d_0 will be counted as 0 on the edge *e* where it shares a channel *k* with another disaster d_1 and as 1 on the edge *e* where it uses a channel *k* that was never been used previously.

3.4 Concept of Virtual Node

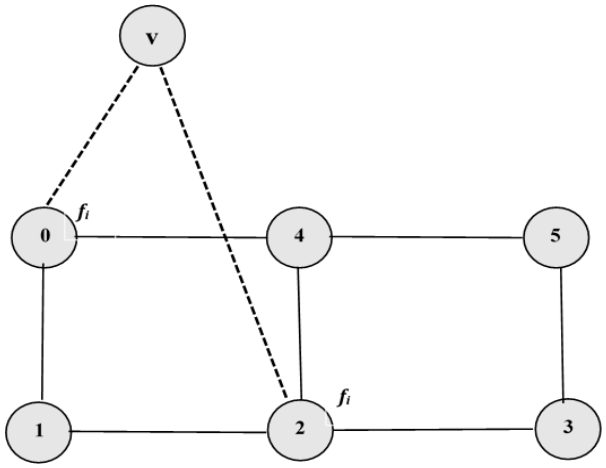


Figure 3.1: Concept of Virtual Node

Since we have multiple copies of file f_i located at different datacenter sites, in our approach we use the concept of *virtual node*. We define *virtual node* as a node which is not part of our actual network, but an imaginary node, which will be connected to all the datacenter sites $s_i$that have a copy of the file f_i. There are *virtual* edges connecting the virtual node to all the sites s_i and there is no cost associated with the channels in these *virtual edges*.

For example, in Fig. 3.1, there is a network of six $nodes$ $(0, 1, 2, 3, 4, 5)$, all of which are actually present in the original *network*. Apart from these six nodes, there is another node i.e. $node\ v$ which is not a part of the original network, but is a *virtual node*. It is clear from the figure that there are two copies of the file f_i which can be requested by any destination node and these two copies are located at $node$ 0 and $node$ 2. Hence, we say that $node$ 0 and $node$ 2 are the two sites containing file f_i. The edges $(v \rightarrow 0)$ & $(v \rightarrow 2)$ are the *virtual* edges and hence are shown with dotted lines.

3.5 Establishment of Primary Lightpath

When we receive a request for communicating file f_i to destination node *t*, the network is already supporting a number of on-going communication requests. The details of all the on-going requests are known to us. Therefore, we already know the routes used by the previous communications to establish their primary lightpaths in fault-free case. We also know the routes and corresponding channels (k) on those routes, used by previous communications to handle each disaster, $d \in D$.

Our objective is to find the primary and the backup paths for establishing the new communication request, such that the $total\ cost$ associated with the network resources will be as low as possible. We calculate the total cost in terms of number of wavelengths (k) used to establish a request's primary path. For establishing a primary lightpath, always a new channel *k* is used on an edge *e* and hence every edge that a primary lightpath traverses, has a cost of 1 associated with it as discussed in Section 3.3.

The primary lightpath must follow the $wavelength\ continuity\ constraint$, i.e. it should use the same channel on every edge it traverses.

As example, we consider the network of Fig 3.2, and assume that there are three available channels ($\lambda_0, \lambda_1, \lambda_2$) on each edge of the network in.

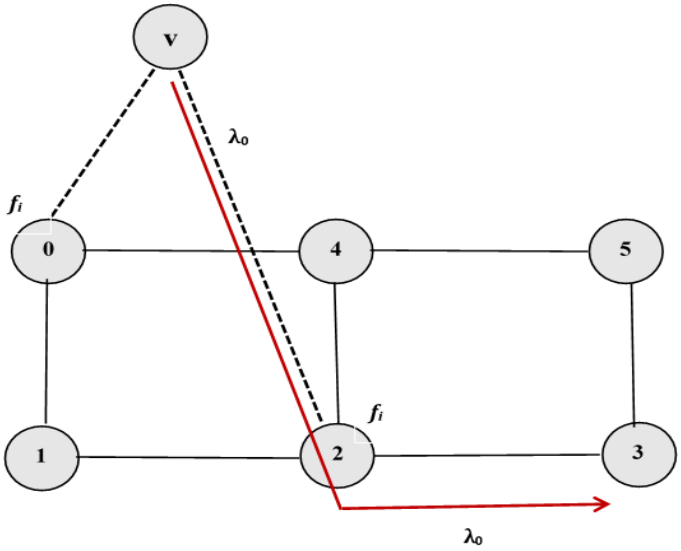


Figure 3.2: Establishment of Primary Lightpath using RWA

If there is a request for communicating file f_i to the destination $node$ 3, one possible primary path (shown in Fig 3.2) is from node *v* (*virtual* node) to node *3* via node *2 (one of the two sites which contain a copy of file* f_i *).* This primary path $v \rightarrow 2 \rightarrow 3$ has a cost of 1 associated with it, since it uses only *one* network edge $2 \rightarrow 3$, and the cost of the virtual edge $v \rightarrow 2$ is 0, as discussed in Sec 3.4. Assuming that all channels on edge $2 \rightarrow 3$ are available, we select the first available channel λ_0, for the primary lightpath. So after establishing this primary lightpath the channel λ_0 on the edge $2 \rightarrow 3$ of the actual network are no longer be available to process any future communication request. But the remaining two channels (λ_1, λ_2) on edge $2 \rightarrow 3$ are still available to be used by any new communication request.

Updating the *edge-channel database* with lightpaths information

After establishing every primary lightpath we update our database namely *"Channel Database"*, with a value of "1" for the channel used on an edge of the primary lightpath. This database contains information of which channel has been used on which edge by any previous primary lightpath. By this updated information we mean that such a channel is no longer available for establishing any new communication request, like in the above example in Fig. 3.2, channel λ_0 on the edge $2 \rightarrow 3$ will be updated with a value of "1" in the *"Channel Database"* and hence the channel λ_0 on edge $2 \rightarrow 3$ won't be available for future communication requests.

This updated channel will also reflect on the *"Channel Database"* that will be used to design the backup paths for all the disasters $d \in D$. For getting a backup path, we must ensure that none of the channels that have been used by some primary lightpath in the previous communications are available to handle any disaster d. We follow the concept of *sharing* channels while designing the backup lightpaths; however such sharing is restricted only among the backup lightpaths and not with any of the primary lightpaths.

3.6 Establishment of Backup Lightpath

Our approach towards designing the backup lightpaths may be viewed as a generalization of $Shared\ Path\ Protection$, according to which two or more communications' backup lightpaths can share the same *channel/wavelength (λ)* among themselves if and only if the backup lightpaths are designed to handle different disasters ($d_1\ and\ d_2$).

Effects of different types of disasters: A disaster d is defined by a set S_d of resources $\{r_1, r_2, \ldots, r_p\}$, where each resource $r_i \in S_d$ represents some component of the network

(e.g. fibers, nodes) that will fail and become faulty due to disaster d . In a wide-area network, there may be an infinite number of disasters occurring at distances much smaller than the length of one optical fiber connecting two nodes. If we consider two disasters d_1 and d_2, it is possible that the effect of disaster d_1 may be more severe than disaster d_2. In this case, we will say that disaster d_1 dominates disaster d_2. This leads to the notion of *dominant* disasters.

Disaster *d* is a *dominant* disaster if there does not exist any other disaster *d',* such that disaster *d'* dominates disaster *d.* We consider that the set of dominant disasters *D* that we have to take into account is predefined, hence we do not define how a particular disaster occurs in our network.

