PERFORMANCE COMPARISON OF TWO DYNAMIC SHARED-PATH PROTECTION ALGORITHMS FOR WDM OPTICAL MESH NETWORKS

A SHARMA

2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

PERFORMANCE COMPARISON OF TWO DYNAMIC SHARED-PATH PROTECTION ALGORITHMS FOR WDM OPTICAL MESH NETWORKS

By

Ameeth Sharma

Studyleader: Professor Dr. F. W. Leuschner

Submitted in partial fulfillment of the requirements for the degree

Master of Engineering (Electronic Engineering)

in the

Department of Electrical, Electronic & Computer Engineering

in the

School of Engineering

in the

Faculty of Engineering, Built Environment & Information Technology

UNIVERSITY OF PRETORIA

July 2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SUMMARY

PERFORMANCE COMPARISON OF TWO DYNAMIC SHARED-PATH PROTECTION ALGORITHMS FOR WDM OPTICAL MESH NETWORKS

by

Ameeth Sharma

Studyleader: Professor Dr. F. W. Leuschner

Department of Electrical, Electronic & Computer Engineering

Master of Engineering (Electronic Engineering)

Finding an optimal solution to the problem of fast and efficient provisioning of reliable connections and failure recovery in future intelligent optical networks is an ongoing challenge. In this dissertation, we investigate and compare the performance of an adapted shared-path protection algorithm with a more conventional approach; both designed for survivable optical Wavelength Division Multiplexing (WDM) mesh networks. The effect of different classes of service on performance is also investigated.

Dedicated path protection is a proactive scheme which reserves spare resources to combat single link failures. Conventional Shared-path Protection (CSP) is desirable due to the efficient utilization of resources which results from the sharing of backup paths. Availability is an important performance assessment factor which measures the probability that a connection is in an operational state at some point in time. It is the instantaneous counterpart of reliability. Therefore, connections that do not meet their availability requirements are considered to be unreliable. Reliability Aware Shared-path Protection (RASP) adopts the advantages of CSP by provisioning reliable connections efficiently, but provides protection for unreliable connections only. With the use of a link disjoint parameter, RASP also permits the routing of partial link disjoint backup paths.

A simulation study, which evaluates four performance parameters, is undertaken using a South African mesh network. The parameters that are investigated are: 1. Blocking Probability (BP), which considers the percentage of connection requests that are blocked, 2. Backup Success Ratio (BSR), which considers the number of connections that are successfully provisioned with a backup protection path, 3. Backup Primary Resource Ratio (BPR), which considers the ratio of resources utilized to cater for working traffic to the resources reserved for protection paths and lastly 4. Reliability Satisfaction Ratio (RSR), which evaluates the ratio of provisioned connections that meet their availability requirements to the total number of provisioned connections.

Under dynamic traffic conditions with varying network load, simulation results show that RASP can provision reliable connections and satisfy Service Level Agreement (SLA) requirements. A competitive Blocking Probability (BP) and lower Backup Primary Resource Ratio (BPR) signify an improvement in resource utilization efficiency. A higher Backup Success Ratio (BSR) was also achieved under high Quality of Service (QoS) constraints.

The significance of different availability requirements is evaluated by creating three categories, high availability, medium availability and low availability. These three categories represent three classes of service, with availability used as the QoS parameter. Within each class, the performance of RASP and CSP is observed and analyzed, using the parameters described above. Results show that both the BP and BPR increase with an increase in the availability requirements. The RSR decreases as the reliability requirements increase and a variation in BSR is also indicated.

Keywords:
shared-path protection, wavelength division multiplexing, mesh network, blocking probability, link failure, path availability, network recovery, dynamic traffic, service level agreement, quality of service.

I dedicate this work

to my parents,

to my teachers,

and to God.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ACKNOWLEDGEMENTS

I would like to thank

- my promotor, Professor F. W. Leuschner for his encouragement,
- Tondi Mangara for his research on the South African Triangular Topology,
- J.P. Delport and colleagues in the Optronics Research Area (DPSS, CSIR) for access to their cluster computing facility,
- The CSIR and the University of Pretoria for respective post-graduate bursaries,
- and my family for their support.

LIST OF ABBREVIATIONS

ASE	Amplified Spontaneous Emission
BP	Blocking Probability
BPR	Backup Primary Resource-Ratio
BSR	Backup Success Ratio
CC	Cable Cut parameter
CWDM	Course Wavelength Division Multiplexing
CSP	Conventional Shared-backup Protection
DBR	Distributed Bragg Reflector
DFA	Doped Fibre Amplifier
DWDM	Dense Wavelength Division Multiplexing
DFB	Distributed Feed Back
EDFA	Erbium Doped Fibre Amplifier
FTP	File Transfer Protocol
Gb	Gigabit
GB	Gigabyte
GHz	Gigahertz
HTTP	Hypertext Transfer Protocol
IETF	Internet Engineering Task Force
ITU-T	International Telecommunications Union
km	Kilometers
Laser	Light Amplification by Stimulated Emission of Radiation
LED	Light Emitting Diode
MB	Megabyte

MTBF	Mean Time Between Failure
MTTR	Mean time To Repair
nm	Nanometer
OADM	Optical Add/Drop Multiplexer
OXC	Optical Crossconnect
OLT	Optical Line Terminal
PIN	Positive Intrinsic Negative
PP	Protection Path
QoS	Quality of Service
RAM	Random Access Memory
RASP	Reliability Aware Shared-backup Protection
RSR	Reliability Satisfaction Ratio
RWA	Routing and Wavelength Assignment
SATT	South African Triangular Topology
SDH	Synchronous Digital Hierarchy
SLA	Service Level Agreement
SMTP	Simple Mail Transfer Protocol
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical Network
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
WDM	Wavelength Division Multiplexing
WP	Working Path
WUT	Wavelength Utilization Table

CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND

The emergence of Wavelength Division Multiplexing (WDM) technology has provided a promising solution to the ever increasing demand for telecommunications bandwidth over recent years [1]. However, as a consequence of the high bandwidth demand, fault tolerance and network recovery have become critical issues. With multiplexed wavelength channels, each transmitting at rates over a gigabit per second, a single fibre failure can result in severe data and revenue losses [2, 3]. Service providers have been forced to seriously consider resilience mechanisms and schemes that could improve the performance of their networks and keep their customers satisfied.

The Service Level Agreement (SLA) is a contract between the service provider or network operator and the customer that stipulates certain Quality of Service (QoS) guarantees [4]. Should the QoS rendered not meet the stipulation, then the operator may bear financial penalties [5, 6]. Therefore, the question of how to provide satisfiable connections to avoid penalty, as well as minimizing cost and resources, is one of the main concerns of the operator [7].

There are several approaches that can be considered to ensure resilience in optical WDM mesh networks. These are based on two basic survivability paradigms [8, 9]:

1. Path/Link Restoration
2. Path/Link Protection

Recovery involves the rerouting of normal traffic by traversing the working path (WP) over a new path called the backup or protection path (PP). In general, restoration is a dynamic scheme whereby spare resources are used to find recovery paths at the time the failure occurs. Restoration schemes therefore have the advantage of being more efficient than protection schemes since they utilize spare capacity only when required [3]. They also have the ability to recover from a range of failures. In contrast, dedicated protection schemes reserve resources in advance to cater for possible failure scenarios. This has the advantage of superior restoration time but at the cost of inefficient resource utilization [10]. To improve resource utilization of dedicated protection schemes, shared-backup protection was introduced to allow the sharing of backup resources between connections when the corresponding working resources are mutually diverse [3]. These two methods could be applied to either the links that make up end-to-end connections or entire paths from sources to their respective destinations.

A major aspect of future intelligent optical networks will be the ability to provide fast provisioning along with efficient network recovery. Therefore, shared-backup protection may be favored due too its speed of recovery, resource efficiency and guarantees provided on its restoration ability. To achieve this, there is a need for a unified control plane and algorithms that will be responsible for the management of Routing and Wavelength Assignment (RWA) protocols and the setup and tear down of connections [11]. Different QoS requirements are also important, since different customers need different levels of fault tolerance and differ in their willingness to pay for a guaranteed service [7]. Providing such services with different levels of reliability will be beneficial to the service provider by improving resource utilization efficiency and allowing service scalability [7].

In recent years, shared-backup path protection has received much attention [12] and there have been many studies conducted and proposals made regarding novel shared-path protection schemes. Shared-backup protection is advantageous since spare resources are more efficiently reserved by sharing backup resources among many connections. Some studies have also considered double or multiple link failure scenarios [13], mixed shared-path protection [14] and others have considered more complicated routing algorithms and more involved cost analysis [15]. There have also been studies that consider partial path protection exclusively [16]. In a survey of dynamic provisioning methods for shared-backup path protection in optical WDM networks it was found that there exists a trade-off between the operational complexity and service blocking performance [12].

1.2 MOTIVATION

Since WDM optical mesh networks carry huge volumes of traffic, maintaining high levels of service availability at an acceptable level of overhead is an important and critical requirement [17]. Recent studies regarding the evaluation of future optical networks have highlighted network reliability and placed emphasis on the performance of the routing and recovery algorithms used therein [15, 18, 19]. Such networks are expected to provide fast, cheap and reliable services to satisfy the ever increasing demand by end users. Network operators and service providers are in a fiercely competitive market, striving for more and more productivity [20]. Therefore, high network performance and reliability are relied upon to reduce operation and maintenance costs and increase revenues. Furthermore, the service providers are contractually obligated in terms of SLA requirements to meet certain levels of service [10]. Investigating differentiated services that cater for different quality of service requirements is an interesting issue which is motivated by how much end users are willing to pay for the quality of service they require [21].

The above reasons provide the impetus and motivation to evaluate the performance of a shared-path protection algorithm called Reliability Aware Shared-path Protection (RASP) and, within the limitations of a simulations environment, to determine whether there are advantages over an algorithm that uses a more traditional approach, such as Conventional Shared-path Protection (CSP). The network performance parameters that are considered have been used in recent studies to evaluate important characteristics and to determine the credibility of such algorithms. In general, a recovery algorithm would be considered advantageous if it results in the network having a higher degree of network integrity (the ability to provide the desired QoS) and a higher degree of survivability (the ability to recover from failures) [10]. With multimedia and real time applications forming a large percentage of today's traffic, specific QoS will be critical. Hence it is important that routing algorithms provide connections that satisfy their QoS requirements, which include fault tolerance and reliability guarantees [22]. They are also expected to make fast and efficient use of available network resources. Studies based on the development of high performance algorithms have shown that such algorithms are an important requirement for future high performance WDM networks. Such networks will benefit the end user by providing fast, cheap and reliable services. From the network and service providers points of view, high network performance and reliability will result in cost effective network operation and maintenance, i.e. higher revenues.

The impact of such a study is also important in the South African context. The South African economy has exhibited impressive growth in recent years and this will certainly put increasing pressure on the existing network infrastructure. Furthermore, with the introduction of a second network operator [23], a more competitive market will drive stakeholders to consider new and improved technologies in an effort to provide cheaper and faster telecommunication services. These reasons have motivated the use of a South African mesh topology as a possible sample network which is used in the simulation study.