A dominant disaster or Shared Risk Group (SRG) is specified in terms of a set of node and links which are affected simultaneously by a single disaster event. We consider that only one disaster occurs at a given time. For our simulations, we consider disasters only on a single datacenter or any other node in the network. We do not consider disasters that occur directly on link/fibers. If a node is affected by a disaster then all the links/fibers which are connected to that node fails and hence a portion of the network will no longer be available for further communication. We define this kind of disaster, which occurs on a particular node, to be the dominant *disasters* for our simulations, because this kind of disasters have more significant effect on the network than single link disasters as a portion of the network becomes inactive. Based on its location, a disaster may be categorized as follows:

a) A disaster at the site itself where file f_i is located (a datacenter node). To handle such a disaster, the file f_i will be transmitted to the destination node t from another site which has a replicated copy of the file f_i. In our example in Fig. 3.3, file f_i is located at *node 2* and *node 0,* so, to handle the disaster occurring at *node 2,* the backup path can be designed from *node*

0 and it can be $v \rightarrow 0 \rightarrow 4 \rightarrow 5 \rightarrow 3$. By doing this, we now get an alternate route to transmit the file f_i to the destination by avoiding all of the edges *going-into* and *going-out* of datacenter *node 2.*

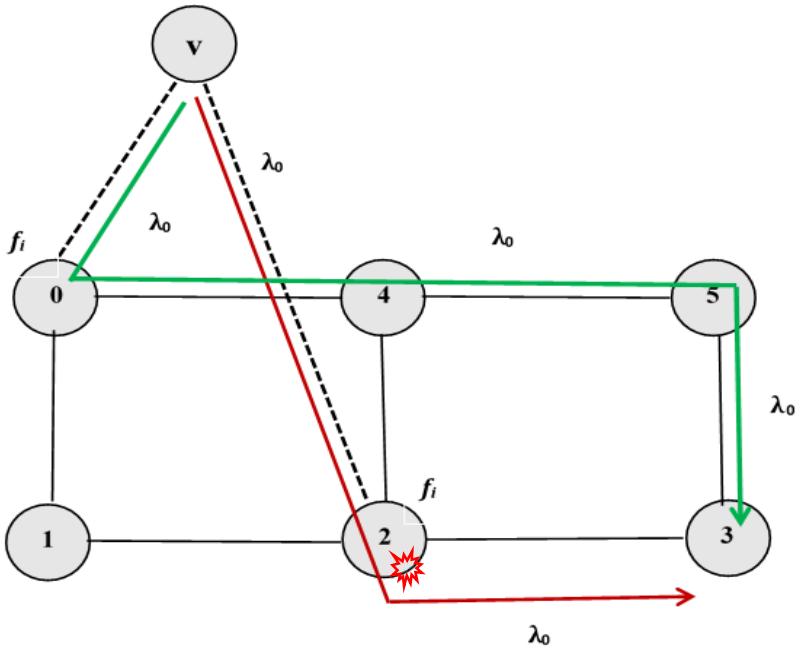


Figure 3.3: Establishment of backup lightpath to handle disaster on a datacenter node

Primary Lightpath	$v \rightarrow 2 \rightarrow 3$
Edges disrupted by disaster at node 2	$2 \rightarrow 4, 4 \rightarrow 2; 2 \rightarrow 1, 1 \rightarrow 2; 2 \rightarrow 3, 3 \rightarrow 2$
Backup Lightpath	$v \rightarrow 0 \rightarrow 4 \rightarrow 5 \rightarrow 3$

Table 3.1: Table showing primary and backup lightpaths

Table 3.1 contains information about the primary and backup lightpaths being used in Fig 3.3. In Fig. 3.3, primary Lightpath is $v \rightarrow 2 \rightarrow 3$. Disaster at *node 2* disrupts edges$(2 \rightarrow 4, 4 \rightarrow 2; 2 \rightarrow 1, 1 \rightarrow 2; 2 \rightarrow 3, 3 \rightarrow 2)$, hence these edges are no longer available for

processing the primary lightpath $v \rightarrow 2 \rightarrow 3$. Backup Lightpath designed to handle disaster at node 2 is $v \rightarrow 0 \rightarrow 4 \rightarrow 5 \rightarrow 3$, where *node 0* is another site of file f_i.

b) A disaster at the destination node itself, i.e, the node requesting file f_i. We say that such a disaster will make the communication request to be failure because if there is a disaster at the destination node itself, then the communication cannot be established.

c) A disaster in an intermediate node of the primary path. Such a disaster can be handled by routing the backup path, possibly from the same source but through different intermediate nodes (avoiding the intermediate node from the primary path that has been affected by disaster d).

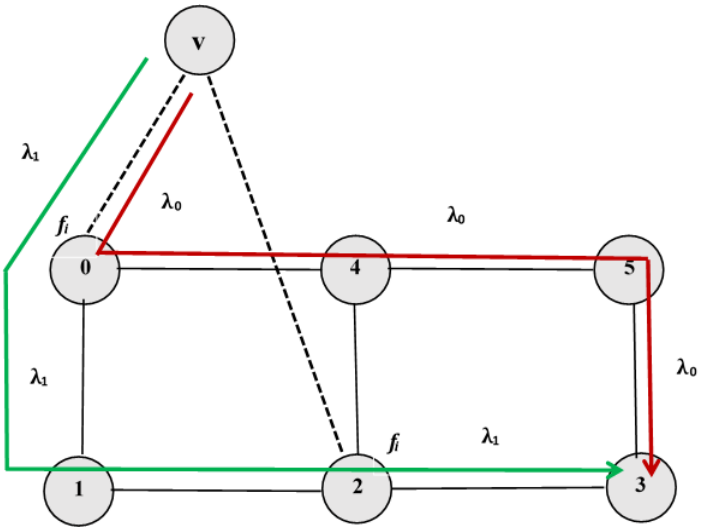


Figure 3.4: Establishment of backup lightpath to handle disaster on an intermediate node

Primary Lightpath	$v \rightarrow 0 \rightarrow 4 \rightarrow 5 \rightarrow 3$
Edges disrupted by disaster at node 4 (intermediate node of primary lightpath)	$0 \rightarrow 4, 4 \rightarrow 0; 4 \rightarrow 5, 5 \rightarrow 4; 2 \rightarrow 4, 4 \rightarrow 2$
Backup Lightpath	$v \rightarrow 0 \rightarrow 1 \rightarrow 2 \rightarrow 3$

Table 3.2: Table showing primary and backup lightpaths

Table 3.1 contains information about the primary and backup lightpaths being used in Fig 3.4. In Fig 3.4, Primary Lightpath is $v \rightarrow 0 \rightarrow 4 \rightarrow 5 \rightarrow 3$. Disaster at *node 4* disrupts edges $(0 \rightarrow 4, 4 \rightarrow 0; 4 \rightarrow 5, 5 \rightarrow 4; 2 \rightarrow 4, 4 \rightarrow 2)$, hence these edges are no longer available for processing the primary lightpath. Backup Lightpath is $v \rightarrow 0 \rightarrow 1 \rightarrow 2 \rightarrow 3$.

3.7 NOTATIONS & ALGORITHMS

In this section we describe the RWA algorithm that has been designed to find a viable primary path in fault-free case and viable backup paths to handle disaster $d \epsilon D$.

We use the following notations in the algorithm;

G: The network topology specified as a graph $G = (N, E)$

N: The set of nodes in the network

E: The set of directed edges of the network. If $i \; and \; j$ are nodes of the network (including datacenters), edge $e \epsilon E$ represents a fiber from node i to node j connecting the two nodes

K: The set of channels k per fiber, where $k \epsilon K$

D: The set of disasters d, where $d \epsilon D$

$\boldsymbol{N_d}$: The set of nodes disrupted due to disaster d

$\boldsymbol{v}$: The virtual node

S **:** The set of datacenter nodes s_i containing a copy of file f_i.

$(\boldsymbol{f_i}, \boldsymbol{t})$: A request where file f_i is the file that is requested and node t is the node requesting the file f_i . When the request is handled, a copy of file f_i from one of the datacenter sites s_i , $s_i \epsilon S$ that contain file f_i will be communicated to destination node t.

NW_ST = The state of the network in terms of the availability of each channel $k \epsilon K$ on each edge $e \epsilon E$. NW_ST can be characterized in terms of the two parameters *state(e, k)* and *disasters(e,k)*, which are defined below, for each channel *k* on each edge *e*.

$\boldsymbol{state}\ (\boldsymbol{e}, \boldsymbol{k})$=

$$\begin{cases} 1, & if\ channel\ k \in K\ on\ edge\ e \in E \text{ has not been used by any previous communication} \\ 2, & if\ channel\ k \in K\ \ on\ edge\ e \in E \text{ has been used previously for a primary path} \\ 3, & if\ channel\ k \in K\ \ on\ edge\ e \in E \text{ has been used previously for a backup path} \end{cases}$$

$\boldsymbol{disasters}\ (\boldsymbol{e}, \boldsymbol{k})$ = The set of disasters $d \in D$ that are handled using channel k on edge e, for any communication request. We note that $disasters\ (e, k)$ = [ ] whenever $state\ (e, k) = 1\ or\ 2$; and is non-empty whenever $state(e, k) = 3$.

$\boldsymbol{INF}$: A very large value for edge-cost indicating the edge cannot be used.

$\boldsymbol{LP}$: The lightpath (primary or backup).

3.7.1 Disaster Aware Dynamic RWA Algorithm

In this section we discuss our heuristic for solving the dynamic RWA problem. Our algorithm tries to allocate primary lightpath and backup lightpaths to a given communication request. A communication request is defined as a request to communicate a file f_i to a node *t*. We call the file f_i as the requested file and the node t as the destination node. Each communication request is denoted as (f_i, t), and for every such communication request, our algorithm finds out a viable primary path and if needed, a viable backup path using dynamic lightpath allocation.

Given a physical network topology $G = (N, E)$, of N nodes interconnected with E bidirectional edges. We are also given set of K channels on each edge of the network ($k \in K$), set of D disasters, $(d \in D)$, such that a single disaster d occurs at a time in a node and as a result the edges entering and leaving that node are disrupted due to disaster d. We are using an optimal replication strategy for file f_i which saves a copy of each file f_i at multiple locations, where each of these locations is a data center. If there are n number of copies of file f_i, we know that

the copies of the file f_i are saved at data centers $(s_i^1, s_i^2, s_i^3, \ldots\ldots s_i^n)$. The replication strategy is such that, for each disaster $d \epsilon D$, at least one copy of each file f_i has a fault-free path to each node in the network that avoids disaster d.

Step 1: $NW\_ST\_CP \leftarrow$ a copy of the current network state NW_ST

Step 2: $request \leftarrow$ read a request from RF $(f_{i\,,}\ t)$

Step 3: *path, k, cost = find_primary_LP* ($(f_{i\,,}\ t), NW\_ST\_CP$)

Step 4: IF *cost* < INF

Step 5: Update NW_ST_CP

a) $state\ (e, k)$= 2, for each edge $e \epsilon E$

Step 6: ELSE

Return REQUEST BLOCKED

Step 7: For each disaster $d \epsilon D$ that affects the primary path

Step 8: *path, k, cost = find_backup_LP* $(f_{i\,,}\ t), NW\_ST\_CP,\ d)$

Step 9: IF *cost* < *INF*

a) Update $NW\_ST\_CP\ state\ (e, k)$ = 3, for each edge e in the backup path that we get from Step 6

b) $disasters\ (e, k)$ = $disasters\ (e, k)$ U d

Step 10: ELSE

Return REQUEST BLOCKED

Step 11: $NW\_ST\ =\ NW\_ST\_CP$

Step 12: Return SUCCESS.

Figure 3.5: Overview of disaster-aware RWA algorithm

Fig 3.5 shows an overview of the RWA algorithm that we have designed for processing each communication request $(f_{i\,,}\ t)$ by setting up a primary lightpath in fault-free case and backup lightpaths to handle each disaster $(d \epsilon D)$. Initially we consider that our network (NW_ST), is completely free from communication, which means that all the channels are

available on all the edges to handle any communication request $(f_{i,}\, t)$. In other words, $state(e,k) = 1$ for all channels $k \epsilon K$ on all edges $e \epsilon E$.

In the Step 1 of our algorithm, we make a copy NW_ST_CP of the current network state NW_ST and update the copied version, NW_ST_CP, after processing every communication request $(f_{i,}\, t)$, so that the actual network state (NW_ST) remains the same in case if a communication request $(f_{i,}\, t)$ gets blocked. A communication will be blocked in two cases; i) if the algorithm is unable to find a viable route using a channel k to establish the primary lightpath, ii) if the algorithm is unable to find viable route using a channel k to set up a backup lightpath to handle disaster d . In Step 3, for a new communication request$(f_{i,}\, t)$, when the new communication request arrives, the RWA algorithm calls the *find_primary_LP* function. The algorithm for setting up the primary lightpath using the function *find_primary_LP* is further explained in Section 3.7.2. Step 4 checks if a suitable primary lightpath has been found. If so, information about available resources is updated by updating NW_ST_CP and the algorithm continues to try to establish suitable backup lightpath(s) as needed. Otherwise, the communication request $(f_{i,}\, t)$ gets blocked (Step 6) and no resources are allocated to this request. In Step 7 - Step 10, the algorithm considers each disaster $(d \epsilon D)$ affecting the primary lightpath, one by one. For each such disaster, it tries to find a suitable backup lightpath (Step 8). If it is successful (Step 9), it updates *NW_ST_CP* to indicate which channels were used for the backup lightpath and then proceeds to find a path to handle the next disaster. If at any point, a suitable backup lightpath cannot be found (for any disaster), the entire communication request is blocked (Step 10). If for a communication request $(f_{i,}\, t)$, our algorithm finds both the primary lightpath and required backup lightpaths to handle each disaster successfully, then the corresponding resources are allocated and the actual network state is updated (Step 11), based

on the allocated resources. In this case, the algorithm returns, indicated that the new communication request was successfully established (Step 12).

3.7.2 Dynamic RWA algorithm to find Primary Path

Step 1: Initialize the parameters.

a) $cost = INF$

b) $path = [\ ]$

c) $channel = -1$

d) v = virtual node for the request (f_i, t)

Step 2: FOR each channel $k \epsilon K$

Step 3: FOR each edge $e \epsilon E$

Step 4: Calculate $cost$ of edge $e \epsilon E$

a) $cost\ (e,k) = 1, if\ state\ (e,k) = 1$

b) $cost\ (e,k) = INF, otherwise$

Step 5: c_k = cost of shortest path p_k based on $cost$ of edge

Step 6: p_k = shortest path from v to t

Step 7: IF $c_k < cost$

a) $path = p_k$

b) $k = k_k$

c) $cost = c_k$

Step 8: Return $path, channel, cost$

Figure 3.6: Overview of function *find_primary_LP*

Fig. 3.6 shows an overview of the function *find_primary_LP* (used in Step 3 of Fig. 3.5) used to find a suitable route and channel for establishing the primary lightpath to handle a communication request (f_i, t). Step 1 of the algorithm initializes the values of the three variables used to identify the primary lightpath. These are:

- *cost*: This is initialized with a very large value *INF*. This variable is used to indicate the cost of the primary path in order to check if it is a viable primary path.
- *path*: This is used to store the actual route to be used by the primary lightpath. The path is specified in terms of a sequence of network nodes starting from the selected data center s_i, containing a copy of the requested file f_i , to the destination node *t*.
- *channel*: This is used to specify the channel *k* that will be used along each edge of the primary lightpath.

When processing a request for transmitting file f_i to a destination node t, it is convenient to visualize a virtual node v and some new virtual edges from v to set of datacenter nodes $s_i \in S$. For each data center node s_i , containing a copy of file f_i , we visualize a single virtual edge from virtual node v to datacenter s_i . When considering a communication request (f_i, t), we note that the length of these virtual edges will be 0.