1.3 OBJECTIVES AND SCOPE

In this simulation study, two dynamic shared-path protection algorithms are investigated, which consider the problem of dynamic survivable lightpath provisioning to combat single link failures. The performance of RASP and CSP is evaluated using a simulation approach. The objective of this study is to therefore show that RASP is advantageous when compared with a more conventional algorithm such as CSP. RASP differs from CSP in two ways.

Firstly, it considers the reliability of a connection to determine whether it is dependable or not. This is done by calculating its path availability and comparing it with the availability requirement for that connection. If the path availability is higher or equal to the requirement, then the connection is regarded as being fault tolerant and is routed without protection.

Secondly, RASP makes use of a disjoint link parameter, which allows protection paths to be partially link disjoint. Although relatively uncomplicated in its approach, RASP attempts to be more efficient by providing shared-backup path protection for only those connections that are not dependable, i.e. those that do not meet their availability requirement, and also making use of partially link disjoint protection paths.

A South African triangular network topology consisting of 19 nodes, situated at major metropolitan areas, is used as the sample network. The performance of RASP and CSP, with respect to blocking resource utilization and reliability success will be evaluated and compared using four performance metrics defined in recent studies [15, 18, 24]. Differentiated QoS [25] will be investigated by evaluating network performance under the constraints of three different classes of availability requirements. Single fibre failures will be considered [1, 26].

The investigation is undertaken using a simulation approach. The software implementation is conceptualized using $MATLAB^{®}$ [27] as the programming platform.

The traffic model, routing and wavelength assignment algorithm, and recovery principles are based on recent publications [2, 15, 18, 24].

In summary, since WDM optical networks offer many advantages, including high capacity and speed to the telecommunication industry, the infrastructure and technology is being adopted worldwide, with South Africa being no exception [23]. Due to the large capacity capability, failures have the potential to cause severe losses in data. This situation has resulted in dynamic recovery schemes gaining much attention [2, 17]. The nature of the problem is that such schemes need to be fast and efficient but also reliable to ensure that QoS requirements are met. To investigate this problem, the following approach has been taken:

1. Design a software program, with the guidance of recent literature [2, 15, 18, 19] and within the constraints of the simulation environment, to simulate two dynamic WDM routing schemes (RASP and CSP) which incorporate shared-backup path protection.

2. To use the simulation platform on a realistically dimensioned South African mesh topology.

3. To make use of performance parameters to evaluate the performance of both schemes under different traffic loads.

4. To investigate the impact of different QoS constraints on performance.

5. Analyze and compare the results to determine behavior trends and identify the possible advantages or disadvantages of either scheme.

1.4 OUTLINE OF DISSERTATION

This dissertation is divided into six chapters beginning with this Introduction. Chapter two deals with a brief overview of optical networking components and concepts. Chapter three presents an overview of network recovery and survivability. Chapter four describes the network simulation study and gives insight into the network model, the connection and availability analysis as well as the routing and cost analysis. Chapter five presents the results of the simulation study with an analysis and discussion of the different performance metrics. A summary and concluding remarks are presented in Chapter six followed by the references and two appendices. Appendix A, presents the Haversine formula and Appendix B presents the simulation results.

CHAPTER TWO

OPTICAL NETWORKS: AN OVERVIEW

2.1 INTRODUCTION

This chapter provides a background to optical networks by introducing some of the important concepts. It presents an overview of the important network components and does not delve into great detail, since such information is available in a number of textbooks [28–30]. Other significant concepts, such as WDM and dynamic routing, are also discussed.

2.2 ENABLING TECHNOLOGY

Optical networks, like other types of telecommunication networks, consist of nodes that are interconnected by links. Nodes incorporate optical transmitters, optical receivers and provide switching capability. The links consist of optical fibres which create pathways (lightpaths) for information to be transmitted. The elements that will be described in more detail include optical fibre, optical transmitters, optical receivers, optical amplifiers and wavelength switching devices.

2.2.1 Optical Fibre

As stated above, optical fibres provide links between nodes in an optical network. It is due to the extremely high bandwidth capability of optical fibre and its immunity to electromagnetic interference that led to the conceptualization and development of optical networking. With these advantageous properties, fibre is however brittle and easily prone to damage. A typical optical fibre consists of a core and a cladding. The core is composed of SiO_2 and the cladding is made

of $B_2O_3 - SiO_2$. The movement of an optical signal through a fibre involves waveguide theory. An important requirement is that the cladding should have a slightly lower optical refractive index that the core. This ensures that the optical signal entering the core is reflected repeatedly by the core/cladding junction as it traverses the length of the fibre.

Optical transmission through a fibre is not ideal and losses do occur. Due to these losses, transmitted signals need to be regularly regenerated with the use of optical amplifiers which will be elaborated upon in a later section. Transmission impairments include dispersion (Chromatic, Polarization Mode and Intermodal), scattering (Stimulated Raman and Stimulated Brillouin), and Four Wave Mixing.

Fibre is available in two variants viz. single-mode and multi-mode. Multi-mode fibres were discovered first and have core diameters ranging from 50 to 85 μm. Although inexpensive, multi-mode fibre suffers from intermodal dispersion which reduces the distance between amplifiers and reduces possible bandwidth.

Single mode fibre was discovered in the mid 1980's. They include smaller core diameters in the range of 8 to 10 μm. Optical energy is transmitted in a single mode, eliminating intermodal dispersion. This increased the distance between amplifiers and allowed higher allowable bit rates.

2.2.2 Optical Sources

Optical sources are required to produce the optical energy needed to transmit information from one point to another in an optical network. Two types of sources are used i.e. LED's (Light Emitting Diodes) and laser diodes. LEDs operating in the 850 nm wavelength range were used in first generation optical systems. In these networks, LEDs were acceptable due to the low bit rates and short internodal distance prevalent at the time. However in today's modern optical networks lasers are used as optical sources.

Semiconductor lasers or laser diodes are devices that convert electrical energy to monochromatic light. All lasers consist of a cavity which includes an active medium that is enclosed on either side by a two mirrors or reflecting surfaces. An example of one such laser cavity is the Fabry-Perot Laser. After excitation of the active medium, light waves are reflected back and forth between the mirrors increasing the intensity of certain wavelengths. The emitted light is a high intensity monochromatic beam. Laser materials are chosen and engineered to produce different wavelengths, for e.g. Gallium Arsenide Phosphide is used for the production

of laser light in the 1300 nm wavelength band.

High bandwidth, high distance light sources need to be single mode single frequency devices and lasers used in early systems had the disadvantage of being both multi-mode and multi-frequency. The Fabry-Perot cavity also suffers from this disadvantage and therefore the architecture had to be modified. A number of approaches were taken to achieve single frequency operation including narrowing of the active region. Cleaved coupled cavity lasers, external cavity lasers, Distributed Bragg Reflector (DBR) lasers and Distributed Feed-back (DFB) lasers where later introduced. These approaches included a resonant element to the cavity that suppressed all other modes and promoted a particular single mode. DBR and DFB lasers are commonly used in high bit rate, long distance applications.

To further increase the bit rate of lasers by narrowing the pulse widths, mode-locking was introduced. Mode-locked lasers are implemented by modulating the gain with the laser cavity with the use of saturable absorbers. The absorbers periodically modulate the laser gain locking the relative amplitudes and phases of the different modes to specific Fourier coefficients, resulting in a series of periodic narrow pulses.

Laser tunability has also become a requirement in today's WDM, multi-wavelength networks. Both fast and slow tunability have specific applications. In applications where different connections need to be setup at different wavelengths or wavebands, slow tunability will suffice, whereas in WDM multiple access applications, rapid tunability is required. Temperature tuning is used in slow tunability (up to 1 nm efficiency) applications and is used for laser frequency stabilization. Fast tunability is achieved with DFB and DBR lasers by changing the refractive index. Another cost effective approach to rapid tunability is the use of multi-wavelength laser arrays. With these arrays, selected wavelengths can be simultaneously transmitted by the activation of one or more lasers.

The last important characteristic of laser sources within optical networks is modulation. High frequency modulation of the laser output is a critical requirement if high bandwidth is to be realized. Two types of modulation are possible i.e. direct modulation and external modulation. Direct modulation basically involves the direct modulation of the stimulation source of the laser medium i.e. the electrical current to the laser diode. Direct modulation in digital systems does however create unwanted effects such as chirp, which is due to the broadening of the pulse spectrum. This leads to dispersion problems and hence direct modulation is usually avoided in long distance high bandwidth networks. Chirp can be avoided by the more costly technique of external modulation. External modulation basically involves a constant output laser that is

modulated by an external, controllable attenuator.

2.2.3 Optical Amplifiers

Due to attenuation, signals transmitted through optical fibre require amplification to ensure that signals are able to reach their destination with the required signal to noise ratio. As an illustration, a 1 Gb/s signal that is not amplified would only be able to travel a few hundred km. This is not acceptable as links in typical long haul networks may span many thousands of km. Three types of amplifiers are used in a typical optical link.

A power amplifier provides the necessary boost to the signal before it is launched in to the fibre. The line amplifier is used to amplify signals travelling through the fibre. These amplifiers are placed a calculated distances along the fibre route to restore the signal to its initial amplitude. Pre-amplifiers are used to provide gain to signals just prior to being detected and received. Pre-amplification is necessary as it improves the performance of optical receivers.

Optical amplifiers come in two variances viz. Semiconductor Optical Amplifiers (SOAs) and Doped Fibre Amplifiers (DFAs). SOAs were discovered during the nineteen eighties followed by Erbium Doped Fibre Amplifiers (EDFAs) later that decade.

The SOA is similar in structure to a semiconductor laser. Even though they provide gain over wide bands, SOAs have a few disadvantages. One of the disadvantages is that semiconductor carrier lifetimes are short which leads to gain fluctuations at high bit rates. This further causes crosstalk effects between simultaneously amplified wavelengths. Secondly, an asymmetrical geometry makes the SOA dependant on polarization. Thirdly, large coupling losses exist at the fibre-SOA interface causing a reduction in the resulting gain.

EDFA's made it possible by the mid nineties to accomplish 9000 km long optical links. EDFAs are fibres that are doped with praseodymium, neodymium and yttrium. Erbium has the important property of wide band amplification with the 1550 nm range. EDFAs use energy from an optical source to pump power into the doped fibre at wavelengths matching the characteristics of erbium. The energy can be coupled in both the forward or backward directions. The disadvantage of EDFAs is that their gain profiles are uneven. This causes uneven amplification of wavelengths within multi-wavelength systems. Furthermore, Amplified Spontaneous Emission (ASE) causes amplifier noise which can lead to self saturation and other non linear cross-talk effects.

Due to the high cost implications in reality, EDFAs are typically placed 20 km to 100 km apart. The maximum power capability of the fibre, non linear fibre effects and receiver sensitivity are all taken into consideration when determining the inter amplifiers distance.

2.2.4 Optical Receivers

The other important component in an optical link is the optical receiver. This chapter has thus far dealt with optical transmitters which generate the optical signals, optical fibre which provides the optical pathway for signals and optical amplifiers which regenerate and restore the data signals to overcome the losses endured due to noise and attenuation. Having reached its destination, optical signals are received and interpreted by optical receivers.

Receivers structurally consist of a photo detector, a preamplifier and detection circuitry. Photo detectors are essentially devices that convert optical signals to electronic signals to allow data processing in the electronic domain. Semiconductor photodiodes are commonly used as detectors in optical systems. In digital systems, these devices produce electrical pulses that are proportional to the incident optical energy. Positive Intrinsic Negative (PIN) photodiodes are one type of detector that has improved responsivity. The PIN diodes can be designed with many advantageous properties such as high quantum efficiency, low diffusion currents and high optical and electrical bandwidths. The other type of detector is the Avalanch photodiode. These detectors overcome the problems of thermal noise and low current gain.

The preamplifiers role is to provide gain to the low levels signals produced by the photo-detector so as to improve the performance of the detection circuitry which follows. The main challenge that exists with pre-amplification is the introduction of noise to the system. Furthermore, signals are highly degraded on arrival to the receiver due to the non linear effects experienced along the fibre and the number of regenerative amplifications. Detection circuitry uses thresholding and fast sampling to discern the pre-amplified detected signal. Bit error rates and receiver sensitivity are important characteristics which are used to determine the performance of receivers.

2.2.5 Optical Switches and Wavelength Converters

Optical switches come in a variety of architectures such as Crossbars, Benes and Spanke. Switching capability within nodes in an optical network is required for the setting up or provisioning of lightpaths. When a call request is received from the source node, a suitable route must be determined. Depending on the available wavelengths, data will be transferred or switched from the input port to the output port of each node along the route to successfully make the connection. This is also the case in survivability schemes which are discussed in a later chapter. In such schemes, backup routes are determined to protect or restore links that were interrupted by unexpected failure. In this application, switches are often referred to as routers or selectors.

Switches are characterized by two main properties i.e. the switching time and the number of ports. There are however a number of other properties that are also important. these include the extinction ratio, which is the ratio of output powers of the switch in the open and closed state, and insertion loss which is the amount of power lost due to presence of the switch.

Wavelength converters are found in nodes and are devices that convert incoming data signals of a particular wavelength to another outgoing wavelength [29]. This is most commonly accomplished through optoelectronic conversion. All-optical techniques, such as optical gating, interferometric and wave mixing also exist.

As transponders, wavelength converters convert the data streams entering a network to wavelengths that are suitable for use within the network. Furthermore, they play a role in routing and wavelength assignment optimizing the utilization of available bandwidth. Wavelength conversion is also required at the boundary between networks, where it becomes important to manage and coordinate the allocation of wavelengths of data-streams moving from one network to the next.

Wavelength converters can be classified by a number of parameters, such as the range of the input and output, which could either be fixed or variable [29]. They also can handle a range of optical powers may also be transparent to the bit-rate or modulation formats.

2.2.6 Wavelength Division Multiplexing (WDM)

Two methods are used to increase the transmission capacity of optical fibre [29]. The first method involves Time Division Multiplexing (TDM). TDM involves the interleaving of low speed data streams into one high speed data stream. The resultant higher bit rates do cause limitations. First generation SONET and SDH networks make use of TDM.

The other method to increase capacity is to make use of WDM which involves the simultaneous transmission of multiple carrier wavelengths over a fibre. Wavelength channels need to be sufficiently far apart to avoid interference. Networks today make use of a combination of WDM and TDM, since the both are complimentary [29].

WDM optical network components include Optical Line Terminals (OLTs), Optical Add/Drop Multiplexers (OADMs) and Optical Crossconnects (OXCs). OLTs are used to multiplex and de-multiplex wavelengths. OADMs provide a cost effective means adding and dropping a wavelength from a pass through data stream [29]. Reconfigurable OADMs are highly desirable by having the capability to select the wavelength that is required to be added, dropped or passed through at any time. OXCs are devices that incorporate switch fabrics to allow for the setup and tear down of connections as required. OXCs provide functions such as service provisioning, protection capability and wavelength conversion.

WDM networks can broadly be divided into Dense Wavelength Division Multiplexing (DWDM) and Course Wavelength Division Multiplexing (CWDM) networks. DWDM is used where high traffic loads are present. In DWDM networks, the channel spacing is typically 0.1 nm. In CWDM networks, specifications are relaxed and costs drastically reduced since the channel spacing is increased to 20 nm.

To maximize the efficiency of fibre transmission and to limit the constraints of impairments, specific wavelength bands are used. Most optical fibre networks operate in either the O-Band (1260 nm to 1360 nm) centered at 1310 nm or the C-Band (1530 nm to 1565 nm) centered at 1550 nm since these bands present the lowest attenuation. Advancement in fibre manufacturing technology has led to the creation of the E-Band (1360 nm - 1460 nm) centered at 1360 nm. Other transmission bands include the S-Band (1460nm - 1530 nm) and the L-Band (1565 nm to 1625 nm). Specialized fibres such as Non-dispersion shifted fibre which is optimized for transmission at 1310 nm and Non-zero dispersion shifted fibre which is optimized at 1550 nm have also been used in application where high bandwidth are required.

2.3 MESH AND RING TOPOLOGIES

Lightpath topologies are usually designed to be 2-connected [29], i.e. having a pair of disjoint routes between any pair of nodes in the network. [10]. Networks can be classified into two main topologies viz. ring or mesh, both of which may comply with the 2-connected design.

An example of a mesh and ring topology is shown in Figure 2.1.

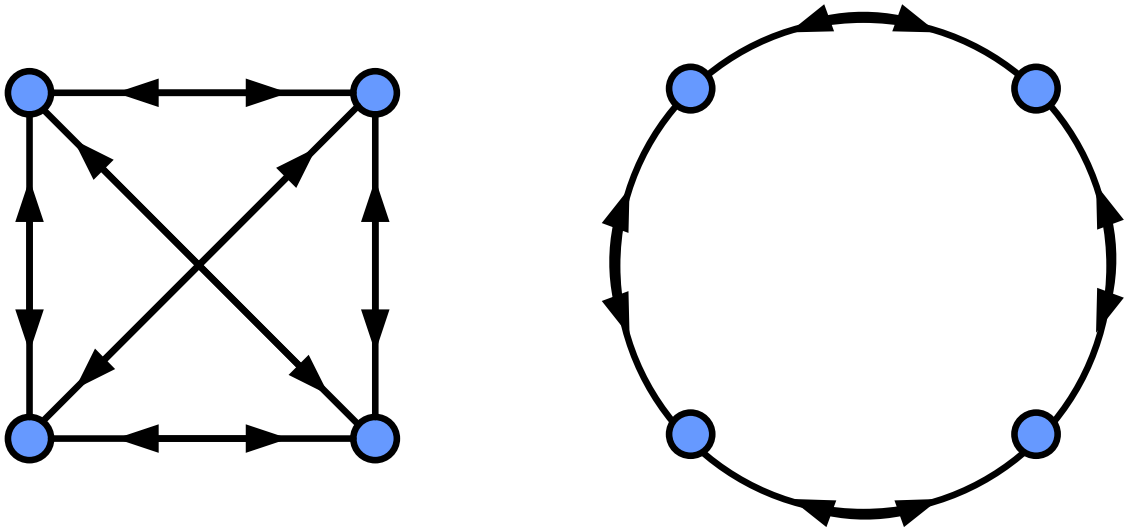

FIGURE 2.1: Mesh and ring topologies

A ring network defines a topology in which the nodes form closed rings which are interconnected. Each node within each ring in the network is connected to its two adjacent nodes. Ring topologies are better suited to networks which cover a small geographical area, while mesh networks are more cost effective for larger networks. Mesh networks also make efficient use of network resources and it is primarily due to these advantages, that mesh topologies have become more popular [30]. Ring networks are sometimes favored due to the reduction in fibre deployment, since for a 2-connected design, rings use the minimum number of possible links.

Another important differences between the two topologies deal with the way traffic is routed. In a ring network, traffic may move within a ring or is routed from ring to ring. In a mesh network, traffic routing is unrestricted and a node may communicate with any other node in the network along diverse routes. In this way, mesh networks provide better resource efficiency than ring networks [10].

Today, more networks are migrating from ring topologies to mesh topologies. This is mainly due to poor scalability of interconnected rings and excessive resource redundancy used in ring based fault management schemes [3]. Furthermore, WDM mesh networks allow for the implementation of various types of protection strategies [31]. The design and operation of mesh WDM networks has therefore received much attention [8].

2.4 NETWORK LAYERS AND PLANES

A layered representation of a network is used to illustrate the functionality and dependance of the different network elements and systems that form part of the network. Standardization bodies such as the ITU-T and the IETF have developed many models [10]. An example of the TCP/IP Stack is shown below in Table 2.1.

TABLE 2.1: An example of a layered network model (TCP/IP Stack) [10].

5	Application Layer
4	Transport Layer
3	Network Layer
2	Data Link Layer
1	Physical Layer

Layer 1 is the physical layer and is composed of the physical components such as optical fibres, transmitters and receivers. It deals with the transmission of data streams over physical links. It also deals with the requirements and all aspects relating to the physical establishment, maintenance and termination of a physical link.

Layer 2 is the data link layer which provides a service to the higher layers by attempting to make the physical link reliable. Its main function is error detection and control which allows higher layers to transmit error-free data over the physical links.

Layer 3 is the network layer and its purpose is to handle data transmission and switching used to connect systems within the network.

Layer 4 is the transport layer which provides a reliable mechanism for the transmission of information between end-points in the network. It also deals with error recovery and flow control ensuring that data is delivered efficiently without errors.

Layer 5 is the application layer which provides the means to applications used by end users to exchange information. Examples of application protocols would be HTTP and FTP for web access and SMTP for email [10].

Network planes are used to identify and organize the different functional blocks within a network. There are three main planes within a network: 1. The Data or User plane, 2. the Control plane and 3. the Management plane. The User plane controls the transmission of user data within the network. Every layer has a User plane. The Control plane controls the setup and tear down of connections. It handles the signalling and routing and also supervises and monitors the network for failures. Every layer has a control plane. The Management plane also exists in every layer and performs layer management, but also exists above all the layers and coordinates the functioning and inter-relations between the different layers.

2.5 DYNAMIC ROUTING

In optical networks, routing algorithms are responsible for finding suitable routes to satisfy lightpath demands. If these demands are known apriori, they are called static demands [32]. Demands are considered dynamic when they arrive unexpectedly with random holding times. Dynamic and static demands collectively result in static and dynamic traffic. The objectives of static lightpath establishment include minimizing network congestion and maximizing single hop traffic.

In dynamic lightpath establishment, the objective is to maximize the average call acceptance ratio or alternatively to minimize the blocking probability [17]. Dynamic requests may or may not be accommodated depending on the availability of resources. Under dynamic traffic conditions, the network state changes and evolves as connections are established and released [30]. Dynamic routing and wavelength algorithms must execute in real time to accommodate these requests if possible. If the connection cannot be provisioned, then it is blocked. Due to the real time nature of dynamic routing, such algorithms are kept relatively simple [17, 30] in an attempt to decrease blocking and reduce recovery times.

CHAPTER THREE
NETWORK RECOVERY

3.1 INTRODUCTION

Fault recovery capability in optical networks is critical, as single failures may affect large volumes of traffic [33]. This section presents an overview of important concepts regarding network recovery and survivability. Optical network failures are covered in Section 2 and fault recovery, Service Level Agreements and Quality of Service are discussed in subsequent sections.

3.2 FAILURES IN OPTICAL NETWORKS

Failure and failure impact studies form an integral part of the design and analysis process of optical networks. Some may argue that one of the most fundamental engineering activities is the study of why things fail and if we cannot prevent failures, then how can we most efficiently recover from them [28].