Once we add these virtual nodes and edges in our network, in Step 2 to Step 4, for each channel $k \in K$ on each edge $e \in E$ of the network, the algorithm calculates the cost of edge *e* based on the channel usage on an edge. In Step 4, the *cost* of using channel *k* on an edge *e* is calculated, where, $cost\ (e,k)$ is 1 if $state\ (e,k)$ is 1. This means that channel *k* on edge *e* is available for setting up a primary lightpath but the cost of using that channel on that edge will be 1. Otherwise, $cost\ (e,k)$ is *INF* which means that channel $k \in K$ is unavailable on the edge *e* and hence cannot be used to set up a lightpath. Using these edge costs *cost (e, k)*, the algorithm calculates the cost of the shortest path from *v* to *t*, using channel *k*, for each $k \in K$ (Step 5 and 6). If the cost of the primary path using channel *k* is lower than the best route found so far, the path, channel and cost are updated accordingly (Step 7). Once all the channels have been checked Step 8 returns the best route, along the corresponding channel and cost. We note that if no feasible

routes could be found, using any of the channels, then the function will return the initial values assigned in Step 1.

3.7.3 Dynamic RWA algorithm to find Backup Path

Step1: Initialize the parameters.

a) $cost = INF$

b) $path = [\ ]$

c) $channel = -1$

d) v = virtual node for request (f_i, t)

Step 2: FOR each channel $k \epsilon K$

Step 3: FOR each edge $e \epsilon E$

Step 4: Calculate $cost$ of edge $e \epsilon E$

IF edge $e \epsilon E$ is affected by disaster $d \epsilon D$

a) $cost\ (e,k) = INF$

ELSE

b) $cost\ (e,k) = 1$, if $state\ (e,k) = 1$

c) $cost\ (e,k) = INF$, if $state(e,k) = 2$ or d in $disasters\ (e,k)$

d) $cost\ (e,k) = 0$, otherwise

Step 5: c_k = $cost$ of shortest path b_k based on $cost$ of edge

Step 6: b_k = shortest path from v to t

Step 7: IF $c_k < cost$

a) $path = b_k$

b) $k = k_k$

c) $cost = c_k$

Step 8: Return $path, k, cost$

Figure 3.7: Overview of function *find_backup_LP*

After getting the primary lightpath for the communication request(f_i, t), the RWA algorithm in Fig. 3.7 shows an overview of the function *find_backup_LP* (used in Step 8 of Fig.

3.5) used to find a suitable route and channel for establishing the backup lightpath to handle a communication request (f_i , t). Step 1 of the algorithm initializes the values of the three variables used to identify the backup lightpath. These are:

- *cost*: This is initialized with a very large value *INF*. This variable is used to indicate the cost of the backup path in order to check if it is a viable backup path.
- *path*: This is used to store the actual route to be used by the backup lightpath. The path is specified in terms of a sequence of network nodes starting from the selected data center s_i, containing a copy of the requested file f_i , to the destination node *t*.
- *channel* : This is used to specify the channel *k* that will be used along each edge of the backup lightpath.

Just like the primary lightpath, when processing a request for transmitting file f_i to a destination node t, it is convenient to visualize a virtual node v and some new virtual edges from v to set of data center nodes $s_i \in S$. Once we add these virtual nodes and edges in our network, in Step 2 - Step 4, for each channel $k \in K$ on each edge $e \in E$ of the network, the algorithm calculates the cost of edge *e* based on the channel usage on an edge. The cost of an edge e i.e., $cost\ (e,k)$ is *INF* if disaster d does affects any edge of the backup lightpath, otherwise for every disaster d that affects the primary lightpath, the *cost* of using channel *k* on a disaster-free edge is calculated (step 4), where, $cost\ (e,k)$ is 1 if $state\ (e,k)$ is 1 which means that channel *k* on edge *e* is available for setting up a backup lightpath, but the cost of using that channel on that edge will be 1. The cost of edge, $cost\ (e,k)$ is *INF*, if $state\ (e,k)$ is 2, which means that the channel *k* on the edge *e* has been used previously by some primary lightpath or if disaster d is in the set $disasters\ (e,k)$, which means that the channel *k* on the edge *e* has been used previously to set up a backup lightpath but to handle the same disaster *d*

from the set of disasters *D* and hence cannot be shared to set up the current backup lightpath. Otherwise the cost of edge, $cost\ (e,k)$ will be *0*, which means that the channel *k* on the edge *e* has been used previously to set up a backup light but for a different disaster and hence the channel *k* can be shared because both the backup lightpaths handle different disasters $d \epsilon D$. Using these edge costs, $cost(e,k)$, the algorithm calculates the cost of the shortest path from *v* to *t*, using channel *k*, for each $k \epsilon K$ (Step 5 and 6). The cost of the backup lightpath (c_k) is calculated based on the minimum number of new channels being used to set up the backup lightpath, as the algorithm tries to *share* as many channels as possible on edges through which it can set up the backup lightpath. If the cost of the backup path using channel *k* is lower than the best route found so far, the path, channel and cost are updated accordingly (Step 7). Once all the channels have been checked, Step 8 returns the best route, along the corresponding channel and cost. We note that if no feasible routes could be found, using any of the channels, then the function will return the initial values assigned in Step 1.

Chapter 4

EXPERIMENTATION AND RESULTS

In this section we discuss our network model, the experiments performed and the results we obtained using our heuristic approach. We present our simulation results for evaluating the performance of the proposed strategy under different disaster scenarios considering different sets of data center nodes in our network. Results reported in this chapter are the average of 5 different runs for each topology, considering different number of data centers and disasters. Our algorithm is able to produce results for practical sized networks. All the simulations were carried out on Intel Core i5 - 4300U CPU 1.9 GHz processor using an integrated development environment (IDE) of Eclipse (version - Mars).

4.1 Simulation setup

4.1.1 Network topology

In this work we have considered several standard and widely used network topologies to perform experiments using our heuristic approach. The size of these topologies range from 6 nodes to 24 nodes; the topologies used include, the 11-node COST-239 network shown in Fig 4.1 [43], the 14-node NSFNET shown in Fig 4.2 [44], the 20-node ARPANET network shown in Fig 4.3 [45] and the 24-node USANET network shown in Fig 4.4 [44].

For each topology, we have performed experiments for different sets of communication requests ranging from 20 to 40 communication requests from a large number of requests available in the request file. The request file that we gave as an input to our algorithm actually contained a total of 200 communication requests. All these 200 requests were generated randomly from the set of 10 files (f_i) and the other nodes of the network which did not belong to the set of data center nodes containing file f_i using random number generator. The simulation was run 5 times with 5 different request files. The results reported are the average of the 5 different simulation runs for each topology considering different source-destination requests every time.