The causes of network failures are diverse and can be broadly distinguished between planned and unplanned outages [10]. Planned outages are usually caused during maintenance operations e.g. upgrade of software or replacement of faulty hardware. These outages are intentional and pre-planned and appropriate measures are taken to minimize their impact on the consumer e.g. maintenance during the night, notification in advance. Unplanned outages such as cable theft and damage caused by natural disasters, occur unexpectedly and defensive measures are necessary.

The second distinction is whether the cause is internal or external to the network. Examples of internal causes are design flaws or component defects. External causes include storms, digging, vandalism etc.

Software systems in optical networks are also prone to failures. These are usually the most difficult to detect and predict due to the infinite possible failure scenarios. Examples include software design flaws, configuration and routing errors and hacker attacks.

By far, the most common cause of outage in optical networks is cable damage [10]. Today's optical networks are widespread, spanning thousands of kilometers which increases the risk of failure. Great efforts have been afforded to the physical protection of optical fibre, by thick sheathing, conduits and by digging deep trenches to lay them. However, despite all these efforts, fibres continue to be damaged and at surprisingly high frequencies.

A study conducted in the USA during the 1990s [28], found that 60 percent of all cable cuts were caused due to cable dig-ups or digging activities that accidentally hit the cable usually during construction work [10]. A variety of other causes were mentioned ranging from vehicle collisions with poles carrying overhead cables to rodents gnawing of cable sheaths to vandalism and sabotage. Human error during maintenance is another less common and unintentional cause of cable cuts.

Besides human intervention, natural causes of damage to optical fibre cannot be ignored. Floods and heavy rains are capable of damaging or washing away bridge crossings. Excess water also increases the probability of hydrogen infiltration resulting in optical power losses. Rock falls and tree falls caused by heavy wind and storms are also prevalent. Earthquakes have also resulted in the large scale damage to network infrastructure.

Equipment failures also affect the availability of optical networks. Studies have shown that node failures occur much less frequently than cable cuts or damage [10]. Some effective defences against node failures include back up power, good site security and spare on-site equipment.

Failures are expected to occur during the lifespan of equipment. Failures cause equipment to alternate between its operational and fault states. To describe this behavior in a probabilistic way, two parameters are used [10]. The Mean Time Between Failure (MTBF) parameter specifies the average time interval between two subsequent failures of the same network element. The Mean Time To Repair (MTTR) parameter refers to the average time needed to repair the network element and return it to its operational state.

3.2.1 Failure Scenarios

Failure scenarios can either be accounted or unaccounted [10]. Unaccounted failures are those that occur rarely. Some of these failures have dramatic consequences and recovery strategies would be too expensive. Others are rare enough to be ignored due to the inability to justify the budgets required to employ recovery mechanisms for them. Unaccounted failure scenarios will not be covered in this study.

A practical strategy is followed whereby the most frequently occurring failures are grouped and classified under accounted failure scenarios. Two common accounted failure scenarios are single link failures and single node failures.

Single link failures occur when a link between two adjacent nodes fails. Due to this, no data is exchanged between this pair of nodes until repairs have been completed. Single node failures occur when a network node fails. When a node fails, all attached links will not transmit data until the failed node is repaired.

In studies, such as this, where single network failures are considered, it is assumed that the single failures are statistically independent and that the MTTR is much shorter than the MTBF, and hence the probability of more than one failure occurring at a time is negligible [10].

3.3 FAULT RECOVERY

In WDM networks, the failure of a network component leads to the failure of all connections traversing through that component. Since each wavelength carries huge volumes of traffic, it is important that these networks are fault tolerant [22]. In survivable networks, the lightpath that carries traffic during normal operation is called the working path or primary path. When the primary path fails, the traffic is rerouted over a new lightpath called the backup path, secondary path or protection path. The purpose of a recovery scheme is to allow the network to continue functioning in the event of a network failure [20].

In these circumstances it would be advantageous for centralized or distributed control systems to make use of available resources to compute an alternate path (backup path). This backup path can then be used to reroute traffic affected by the failure. The backup paths may either be pre-computed or computed at the time the failure has occurred [3, 17, 34].

Furthermore, the backup path may either be path based (i.e from the source to destination node) or link based (i.e. from a node preceding the failure to a node succeeding the failure) Therefore, survivability paradigms can broadly be classified using four criteria, viz. Execution, Computation, Rerouting and Resources . This is illustrated below in Figure 3.1.

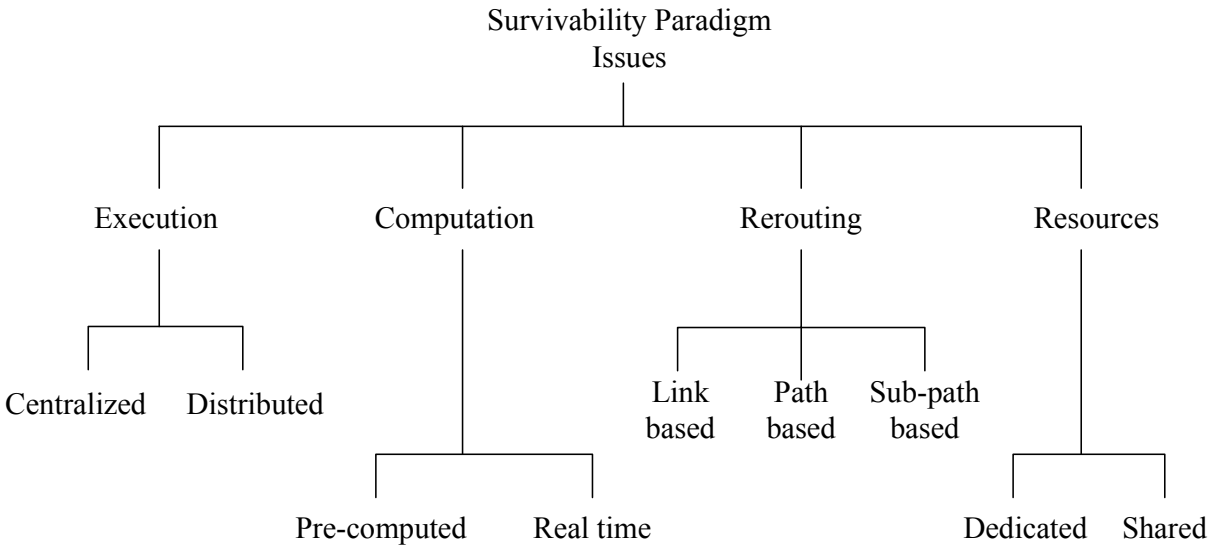

FIGURE 3.1: Classification of survivability paradigms [20].

3.3.1 Execution

A network may employ a survivability scheme that is executed and controlled either centrally or in a distributed manner [35]. Centralized schemes involve a central controller which receives requests generated by source nodes. The controller then does the routing and wavelength assignment while maintaining and updating the status of the network. This requires frequent communication between the central controller and the nodes, to update them, which results in overheads which may become problematic as the network size increases [20]. In distributed schemes [22], no central controller is present. The network operates like a two level network with a data network for physical transmission and a 'shadow' control plane having the same topology as the physical network, which is used for exchanging control signals [35].

3.3.2 Computation

Backup or recovery paths may be computed prior or subsequent to the failure occurrence. Protection schemes use the the pre-computed approach to calculate backup paths before the failure has occurred. Restoration schemes alternatively calculate the backup path in real time, after the failure has occurred [3, 20, 36]. Protection schemes have the advantage of offering fast recovery due to the pre-computation of backup paths. Restoration schemes have the advantage of efficient resource utilization since backup paths are computed only once a failure has occurred. However due to the time necessary for identifying the failure, determining the current network status and finding a recovery path, these restoration schemes are slow and unattractive [20]. Centralized schemes which involve pre-computed routes are conducive for practical implementation [10].

3.3.3 Rerouting

The backup paths may either be path based or link based [17]. Link based approaches employ local detouring of disrupted traffic around the failed link. Path based rerouting methods provide end to to end detouring by computing backup paths from the source node to the destination node [10]. Illustrations of link based and path based recovery are shown in Figure 3.2 and Figure 3.3 respectively. In Figure 3.2, working link 2-3 is protected by protection path 2-5-6-3. In Figure 3.3, working path, 1-2-3-4 is protected by protection path 1-5-6-4.

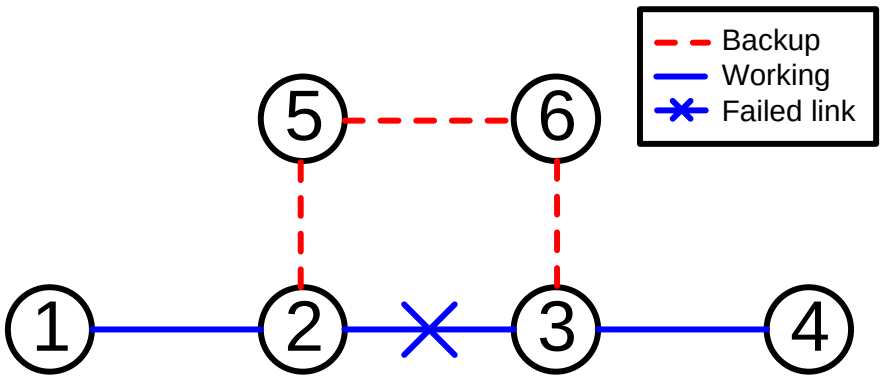

FIGURE 3.2: Link based recovery.

Recently, the idea of sub-path protection in mesh networks has been proposed [3]. Sub-path protection involves dividing the Working Path (WP) into a number of segments and protecting each segment separately [37]. Compared with path protection, sub-path protection can achieve high scalability and fast recovery for a modest sacrifice in resource efficiency [3].

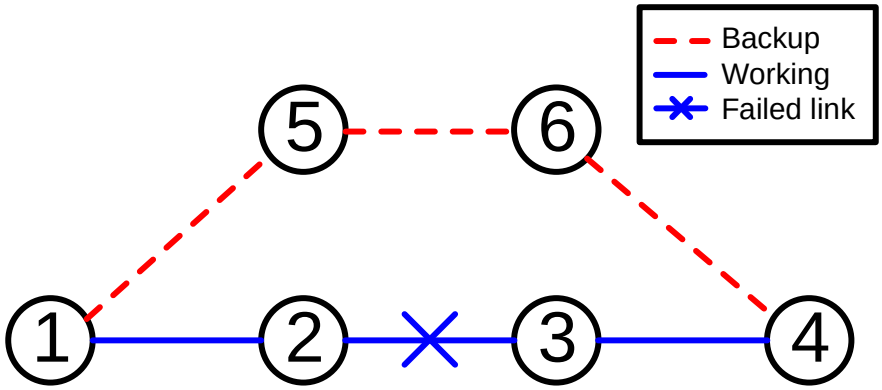

FIGURE 3.3: Path based recovery.

3.3.4 Resources

Survivability schemes also differ with respect to how backup capacity or resources are utilized. Dedicated techniques specify that each primary path should have its own dedicated backup path. [19]. In protection schemes, dedicated protection is classified into 1+1 and 1:1 schemes [25]. In 1+1 protection, data is transferred to both the primary and secondary paths simultaneously, whereas in the 1:1 case, data is sent over the primary path only and the secondary path may be used for other low priory traffic [10]. When the failure does occur, traffic is switched over to the backup path. In comparison, 1:1 protection is slower, although it makes more efficient use of capacity. Dedicated schemes therefore use twice the bandwidth required for transmitting data to protect connections against single link failures [22].

In the shared case [38], primary paths may share the same backup resources as long as the primary paths are link and node disjoint. Shared approaches employ backup multiplexing [10]. In M:N protection, to improve link utilization, resources are shared among backup paths and these schemes are categorized as shared-backup path protection schemes. Shared protection may also come in the form of segment shared protection, where protection segments may be shared between different working paths [38].

An example of shared-backup path protection is shown in Figure 3.4. In the figure, two link disjoint working paths (1-5-6-3 and 1-7-8-4) share common protection links (1-2 and 2-3). The disjoint constraint for the working paths is to ensure that if one of the working paths should fail, then that connection would be able to use the protection path to recover. If the working paths shared a common working path link and if that common link failed, then only one of the working paths would be able to recover. This problem becomes worse as the number of working paths that share protection bandwidth, increases.

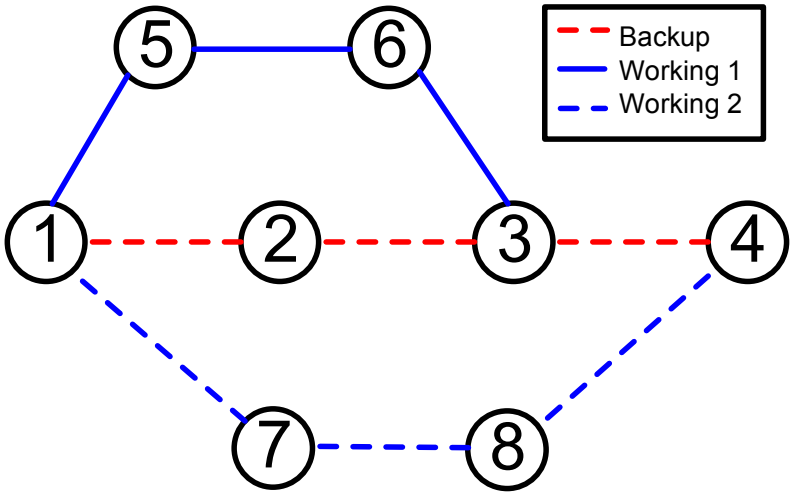

FIGURE 3.4: Example of Shared-backup path protection.

3.4 SERVICE LEVEL AGREEMENTS (SLAS) AND QUALITY OF SERVICE (QOS)

Network survivability in general, gauges the ability of the network to support a committed QoS continuously in the presence of various failure scenarios [39, 40]. QoS service is a combination of several qualities or properties of a service, such as [41]:

- Availability, which is the percentage of time that the service is in it operational state,
- Security, which refers to the authentication, confidentiality, data integrity and resistance to attacks,
- Response time, which refers to time taken to respond to service requests, including connection setup time and recovery time and service mean down time,
- and Throughput, which refers to the rate at which services can process requests.

Reliability parameters mainly include availability and restoration time [9]. Service disruption time and the quantity of data lost due to disruption are also metrics used to capture the routing dynamics in telecommunication networks [4]. Telecom carriers and network service providers offer contracts or SLAs to their customers providing details of the QoS that their customers may expect and the penalties that may result from the SLA being violated [36, 41]. SLAs typically specify the minimum availability of service and the maximum downtime that is acceptable. The more stringent the reliability requirements, the more expensive the service will be [10].

CHAPTER FOUR
NETWORK SIMULATION

4.1 INTRODUCTION

This chapter describes the simulation study that was undertaken and is divided into three sections. Section 2 presents the simulation model used. Section 3 covers the South African network topology and Section four describes the software algorithms with the aid of flowcharts.

4.2 SIMULATION MODEL DESCRIPTION

4.2.1 Network Model

The network topology is defined by $D(N, L, W)$ where N is a set of nodes, L is a set of bi-directional links and W is a set of wavelengths per link. Each link is allocated a capacity of eight wavelength channels. In this study, it is assumed that all nodes have wavelength conversion capability [15, 18, 19]. It is also assumed that only one connection request arrives at any point in time. An availability matrix, A_N, contains the availability values of all the links in the network. The network state is described by two other such matrices, λ_w and λ_p, which respectively store the number of working and protection wavelengths being utilized on each link at any point in time. λ_p also includes information about which connections are sharing protection wavelengths on a particular link. The traffic is dynamic and connection requests arrive without knowledge of subsequent arrivals. A connection request is described by $r(s, d, a_r)$, where s is the source node, d is the destination node and A_{req} is the availability requirement of the request.

4.2.2 Traffic Model

Telecommunication traffic is defined as the average number of connections in progress [42]. A widely accepted approach to dynamic traffic modelling has been adopted in which the arrival of connection requests is a Poisson Process with a constant arrival rate β [15, 18]. The arrival rate is the rate at which connection arrival requests are received by the network per unit time. Using a constant arrival rate assumes that the network is in a statistical equilibrium i.e. sudden changes in arrival rate can be ignored [42]. This model is popular since it realistically describes the arrival of connection requests, which are independent of each other. The connection holding time refers to the time duration of a connection from its establishment to its termination. Holding times are random and a negative exponential distribution, with a mean of $1/\mu$, is used for the holding time. Traffic is measured in Erlangs which is a dimensionless unit. In general, Erlang traffic ($T_{Erlangs}$) is defined by the equation

$$T_{Erlangs} = \beta \frac{1}{\mu} , \tag{4.1}$$

where β refers to the connection arrival rate and $1/\mu$, the mean holding time [42]. In this study, the time measurements have been normalized by assuming $1/\mu = 1$ so that the network traffic load can be considered, in units of Erlang, as being equivalent to β [12, 19]. Source and destination nodes are randomly chosen allowing more than one connection to be established between any pair of nodes. A waiting queue is not implemented and if the algorithm cannot establish a connection, then it is rejected or blocked immediately. The total number of arrival requests simulated is 10^6 [19].

4.2.3 Connection Availability Analysis

In this study, reliability is measured using availability [13] since availability denotes the percentage of time that a component or connection will be in its normal operating state at any random point in time [41, 43]. Here, availability is defined and calculated for an end to end connection that is established as either a working path or a combination of working and backup paths [44]. Connection requests that meet their fault tolerance requirements (A_{req}) are called dependable connections [17]. Furthermore, availability is an important decision criterion, used in network planning and dimensioning studies as it is often indicated in SLAs between service providers and customers [36, 45]. It is assumed that only one link fails at a time and that the

MTBF and the MTTR are independent, memoryless processes [2, 15]. Due to the greater effect that link failures have on network performance, the availability of other network components, such as nodes and amplifiers, has been neglected and assumed to be 1.

The following notation is used in the definitions and equations that follow.
ij: the link connecting nodes i and j which is represented by two unidirectional fibres.
a_{ij}: the availability of ij.
c_{ij} : the cost of ij, which is determined by the availability of ij and and/or the wavelength assignment on ij.
a_{path}: the availability of an arbitrary path, consisting of a series of connected links.
a_{LDP}: the availability of a link disjoint pair or working and protection paths.
a_{wp}: the availability of a link disjoint working path.
a_{pp}: the availability of a link disjoint protection path.
a_{PLDP}: the availability of a partial link disjoint pair of working and protection paths.
S_1: a set of links common to both the working and protection paths of a partial link disjoint path pair.
S_2: a set comprising the links of all link disjoint path segments of a partial link disjoint path pair.
LDP_k: the k_{th} link disjoint segment of a partial link disjoint path pair.
wp_k: the working path of the k_{th} link disjoint path segment.
pp_k: the protection path of the k_{th} link disjoint path segment.
ξ: the link disjoint parameter defined between 0 and 1.
θ: the spare capacity usage factor.

4.2.3.1 Availability of a Link Disjoint Path Pair

An example of a link disjoint path is given in Figure 4.1. The example consists of a working path 1-2-3-4 and a backup path 1-5-6-4. CSP makes use of a pair of link disjoint paths when provisioning a connection. RASP may also establish a pair of link disjoint paths, but if one is not available then it has the advantage of provisioning a partial link disjoint path as well [18, 21]. An example of a partial link disjoint path is given in Figure 4.2.

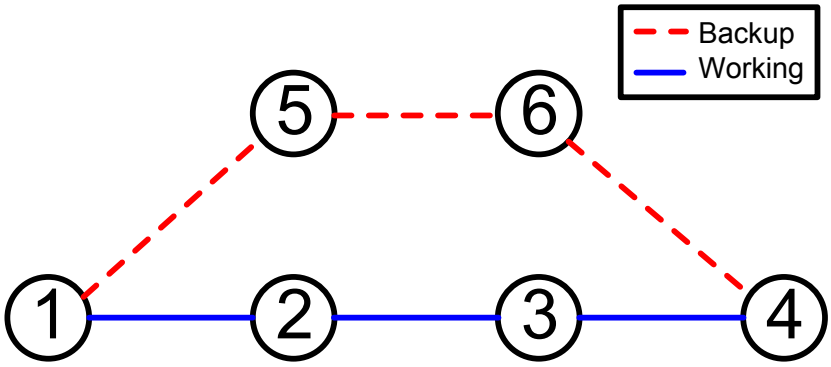

FIGURE 4.1: Link disjoint working and backup path pair.

As defined, the availability of an arbitrary path, consisting of a number of interconnected links from source to destination, is given by the equation

$$a_{path} = \prod_{ij \in path} a_{ij} \ . \tag{4.2}$$

Following from Eq. 4.2 and with respect to Figure 4.1, the availability of the working and protection paths, a_{wp} and a_{pp}, are respectively described by the equations

$$a_{wp} = a_{12}\ a_{23}\ a_{34} \tag{4.3}$$

and

$$a_{pp} = a_{15}\ a_{56}\ a_{64} \ . \tag{4.4}$$

The calculation for the availability of a link disjoint path pair [7, 15] is given by the equation.

$$\begin{aligned} a_{LDP} &= 1 - (1 - a_{wp})(1 - \theta.a_{pp}) \\ &= a_{wp} + a_{pp} - a_{wp}\ a_{pp} \ . \end{aligned} \tag{4.5}$$

In the following explanation, this connection will be referred to as R, comprising working path A and protection path P. In Eq. 4.5, the spare capacity usage factor, θ, is introduced. Since RASP and CSP are both shared-backup protection schemes, the backup resources maybe shared with other connections on condition their working paths do not traverse common links [10].

The spare capacity usage factor is therefore used to denote the probability that connection R can use the resources of P to recover from the failure of A. This probability is determined by the probability that the other connections sharing backup resources with R, will not fail before the failure of R occurs [15]. The value of θ is inversely proportional to the number of connections sharing the backup resources with R. In the multiple failure case, the greater the number of connections sharing backup resources, the higher the probability that the resources will be utilized by one of them, resulting in all the others not having any protection. Since this study considers single link failures, only one connection may fail at any time resulting in θ having a value of 1 [15].