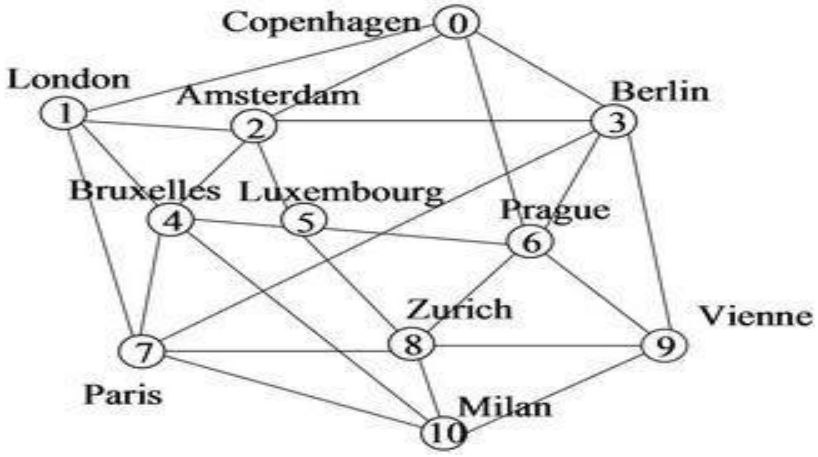


Figure 4.1: COST-239 network (11 - node topology)

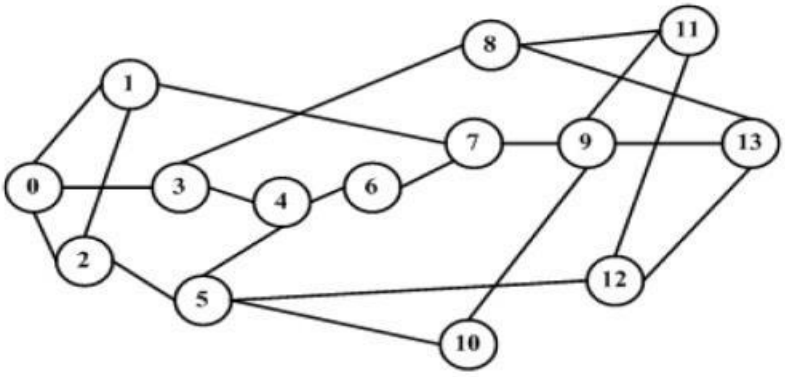


Figure 4.2: NSFNET network (14 - node topology)

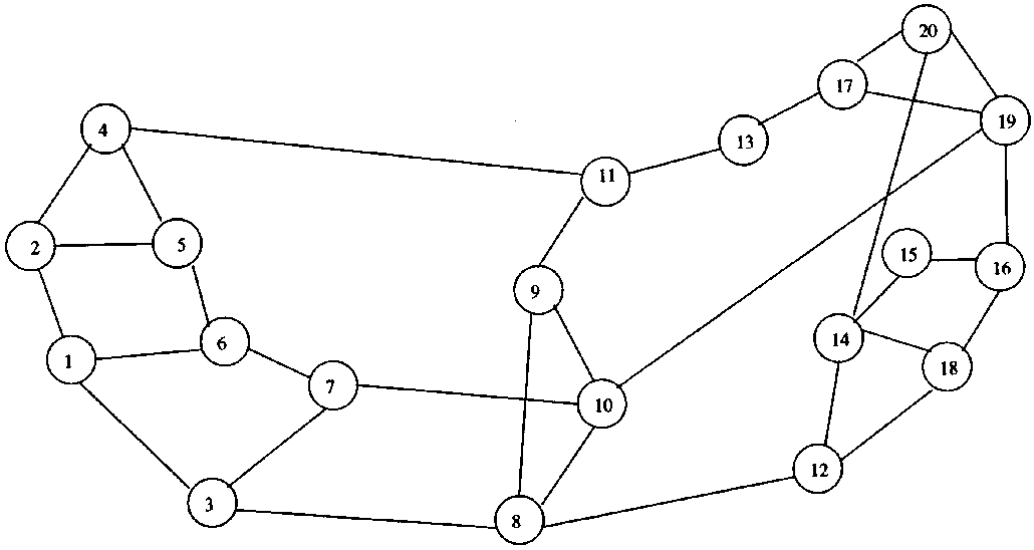


Figure 4.3: ARPANET network (20 - node topology)

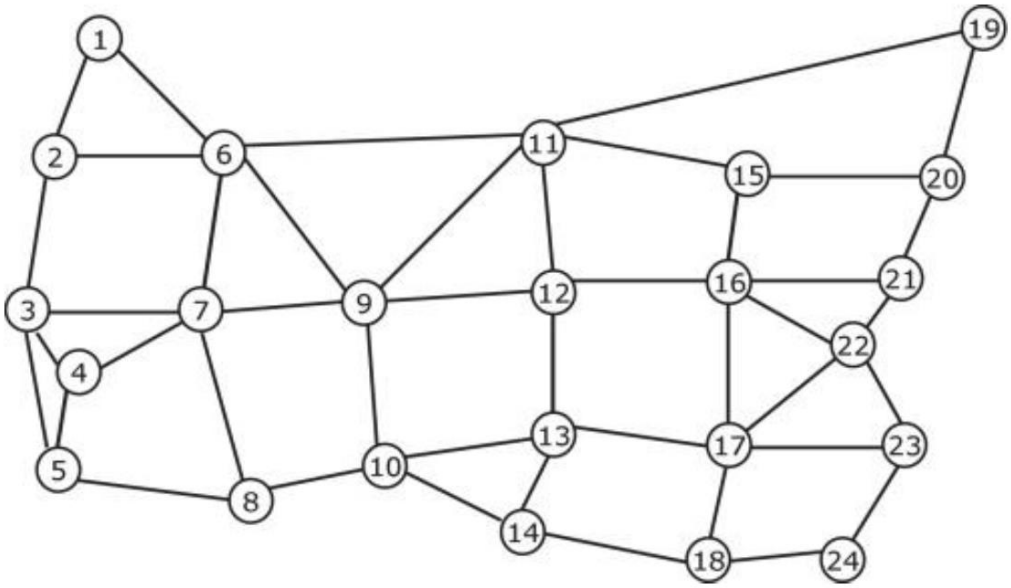


Figure 4.4: ARPANET network (24 - node topology)

4.1.2 Algorithm inputs

The following parameters were given as inputs to our RWA algorithm:

i) **Network Topology:** The network topology (N, E) consists of a defined set of nodes (N) and fiber links (E) connecting the nodes of the network.

ii) **File-Replication Information:** We assume that copies of the relevant content (i.e. the files) have been distributed at selected data centers, based on some pre-defined replication strategy. For our simulations, we have used a simple replication strategy that places each file f_i at a minimum number of data centers such that,

- There is at least one copy of f_i available under all disaster scenarios.
- There is at least one fault-free path available from a surviving copy of f_i to each surviving node, under all disaster scenarios.

iii) **Request-file:** The request-file consists of a large number of communication requests from a source to a destination node. The source is a file f_i located at some data center node and the destination is a node that requests a copy of the file f_i. We ensure that the destination is never one of the data center nodes that contain f_i.

iv) **Disaster Nodes:** We consider different set of disasters and assume that one disaster occurs from the set of disasters at a time. When a disaster occurs, the edges entering or leaving the node affected by the disaster are not available for that particular disaster to process a communication request.

4.2 Performance evaluation of Phase I & II

The simulations are carried out in 2 phases. During Phase I, we try to accommodate a given number of communication requests using our algorithm. We consider a database which contains information about the optical fibers connecting the network nodes and the wavelengths

on each fiber. We name this database as *edge-channel database* and it contains information about each wavelength's availability on each fiber connecting the nodes of the network. Our algorithm tries to find a primary path to handle the communication in the fault-free case and the backup paths to handle the communication to handle each disaster. If a communication request is able to find a route and an available channel (wavelength) on that route using some *Routing and Wavelength Assignment* algorithm of WDM network, then the communication request is successful and resources are allocated to handle the request. We update the resource allocation information in the database (*edge-channel database*). A communication request (f_i , t) specifies the file f_i is to be communicated to the requesting node *t*. File f_i is available from any data center that has a copy of the file. A connection request (f_i , t) will be successful only if;

1) A primary lightpath can be established from a data center having a copy of f_i to the requesting node *t*.

2) Resources are available for backup lightpaths from a data center having a copy of f_i (not necessarily the same data center used as the source of the primary lightpath) to *t* under all possible disasters that disrupt the primary lightpath. Resources for the backup lightpaths may be shared with those for other backup lightpaths, as explained in Sec 3.2.

During Phase II, we assume that the network has a set of ongoing communications already established from Phase I. We randomly generate new communication requests and report the blocking probability for establishing a new connection with a set of existing connections already established over the network. We considered two categories of disasters as follows:

- Category I: Disasters can only affect the DC nodes
- Category II: Disasters can affect any node in the network

In the following subsections, we discuss our simulation results. Each value is obtained by taking the average over 5 simulation runs.