4.2.3.2 Availability of a Partial Link Disjoint Path Pair

Figure 4.2 gives an example of a partial link disjoint path pair. The example shows that the connection comprises three sub-paths, i.e. 1-2, 3-4 and 2-5-3. Sub-paths 1-2 and 3-4 include fibres that share protection and working links. The remaining sub-path, 2-5-3, which is link disjoint consists of a working path 2-3, and protection path 2-5-3. Therefore consistent with their definitions, in this example s_1 consists of links 1-2 and 3-4, and s_2 consists of links 2-3, 2-5 and 5-3. Furthermore, $k = 1$, since there is only one sub-path which is link disjoint.

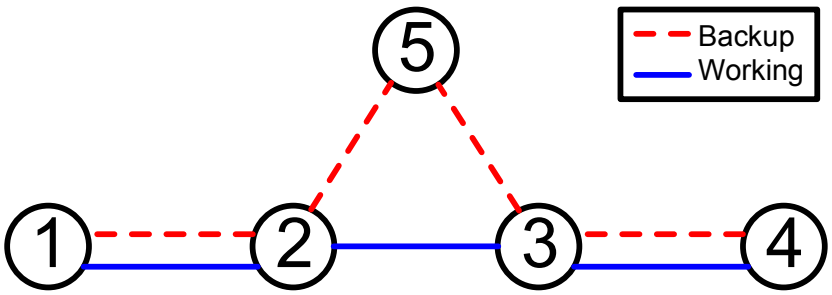

FIGURE 4.2: Partial link disjoint working and backup path pair.

The calculation for the availability of a partial link disjoint path is given by developed in the following equations [7, 15].

$$a_{PLDP} = a_{S_1}\, a_{S_2}\ . \tag{4.6}$$

$$\begin{aligned} a_{S_1} &= \prod_{ij \in S_1} a_{ij} \\ &= a_{12}\, a_{34}\ . \end{aligned} \tag{4.7}$$

$$a_{S_2} = \prod_{LDP_k \in S_2} a_{LDP_k}\ . \tag{4.8}$$

$$\begin{aligned} a_{LDP_k} &= 1 - (1 - \prod_{ij \in wp_k} a_{ij})(1 - \prod_{ij \in pp_k} a_{ij}) \\ &= \prod_{ij \in wp_k} a_{ij} + \prod_{ij \in pp_k} a_{ij} - \prod_{ij \in wp_k} a_{ij} \prod_{ij \in pp_k} a_{ij} \\ &= \prod_{ij \in wp_k} a_{ij} + \prod_{ij \in pp_k} a_{ij} - \prod_{ij \in (wp_k \cup wp_k)} a_{ij} \\ &= a_{23} + a_{25}\, a_{53} - a_{23}\, a_{25}\, a_{53}\ . \end{aligned} \tag{4.9}$$

4.2.4 Routing and Cost Analysis

A routing algorithm is used to establish an appropriate path from a source to a destination. There are two main objectives of network routing. One is to maximize network throughput by providing as many connections as possible. The other is to minimize the cost of these paths, by providing least cost paths [20].

Both CSP and RASP employ the concept of shared-backup path protection where, in order to protect WPs from single link failures and make efficient use of resources, the protection wavelengths may be shared by different PPs only if their respective WPs are link disjoint [12]. Lightpaths are established and taken down dynamically using the traffic model described above [18, 46]. Dijkstra's least cost routing algorithm is used to search the network to discover possible working and protection paths to satisfy requests [7, 18, 19, 37]. Dijkstra's algorithm uses information provided by the arrival request, r, such as the source node, s, and destination node, d, as well as additional information about the current network state provided by λ_w and λ_p. In order to find a least cost path, A_N is required to compute the cost of each link.

When a connection request arrives, both CSP and RASP respond by searching for a suitable working path. The cost of each link is first computed using Eq. 4.10.

$$\mathrm{c_{ij}} = \begin{cases} +\infty & \text{if } SW_{ij} = 0 \ , \\ -\ln a_{ij} & \text{otherwise} \ . \end{cases} \tag{4.10}$$

As mentioned, each link is assumed to have a capacity of eight wavelength channels. SW_{ij} represents the number of free wavelengths present on link ij and is equal to zero when all eight wavelengths are being utilized. A link cost of infinity excludes a particular link from the search for a route. Eq. 4.10 results in a link having a high cost when its availability is low and vice versa. When Dijkstra's algorithm is unable to find a suitable working path, due to the lack of resources, the connection is blocked [18, 46].

Satisfying a connection request may also involve finding a suitable backup path. RASP, being a reliability aware algorithm, is able to establish a working path without protection if the availability of the working path is greater or equal to the availability requirement of the connection. CSP does not consider reliability requirements and compensates by routing every connection with a link disjoint protection path. RASP allows partial link disjoint protection which improves the probability of finding a protection path (PP). When a backup protection path is required, the cost of each link is computed using Eq. 4.11.

$$\mathrm{c_{ij}} = \begin{cases} -\ln \xi a_{ij} & \text{if } ij \in WP \ , \\ -\ln a_{ij} & \text{if } SW_{ij \notin WP} \neq 0 \ , \\ +\infty & \text{if } SW_{ij \notin WP} = 0 \ . \end{cases} \tag{4.11}$$

The main difference between Eq. 4.10 and Eq. 4.11 is that consideration is given to whether a specific link has been used in the working path of the connection. A link disjoint parameter, ξ, is introduced which is defined between 0 and 1. ξ is used to determine how disjoint the resulting protection path is. CSP uses a value of zero for ξ resulting in a cost of ∞, when $ij \in$ WP. This excludes all working path links from the search, resulting in fully link disjoint protection paths. RASP uses a non-zero value for ξ which results in a higher cost for all $ij \in$ WP but does not exclude it from the search for a protection path. RASP assumes a value of 0.01 for ξ [18] and will attempt to find a disjoint path but if one is not possible will attempt to find a path that is partially disjoint by utilizing protection bandwidth on one or more of the working path links.

In this way, the use of ξ does not allow RASP to favor a non-disjoint or partially disjoint path, but permits one should there be no alternative. ξ, therefore increases the probability of finding a suitable PP for a WP that is unreliable. Once a possible backup path is found, RASP will again calculate the availability of the path pair to ensure that it meets the availability requirement of the connection before being established. In the case of CSP, the path pair is established immediately, without any verification of reliability.

4.3 SOUTH AFRICAN NETWORK TOPOLOGY

In accordance with global trends, telecommunication traffic carried by South Africa's backbone network is on the increase. In future, voice will account for only a small portion of the total traffic, with data being the predominant component. The following section describes a network topology that was proposed as a reference scenario for studies on photonic networks in South Africa. The study was conducted by T. Mangara as part of his post graduate studies at the University of Pretoria. In his study, Mangara presents a hypothetical highly meshed fibre optic backbone network for South Africa. The network's physical topology can be seen in Figure 4.3 below.

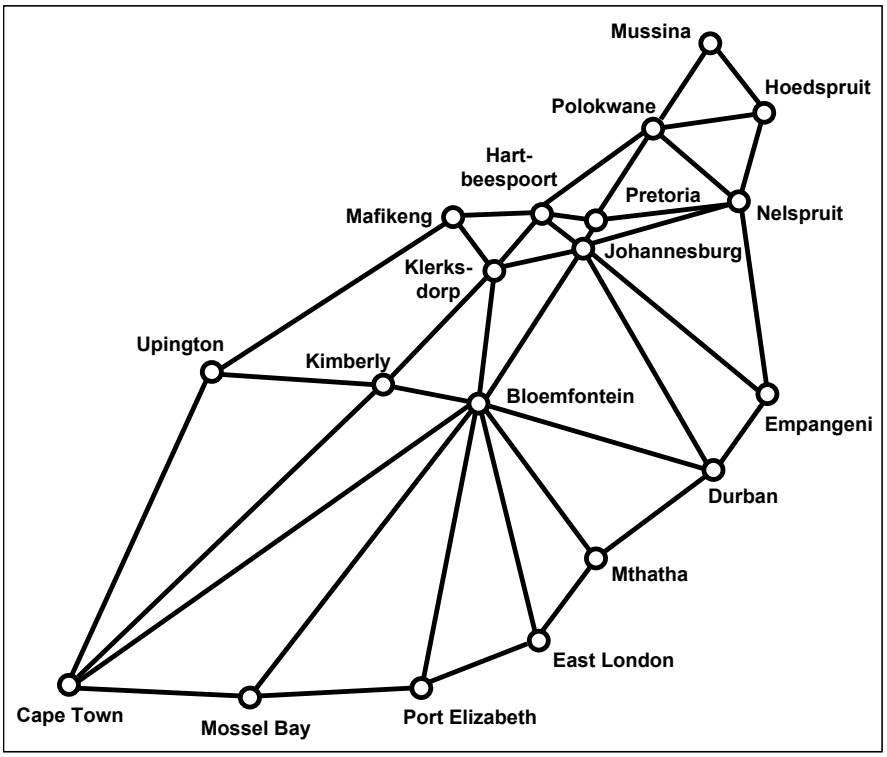

FIGURE 4.3: South African Triangular Topology.

TABLE 4.1: SATT's topological parameters.

Parameter		Value
Nodes		19
Links		40
Nodal Degree	Minimum	2
	Maximum	9
	Average	4.2
Node and Line Connectivity	Minimum	2
	Maximum	5
	Average	3.4
Fibre Distance	Minimum	47.7
	Maximum	1326.4
	Average	476.7
Network Diameter	Km	5092.8
	Hops	11

The network was derived from an investigation of currently deployed network topologies [23]. The triangular topology consists of 19 nodes located at major South African metropolitan areas. These nodes are interconnected by 40 bidirectional fibre links and are assumed to have wavelength conversion capability [15, 18]. Due to the highly meshed triangular architecture, the network was named the South African Triangular Topology (SATT).

4.3.1 SATT Network Parameters

A summary of some topological parameters of SATT are tabulated in Table 4.1, which shows the maximum, minimum and average nodal degree and the node and line connectivity of SATT. The nodal degree of a node refers to the number of links that are incident to that particular node. The node and line connectivity is determined for a pair of nodes, and refers to the number of node disjoint and link disjoint paths that can be determined between this pair of nodes. Connectivity is a useful parameter which provides a measure of survivability. The network diameter is the length of the longest node and link disjoint paths between any two nodes. Distance can be measured both in kilometers (km) and number of links or hops.

A further requirement for the characterization of SATT is the physical length of the optical

TABLE 4.2: SATT's availability statistics.

Stat	Value
Minimum	0.991924
Maximum	0.999709
Average	0.997097
Median	0.997436

links which could either be underground or overhead. SATT has been dimensioned using the Haversine formula [see Appendix A] which uses longitude and latitude co-ordinates to determine the great circle distance between two points on the earth's surface. This distance between two interlinked nodes is referred to as the airline distance (D_{air}) [47]. From the airline distance, the optical fibre distance (D_{fib}) for each link can be calculated using the relations in Eq. 4.12 [48].

$$D_{fib} = \begin{cases} 1.5\ D_{air} & if D_{air} < 1000km\ , \\ 1500km & if 1000km < D_{air} < 1200km\ , \\ 1.25\ D_{air} & if D_{air} > 1200km\ . \end{cases} \tag{4.12}$$

Using the physical lengths provided by the Haversine formula [see Appendix A], the MTBF of each fibre link was calculated using Eq. 4.13 [10].

$$\text{MTBF}_{\text{fibre}}(hours) = \frac{CC \times 365 \times 24}{\text{length of fibre}}\ . \tag{4.13}$$

The Cable Cut parameter (CC) is defined as the average length of cable that results in a single cable cut per year. For terrestrial optical fibre, CC assumed to be 450 km [10, 45].

$$a_{ij} = 1 - \frac{\text{MTTR}_{\text{Link}}}{\text{MTBF}_{\text{Link}}}\ . \tag{4.14}$$

Using a MTTR of 24 hours [10] and a MTBF, as calculated in Eq. 4.13, a realistic value for the availability of each fibre link was calculated using Eq. 4.14. Table 4.2 summarizes the availability results for SATT.

4.4 SOFTWARE DESCRIPTION

4.4.1 General Procedure (RASP and CSP)

The software was divided into a series of subprograms with each one responsible for executing a smaller function or task. In this section, the general flow of the software implementation is described. A flowchart showing the general process followed by both CSP and RASP is shown in Figure 4.4.

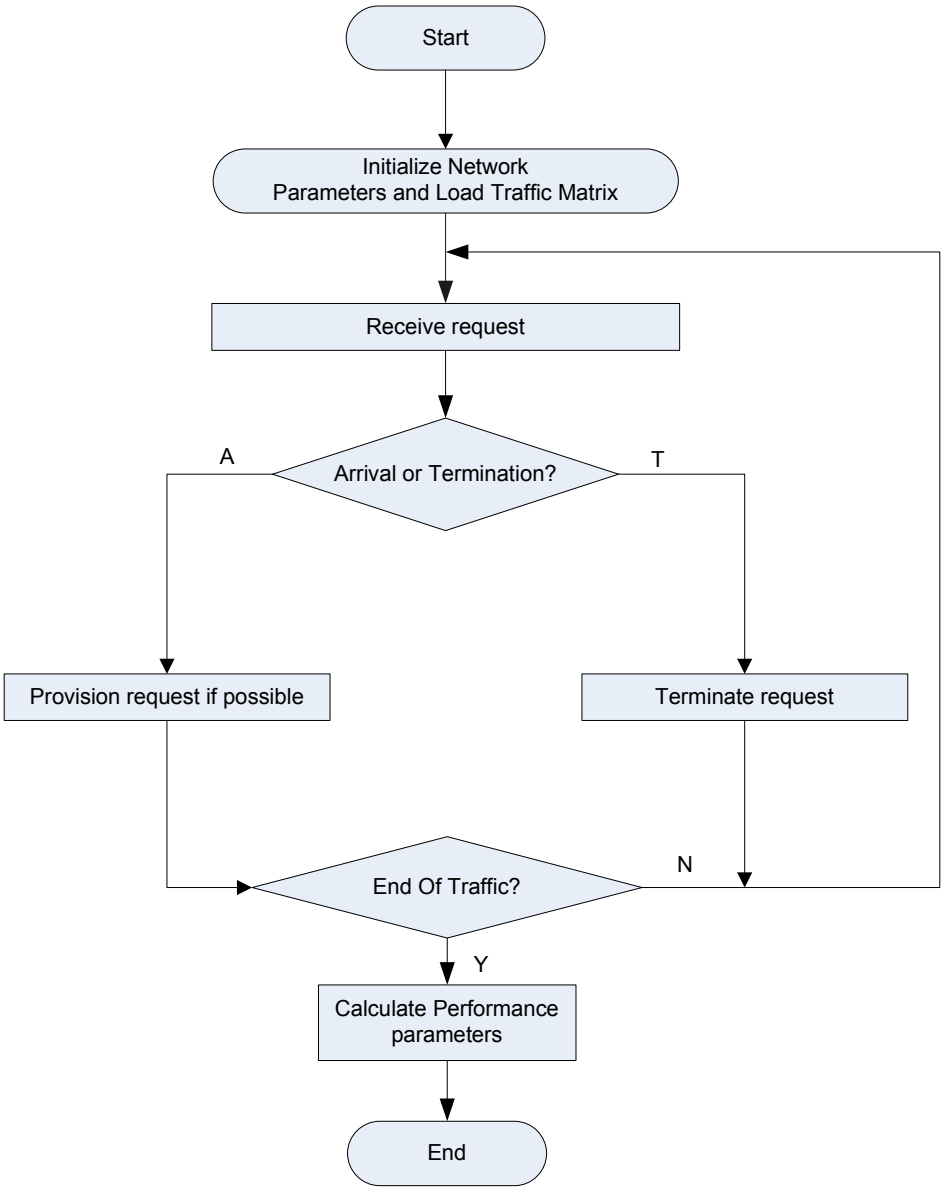

FIGURE 4.4: Flowchart showing the general process followed by both the RASP and CSP algorithms.

The program first goes through initialization. In this step, all constant network parameters such the number of wavelength channels per link, CC and MTTR are allocated values. A number of counters are initialized to zero. Matrices, such as the availability matrix (which contains the availability of all links) and the cost matrix (which includes the cost of each link), are also initialized.

After initialization, the traffic matrix is loaded. The traffic matrix is generated (using the model described earlier) to create a list of sequential connection requests. These requests may either be an arrival request, which requests the establishment of a connection or a termination request, which requests the tear down of connection. In every row of the the traffic matrix, a connection is described by five attributes:

1. Identifier
2. Type
3. Source node
4. Destination node

The Identifier attribute is used identify a connection and is stored along with any routes belonging to that connection. Every request appears twice in the traffic matrix, firstly for its arrival and secondly for its termination. The Type attribute indicates whether the connection request is an arrival request or a termination request. The source and destination nodes are then also identified.

Once a connection request has been received, the Type attribute is used to determine whether the program should attempt to establish the request or tear it down. The program subsequently stays in this loop, accepting connections requests sequentially. Arrival requests are attributed a random availability requirement (A_{req}) which is used to determine whether the respective connections are dependable or not. As connections are accepted and terminated, the network state (which is defined by the the wavelength usage on every link) is modified. Specific software counters which collect valuable statistics are also incremented. Some examples of statistical data that is progressively updated as the simulation runs, include the total number of blocked connections, the total number of connections accepted with protection path, the total wavelengths used for working paths etc. Once all the connection requests in the traffic matrix have been executed, the loop ends and the statistical data is used to calculate the performance parameters.

4.4.2 RASP Arrival Request

Figure 4.5 is used to describe the series of actions that take place, in an attempt to provision an arrival request when using RASP.

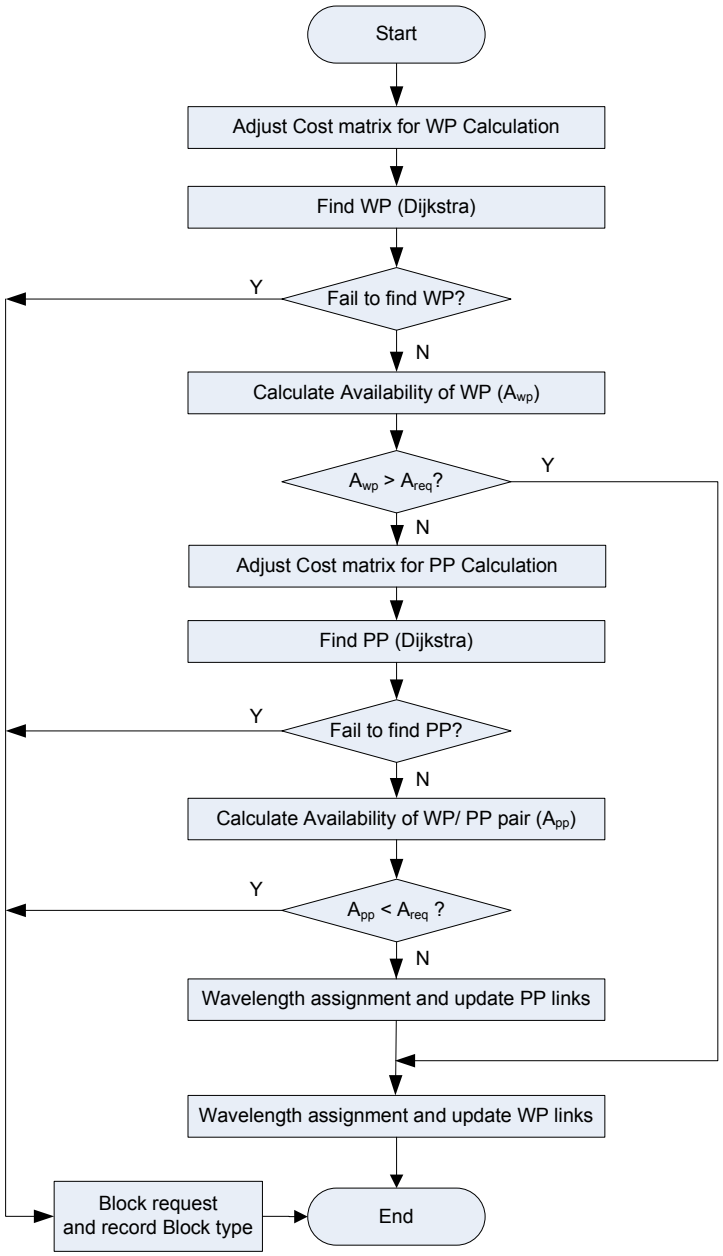

FIGURE 4.5: Flowchart showing the RASP connection arrival procedure.

In this figure, A_{wp} and A_{pp} refer to the availabilities of the working and protected paths respectively. A_{req} refers to the availability requirement of the connection. When the a connection request is received, the cost matrix is first adjusted using Eq. 4.10. The calculation requires the availability of each link and the number of free wavelengths on each link. Dijkstra's algorithm is then used to find the shortest path between the source and destination nodes.

If a suitable WP is not found, it is blocked. A blocked connection is handled by using the connection identifier to locate and delete the expected termination request of that connection from the traffic matrix. Once blocked, the blocked connections counter as well as the counter which records the the type of failure are incremented.

If a suitable path is found, then A_{wp} is calculated using Eq. 4.2. If $A_{wp} > A_{req}$, then the connection is considered dependable and is established. To establish a connection, the wavelength utilization tables (WUT), on each of the links belonging to WP, are updated by recording the Identifer next to the utilized wavelength in λ_w. The free wavelength counters on each of these links are correspondingly decremented. The counter, for the total number of WP wavelengths used, is incremented by the number of links used by the WP.

If $A_{wp} < A_{req}$, then RASP will attempt to find a protection path for this connection, since a protected connection has a higher availability than an unprotected one, which could make it dependable. The cost matrix is then updated, using Eq. 4.11. RASP utilizes a value of 0.01 for ξ to enable the possibility of partial link disjoint protection paths. Dijkstra's algorithm is recalled with the updated cost matrix to find a suitable protection path. If one is not found, then the connection is blocked and the same block procedure, as described above, is followed. If a PP is found, then A_{pp} is calculated using either Eq. 4.5 or Eq. 4.6. A_{pp} is then compared with A_{req} to make sure that the protected connection is dependable. If not then the connection is blocked. If the protection path is dependable, then the path pair is established. The working and protection paths are recorded in λ_w and λ_p respectively. On a link where an already used protection wavelength is being shared, the connection identity is recorded. In addition, when a new protection wavelength is needed on that link, the free wavelength counter is also decremented by one. All other counters are also updated. An example of a WUT for a specific link is given in Table 4.3.