Table 4.1: Comparison of Avg. BP in COST-239 network (8 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability (BP)of 5 Runs in Phase II
11 Node COST-239 Network / 8 channels	[0, 5]	I	27	0.00
			31	0.01
			32	0.25 (high)
		II	27	0.00
			31	0.03
			32	0.27 (high)
	[0, 5, 10]	I	59	0.00
			61	0.02
			63	0.125 (high)
		II	59	0.00
			61	0.07
			63	0.36 (high)

In Table 4.1, we have shown the simulation results for the 11 node COST-239 network. As an input to run the simulation, we have provided the network topology, the $network\_state$ from Phase I, information about the number of channels per fiber in the overall network (for this experiment we considered 8 channels per fiber and there were a total of 26 bidirectional fibers connecting the nodes of the network), the number of data center nodes, where we considered two

sets of data centers, one set containing data center nodes [*node0, node5*], another set containing data center nodes [*node0, node5, node10*]. For data center nodes [*node0, node5*], when we considered the Category I disasters and ran a total of 27 requests from each of the 5 request files in Phase I, we got an average BP of 0.00, which otherwise means that none of requests in Phase II were blocked during this case. We then increased the number of requests in Phase I slightly up to 31 and we got an average BP of 0.01 in Phase II, which is quite an acceptable BP and means that some of the requests from the total number of requests in Phase II were blocked. We have also reported that the average BP that was quite high when we ran a total of 32 requests from each of the 5 request files in Phase I. From this experiment we can say that the BP in Phase II is highly affected even with a slight increase in the number of static requests in Phase I.

Similarly when we ran our simulation for the same set of data centers nodes [*node0, node5*] but with disasters possible at any node (i.e. Category II), we got an average BP of 0.05 for 31 requests. From this result we can say that, keeping the number of requests the same (i.e. 31) in Phase I, but by increasing the number of disasters from Category I to Category II, increased the BP (from 0.01 to 0.05), but a more important factor affecting the BP is the number of ongoing connections currently established on the network.

We performed the next set of experiments in a similar way, but this time we increased the number of data center nodes from 2 data centers [*node0, node5*] to 3 data centers [*node0, node5, node10*]. From this experiment it is clear that increasing the number of data center nodes in the network significantly reduces the BP, since there is more flexibility in terms of how the RWA may be carried out.

Table 4.2: Comparison of Avg. BP for COST-239 network (16 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
11 Node COST-239 Network / 16 channels	[0, 5]	I	32	0.00 (was high in 8 channels)
			60	0.00
			63	0.02
			65	0.20 (high)
		II	32	0.00 (was high in 8 channels)
			60	0.01
			63	0.07
			64	0.41 (high)
	[0, 5, 10]	I	63	0.00 (was high in 8 channels)
			120	0.00
			124	0.02
			125	0.25 (high)
		II	63	0.00 (was high in 8 channels)
			120	0.00
			124	0.09
			125	0.47 (high)

Table 4.2 shows simulation results for the same COST 239 network topology, but with 16 available channels per fiber. When we had 8 channels per fiber, we were able to accommodate a total of 31 requests in Phase I to get a minimum BP in Phase II (Table 4.1), but in the scenario considered in Table 4.2, we are now able to accommodate more requests in Phase I i.e. 63

requests, for which we get the average BP of five runs to be 0.02. It is clear that increasing the number of channels allows us to accommodate more communication requests during Phase I, before we see any blocked connections during Phase II. However a similar trend is evident (as for the case with 8 channels per fiber), where there seems to be a threshold for number of ongoing connections. If the number of established connections is increased beyond this threshold for Phase I, then the BP increases sharply in Phase II.

When we compare results in Table 4.2 with results in Table 4.1, we can see that when we has considered 32 requests in Phase I with 8 channels per fiber in Table 4.1 (2 Data center locations) the average BP was quite high, which was 0.25, but as we increased the number of channels in the same COST 239 network from 8 to 16 channels per fiber, the average BP for 32 requests was still 0.00. Similarly, we have reported the results that we go by considering 3 data center locations in our network, data centers [*node0, node5, node10*]. Increasing the number of data centers, our network was able to accommodate almost double the number of requests in Phase I, than it was able to accommodate for 2 number of data centers [*node0, node5*]. Considering 3 data center locations and a total of 124 requests in Phase I, we were able to achieve a an average BP of 0.02 in Phase II, but as soon as we increased the number of requests in Phase I to 125, the average BP increased sharply, which we had also noticed in the results in Table 4.1.

Tables 4.3 & 4.4, 4.5 & 4.6 and 4.7 & 4.8 shows the simulation results for the 14-node, 20-node and 24-node topologies respectively. For each topology, the first Table (i.e. Table 4.3, 4.5 and 4.7) reports the results with 8 available channels per fiber and the second Table (i.e. Table 4.4, 4.6 and 4.8) reports the results with 16 available channels per fiber. The results for these topologies follow the same pattern as for the 11-node topology. However, the actual

number of requests that can be accommodated, and the threshold at which the BP increases steeply are different in each case.

Table 4.3: Comparison of Avg. BP for 14-node NSFNET network (8 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
14 Node NSFNET Network / 8 channels	[5, 9]	I	25	0.00
			26	0.01
			28	0.15 (high)
		II	25	0.01
			26	0.03
			28	0.22 (high)
	[0, 5, 9]	I	40	0.00
			42	0.04
			43	0.36 (high)
		II	40	0.00
			42	0.09
			43	0.42 (high)

In Table 4.3 we have reported the simulation results for 14-node NSFNET network by considering 8 channels per fiber. There were a total of 21 bidirectional fibers connecting the 14-node NSFNET network nodes. Like the 11-node COST 239, network, a similar trend has been followed in the experimentation with 14-node NSFNET. If we compare the results in Table 4.3 with Table 4.1, we can see that the number of requests that are handled in Phase I of Table 4.3

(for both the set of data center locations) is lesser than the number of requests that are handled in Phase I of Table 4.1 (for both the set of data center locations). From this comparison we understand that, even after increasing the number of nodes in our network from 11 (COST-239) to 14 (14 – NSFNET), our network was able to accommodate fewer requests, because the number of edges (fibers) were more in the 11- node network as compared to the 14 - node network.

Table 4.4: Comparison of Avg. BP for 14-node NSFNET network (16 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
14 Node NSFNET Network / 16 channels	[5, 9]	I	28	0.00 (was high in 8 channels)
			45	0.00
			47	0.03
			49	0.27 (high)
		II	28	0.00 (was high in 8 channels)
			45	0.01
			47	0.04
			49	0.42 (high)
	[0, 5, 9]	I	43	0.00 (was high in 8 channels)
			78	0.00
			80	0.01
			82	0.17 (high)
		II	43	0.00 (was high in 8 channels)
			78	0.03
			80	0.02
			82	0.20 (high)

In Table 4.4 we have reported the simulation results for 14-NSFNET by considering 16 channels per fiber. According to the simulations that we ran, the average BPs reported were 0.03

and 0.04 when we considered 47 requests, respectively for Category I and Category II disasters, considering 2 data center nodes [*node 5, node 9*]. As expected, when we increased the number of data center nodes from 2 to 3 [*node 0, node 5, node 9*], our network was able to accommodate more requests in Phase I i.e. 80 requests and still reporting lesser BPs of 0.01 and 0.02 respectively for Category I and Category II disasters.

Table 4.5: Comparison of Avg. BP for 20-Node ARPANET network (8 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
20 Node ARPANET Network / 8 channels	[1, 7, 18]	I	19	0.00
			20	0.01
			23	0.15 (high)
		II	19	0.02
			20	0.05
			23	0.46 (high)
	[1, 7, 10, 18]	I	31	0.00
			32	0.07
			34	0.46 (high)
		II	31	0.03
			32	0.09
			34	0.55 (high)

In Table 4.5 we have reported the simulation results that we got after running our algorithm on 20 - Node ARPANET network, considering 8 channels per fiber. For this experiment we had total 32 bidirectional edges connecting the 20 nodes of the ARPANET

network. But unlike the previous experiments of COST-239 and 14-Node NSFNET, we incremented the number of disasters in 20 – Node ARPANET network, due to which the number of backup lightpaths had increased in Phase I and hence the number of requests handled were less. Our experiments reported an average BP of 0.01 and 0.07, when we considered 20 requests in Phase I respectively for Category I and Category II disasters, considering 3 data center nodes [*node 1, node 7, node 18*] and the average BP was 0.07 and 0.09 when we considered 32 requests in Phase I respectively for Category I and Category II disasters considering 4 data center nodes [*node 1, node 7, node 10, node 18*].

In Table 4.6 below, we have reported the simulation results that we got after running our algorithm on 20 - Node ARPANET network, considering 16 channels per fiber. Since we increased the number of channels from 8 to 16, our network was able to handle more requests. In Phase II, our experiment reported an average BP of 0.04 when we considered 60 requests in Phase I, and when we slightly reduced the number of requests to 59, we got an average BP of 0.02. This reduction in the number of requests was due to the increase in the number of disasters from Category I to Category II. When we do an overall comparison of Table 4.6 (20 Node ARPANET network/ 16 channels) with Table 4.5 (20 Node ARPANET network/ 8 channels), we see that the number of requests handled in Table 4.6 is more than that in Table 4.5 due to the increment in the number of channels per fiber.

Table 4.6: Comparison of Avg. BP for 20-Node ARPANET network (16 channels per/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
20 Node ARPANET Network / 16 channels	[1, 7, 18]	I	23	0.00 (was high in 8 channels)
			59	0.00
			60	0.04
			62	0.47 (high)
		II	23	0.00 (was high in 8 channels)
			59	0.02
			60	0.10
			62	0.47 (high)
	[1, 7, 10, 18]	I	34	0.00 (was high in 8 channels)
			77	0.00
			81	0.03
			82	0.22 (high)
		II	34	0.00 (was high in 8 channels)
			77	0.00
			81	0.09
			82	0.34 (high)

Table 4.7: Comparison of Avg. BP for 24-Node USANET network (8 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
24 Node USANET Network / 8 channels	[0, 8, 15]	I	24	0.00
			40	0.04
			41	0.42 (high)
		II	24	0.03
			40	0.09
			41	0.55 (high)
	[0, 8, 15, 23]	I	44	0.00
			51	0.05
			52	0.20 (high)
		II	44	0.01
			51	0.07
			52	0.71 (high)

In Table 4.7, we have reported the simulation results after we carried out similar experiments in 24 - Node USANET network, considering 8 channels per fiber.

In Table 4.8 below, we have reported the simulation results after we carried out similar experiments in 24 - Node USANET network, considering 16 channels per fiber. The average BP reported followed a similar trend like the previous experiments. By comparing the experimental results from 24 - Node USANET network with previous experiments, we conclude that the location of the data centers in a network also plays a vital role in accommodating number of request in the network.

Table 4.8: Comparison of Avg. BP for 24-Node USANET network (16 channels/fiber)

Network Topology	Data center locations	Disaster Category	No. of requests in Phase I for 5 Runs	Average Blocking Probability of 5 Runs in Phase II
24 Node USANET Network / 16 channels	[0, 8, 15]	I	41	0.00 (was high in 8 channels)
			87	0.01
			88	0.15
			89	0.20 (high)
		II	41	0.00 (was high in 8 channels)
			87	0.04
			88	0.17
			89	0.36 (high)
	[0, 8, 15, 23]	I	52	0.00 (was high in 8 channels)
			90	0.00
			92	0.03
			93	0.23 (high)
		II	52	0.00 (was high in 8 channels)
			90	0.00
			92	0.04
			93	0.42 (high)

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In today's world huge amount of important data are stored in the data centers and these data are transmitted to the users across the entire network through different communication schemes. Natural disasters and man-made attacks cause large scale damage to these communication networks. When such disasters occur, specific geographic area of the communication networks are affected, as a result of which, the network components like nodes and fibers (links) gets disrupted, which in turn leads to failure of communications across the network. To handle such situations, there is a need to develop an efficient communication scheme to handle requests for communication that includes provisions to handle disaster. In this work we have presented a new approach to the problem of developing efficient communication schemes to handle dynamic communication requests in data center (DC) networks. Our approach provides an efficient heuristic based solution to this problem. To the best our knowledge, this is the first work to provide a heuristic solution to the problem of handling dynamic communication requests in data center networks. The main objective of our work is to minimize the average blocking probability (BP) for the new communication requests that arise dynamically in a network when the network already has some pre-existing communications going on in it, while trying to minimize the overall resource (channels) usage

for each communication request. This is achieved by sharing resources (channels) to the maximum possible extent while allocating the backup lightpaths. To incorporate the sharing scheme, we considered that the backup lightpaths will be used if and only if a disaster affects the primary lightpaths. Since, the backup lightpaths are not active always, hence they can share resources among themselves when required. We ran our simulation on several network topologies, and with different sets of communication requests. Our approach reported the average blocking probability (BP) under different traffic conditions in the network, different number disasters and different number of channels per fiber.

5.2 Future Work

Directed attacks or natural disasters pose a serious threat to the safety of user data which are located in the data centers, hence making disaster survivability in communication networks a major challenge. The advent of cloud services delivered by data center networks gives novel opportunities to provide protection against disasters in a cost-effective way. In this work, we have presented an efficient heuristic to ensure the survivability of a dynamic communication request in presence of a disaster. Since our objective was to establish dynamic lightpaths (primary and backup), using minimum number of fibers, hence while generating the lightpaths using routing and wavelength assignment, we had to ignore the constraint of optical reach. In well-connected networks, our approach is guaranteed to identify survivable solutions. A promising direction to our future work can include, handling static communication requests by taking into account the constraint of optical reach.

BIBLIOGRAPHY/REFERENCES

[1] K. Chen, C. Guo, H. Wu, J. Yuan, Z. Feng, Y. Chen, S. Lu and W. Wu, "DAC: generic and automatic address configuration for data center networks," IEEE/ACM Trans. Networking, vol. 20, no. 1, pp. 84-99, 2012.

[2] M. Bari, R. Boutaba, R. Esteves, L. Granville, M. Podlesny, M. Rabbani, Q. Zhang, and M. Zhani, "Data center network virtualization: a survey," IEEE Communications Surveys & Tutorials, pre-published, pp. 1-20, 2012.

[3] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "The Resilience of WDMNetworks to Probabilistic Geographical Failures," in Proc. IEEE INFOCOM, Shanghai, China, Apr. 2011, pp. 1521–1529.

[4] Jun Zheng & Hussein T.Mouftah, " Optical WDM networks, concepts and Design," IEEE press, John Wiley –Sons, Inc, Publication, vol. 3, no. 4, pp.1-4, April 2013.

[5] H. Zang, J. P. Jue, and B. Mukherjee. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. Optical Networks, 1(1):47{60, January 2000.

[6] Bandyopadhyay, S., Jaekel, A. & Varanasi, S., (2014). Impairment-Aware Dynamic Routing and Wavelength Assignment in Translucent Optical WDM Networks. In Distributed Computing and Networking (pp. 363-377). Springer Berlin Heidelberg.

[7] Saradhi, C., and Subramaniam, S. 2009. Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges. Communications Surveys and Tutorials, IEEE 11, 4, 109-130.

[8] Z. J. Mouftah, H., Optical WDM Networks:Concepts and Design Principles. Wiley-IEEE Press, 2004.

[9] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks, Part I—Protection," in Proc. IEEE INFOCOM'99, vol. 2, New York, NY, Mar. 1999, pp. 744–751.

[10] Fred B. Schneider. Implementing Fault-Tolerant Services Using the State Machine Approach: A Tutorial. ACM Computing Surveys, 22(4):299–319, 1990.

[11] F. L. Presti, C. Petrioli, and C. Vicari, "Dynamic replica placement in Content delivery networks," in Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2005. 13th IEEE International Symposium on, pp. 351–360, IEEE, 2005.

[12] Gagnaire, M, Zahr, S., Impairment-aware routing and wavelength assignment in Translucent networks: state of the art- *Communications Magazine, IEEE, 2009 -ieeexplore.ieee.org.*

[13] T. A. Neves, L. M. Drummond, L. S. Ochi, C. Albuquerque, and E. Uchoa, "Solving replica placement and request distribution in Content distribution networks," Electronic Notes in Discrete Mathematics, vol. 36, pp. 89–96, 2010.

[14] S. Ferdousi, F. Dikbiyik, M. F. Habib, M. Tornatore, and B. Mukherjee, "Disaster-aware data center placement and Dynamic content management in cloud networks," Journal of Optical Communications and Networking, vol. 7, no. 7, pp. 681–694, 2015.

[15] Bandyopadhyay, S.: Dissemination of Information in Optical Networks: From Technology to Algorithms. *Springer (2008).*

[16] http://computer.howstuffworks.com/fiber-optic6.htm

[17] Farrell G., (2002). *Optical Communication Systems, presentation. Dublin Institute of Technology, School of Electronic and Communications Engineering.*

[18] Azodolmolky, S., Klinkowski, M., Marin, E., Careglio, D., Sol- Pareta, J., and Tomkos, I. 2009. A Survey on Physical Layer Impairments Aware Routing and Wavelength Assignment Algorithms in Optical Networks. *Elsevier Computer. Network., vol. 53, no. 7, pp. 926944, May 2009.*

[19] http://www.radio-electronics.com/

[20] Chen, W., Evans, J. S., Shieh, W., and Tucker, R. S. Optical Signal-to-Noise Ratio Monitoring Using Uncorrelated Signal-Spontaneous Beat Noise. *ATNAC 2004 (2004):150-155.*

[21] Ali, K. S., Rasheed, I., Sial, M. F. A., & Mehboob, T. (2012, May). Evaluation of Optical receiver sensitivity–bit error rate (BER)/Q factor. *In International conference on Computer and communication technologies (ICCCT '2012) (pp. 26-27).*

[22] Al Zahr, S., Doumith, E. A., & Gagnaire, M. (2011, May). An exact approach for Translucent WDM network design considering scheduled lightpath demands. *In Telecommunications (ICT), 2011 18th International Conference on (pp. 450-457).IEEE.*

[23] Datta, D., Feng, H., Heritage, J. P., Mukherjee, B. & Ramamurthy, B., (1999). Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks. *Journal of Lightwave Technology, 17(10), 1713.*

[24] S. Azodolmolky, M. Angelou, I. Tomkos, T. Panayiotou, G. Ellinas, N. Antoniades, Impairment-aware optical networking: A survey, WDM Systems and Networks, Optical Networks (2012) 443 – 479.

[25] Simmons, J. M. (2006). Network design in realistic" all-optical" backbone networks. *Communications Magazine, IEEE, 44(11), 88-94.*

[26] Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993*). Network Flows: Theory, Algorithms, and Applications*

[27] Aneja Yash, Bandyopadhyay Subir, Rahman Quazi: Optimal Regenerator Placement In Translucent Optical Networks. *Elsevier (2014)*

[28] Bandyopadhyay, S., Rahman, Q., Banergee, S., Murthy, S., and Sen, A. 2009. Dynamic Lightpath Allocation in Translucent WDM Optical Networks. *Communications,2009, ICC09. IEEE International Conference, 1 - 6.*

[29] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the Vulnerability of the fiber infrastructure to disasters," in Proc. IEEE INFOCOM, Rio de Janeiro, Brazil, Apr. 2009, pp. 1566–1574.

[30] G. Ellinas, A. G. Hailemariam, T. E. Stern, and L. Fellow, "Protection Cycles in Mesh WDM Networks," vol. 18, no. 10, pp. 1924–1937, 2000.

[31] Aneja, Y. Jaekel, A., Bandyopadhyay, S. 2007. Some studies on path Protection in WDM networks. Photonic Net. Com., 2007.

[32] Bao, N.-H., Li, L.-M., Luo, H.-B., Yu, H.-F., and Zhang, Z.-Z. 2012. Impairment aware sharing constraint relaxed path protection in translucent optical Networks. Opt. Eng. 51(4), 045002 Apr 06, 2012. doi:10.1117/1.OE.51.4.045002.

[33] M. F. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, "Design of Disaster-Resilient Optical Datacenter Networks," vol. 30, no. 16, pp. 2563–2573, 2012.

[34] J. Xiao, B. Wu, X. Jiang, P. Ho, and S. Fu, "Data Center Network Placement and Service Protection in All-Optical Mesh Networks," no. 2013, pp. 88–94.

[35] J. Yao, P. Lu, L. Gong, Z. Zhu, and S. Member, "On Fast and Coordinated Data Backup in Geo-Distributed Optical Inter-Datacenter Networks," vol. 33, no. 14, pp. 3005–3015, 2015.

[36] Y. Chen, R. H. Katz, and J. D. Kubiatowicz, "Dynamic Replica Placement for Scalable Content Delivery."

[37] F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Minimizing the risk from disaster failures in optical backbone networks," Journal of Lightwave Technology, vol. 32, no. 18, pp. 3175–3183, 2014.

[38] S. S. Savas, F. Dikbiyik, M. F. Habib, M. Tornatore, and B. Mukherjee, "Disaster-aware service provisioning with manycasting in cloud networks," Photonic Network Communications, vol. 28, no. 2, pp. 123–134, 2014.

[39] I. Chlamtac et al., "Lightnets: topologies for highspeed optical networks," IEE/OSA Journal of Lightwave Technologv, vol. 11, (May/June, 1993), pp. 95 1-961.

[40] Gerstel et al., "Dynamic wavelength allocation in WDM ring networks," IEM Research Report RC 20462, May, 1996.

[41] Gerstel et.al. "Dynamic channel assignment for WDM optical Networks with little or no wavelength conversion," Proc. 34th Annual Allerton Conf (Monticello, IL, Oct. 1996), pp. 32-43.

[42] R. Ramaswami et al., "Design of logical topologies for wavelength routed all-optical networks", Proc. IEEE INFOCOM'95, (Boston, Apr. 1995), pp 1316-1325.

[43] Hashimoto, M. and Miura, K., 2012. Layer-Wise Topology Design for Cost Effective IP-Optical Networks.

[44] Yang, H., Zhang, J., Ji, Y., Tian, R., Han, J. and Lee, Y., 2015. Performance evaluation of multi-stratum resources integration based on network function virtualization in software defined elastic data center optical interconnect. Optics express, 23(24), pp.31192-31205.

[45] Andrei, D., Yen, H.H., Tornatore, M., Martel, C.U. and Mukherjee, B., 2009. Integrated provisioning of sliding scheduled services over WDM optical networks [Invited]. Journal of Optical Communications and Networking, 1(2), pp.A94- A105.

VITA AUCTORIS

NAME: Ruchisree Das

PLACE OF BIRTH: Dimapur, India

YEAR OF BIRTH: 1987

EDUCATION: Pranab Vidyapith Higher Secondary School, Dimapur, India, 2006

G.H. Raisoni College of Engineering (RTMNU), Nagpur, India

B.Eng., Information Technology, 2010

University of Windsor, Windsor, Ontario

M.Sc., Computer Science, 2016