From the example the following can be deduced:

1. A total of seven out of eight wavelength channels are being used on this link.

2. Four are used for WPs and three are used for PPs.

Table 4.3: An example of Wavelength Utilization Tables (WUTs) in λ_w and λ_p for a specific link

λ_w	
Channel number	Connection Identifiers
1	5
2	13
3	20
4	36
5	-
6	-
7	-
8	-
Total	4

λ_p	
Channel number	Connection Identifiers
1	[20, 32,6]
2	2
3	[9,40]
4	-
5	-
6	-
7	-
8	-
Total	3

3. The WPs of connections 2, 6, 9, 20, 32 and 40 are link disjoint with this link.

4. Connection 20 uses a partial link disjoint protection path.

4.4.3 CSP Arrival Request

Figure 4.6 describes the procedure followed when an arrival request is received by the CSP algorithm. There are three main differences when comparing it with Figure 4.5. The first difference is that A_{wp} is not evaluated. This step is omitted in CSP, since all connections are established as protected path pairs. The second difference is that the cost matrix is adjusted using Eq. 4.11 with ξ=1 instead of ξ=0.01 (as was used in the case of RASP). This forces Dijkstra's algorithm to find a link disjoint protection path, by making the availability of all working path links equal to infinity and excluding them from the search. The third difference is that the calculation of A_{pp} is absent. A_{pp} is calculated in the CSP simulation and although it is not used by CSP, it is done for statistical purposes and used later in performance evaluations. If a connection is established, A_{pp} is used to determine whether or not the connection meets its hypothetical QoS requirement. The other aspects of the CSP flowchart are similar to those of the RASP flowchart and are discussed in the previous section describing the RASP arrival process.

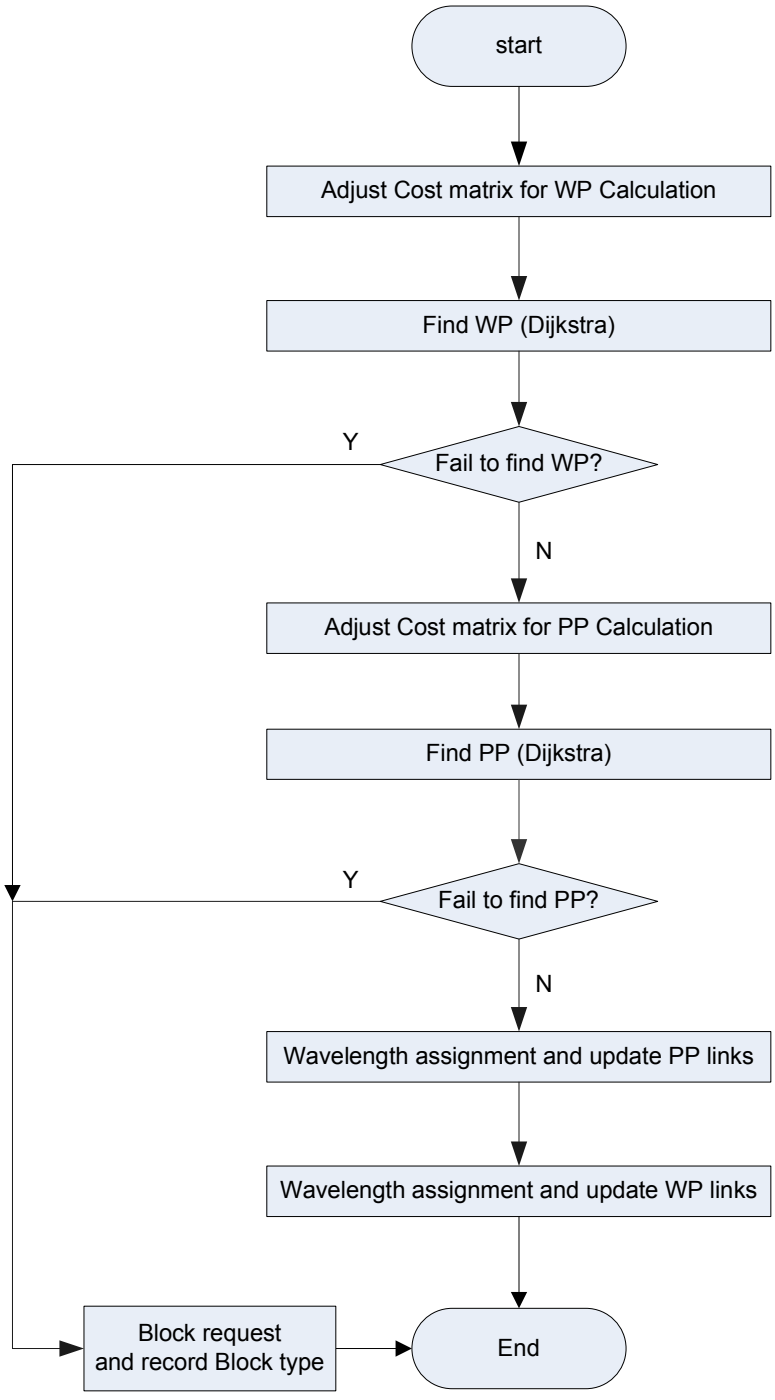

FIGURE 4.6: Flowchart showing the CSP connection arrival procedure.

4.4.4 RASP Termination Request

Figure 4.7 describes the termination process of the RASP algorithm. When a termination request arrives, the first step is to determine whether the connection being terminated has a PP or not. This is achieved by scanning λ_p to check whether the connection identifier is present. If a PP is not found, then all the respective WUTs in λ_w are updated by removing all instances of the terminated connection's identifier. The respective free wavelength counters, on each link that were used by the WP are incremented by 1. If a PP is found then both $\lambda_w$and λ_p and counters are updated by removing instances of the identifier and updating all necessary counters. The free wavelengths counters corresponding to WP links are decremented by one. The values corresponding to protection links are decremented only if the respective protection wavelength was not being shared. On a link where the protection wavelength was shared, the connection identity is removed from that wavelength and the free wavelength counter is not changed.

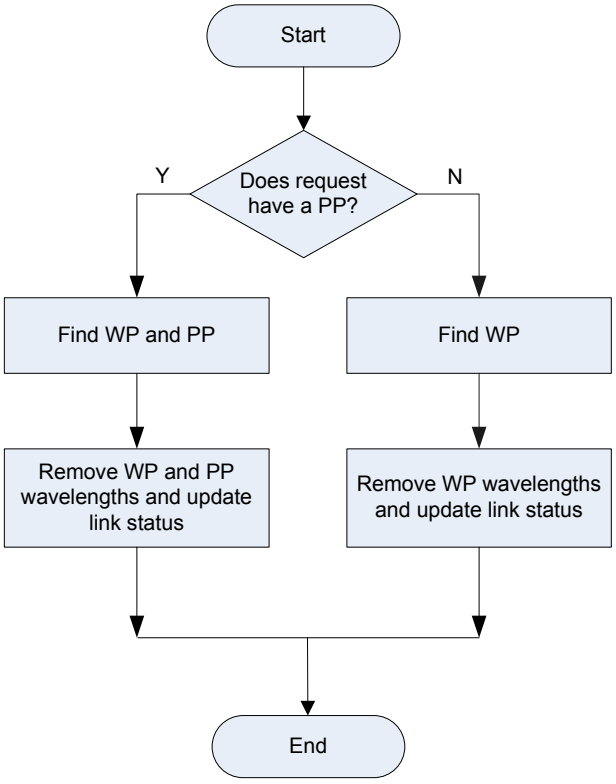

FIGURE 4.7: Flowchart showing the RASP connection termination procedure.

4.4.5 CSP Termination Request

Figure 4.8 describes the termination process of the CSP algorithm. When compared with Figure 4.7, the only difference that is observed, is that CSP does not need to determine whether a PP path is used or not, since by definition all established connections are routed with a PP. All WUTs related to the WP and PP links are updated as described earlier.

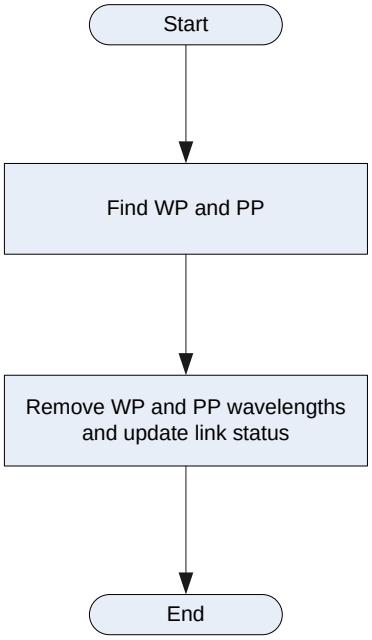

FIGURE 4.8: Flowchart showing the CSP connection termination procedure.

CHAPTER FIVE
RESULTS

5.1 INTRODUCTION

In this chapter, the results of the simulation study are presented. The metrics which are used to evaluate performance are first described. Results are presented in four experiments. In each experiment, the graphical results of the performance metric under review are discussed as a series of observations. Experiment 1, considers a general comparison of performance between RASP and CSP. The availability requirement (A_{req}) is uniformly distributed in a large range from 0.99 to 1 [7, 15]. The objective of this investigation is to obtain a general evaluation of the performances of both algorithms.

In addition to Experiment 1, the study also considers the effect of varying availability requirements on performance. Connection reliability can be measured via the metric of connection availability [15], which can be used as a QoS parameter [21]. Even though WDM networks are able to offer high bandwidth with high reliability guarantees, high levels of service may be costly and not required by all clients. In such instances, some clients would value the option of more affordable contracts with SLA's that stipulate lower levels of reliability. Therefore in Experiments 2,3 and 4, three availability requirement ranges are considered to simulate three classes of service [21]. These are summarized in Table 5.1. In the respective simulations, the availability requirements are uniformly distributed from the lower to higher bounds of the range. A list of all obtained results can be found in Appendix B.

TABLE 5.1: Three classes of availability requirements.

Range	Minimum	Maximum
Low	0.99	0.9995
Medium	0.995	0.99999
High	0.9995	1

5.1.1 Performance Metrics

Four parameters are used to evaluate the performance of the CSP and RASP. The network load is varied from 20 to 100 Erlangs in increments of 5 Erlangs. The four performance metrics that are considered are Blocking Probability (BP), Reliability Satisfaction Ratio (RSR), Backup Primary Resource Ratio (BPR) and Backup Success Ratio (BSR). In addition to these performance parameters, the reasons for the blocking of connections are studied to further understand and explain the trends in the obtained performance results.

5.1.1.1 Blocking Probability

BP is a performance metric that is widely used to measure the performance of dynamic routing algorithms [30]. It is defined as the percentage of blocked connections over the total number of connection arrival requests [19]. This is described by the equation

$$\text{BP} = \frac{\text{Number of blocked connection requests}}{\text{Total number of connection requests}} . \tag{5.1}$$

In general, connections are either blocked due to insufficient resources or due to connections not meeting the reliability requirements (in the case of RASP). A low BP is considered advantageous because it would indicate that a low percentage of connections are blocked or conversely that a high percentage of connection requests are accepted and provisioned.

5.1.1.2 Backup Primary Resource Ratio

BPR is defined as the percentage of utilized backup wavelength links over the total number of utilized primary wavelength links as described by the equation [2, 15]

$$\text{BPR} = \frac{\text{Total number of links used for backup connections}}{\text{Total number of links used for primary connections}} \, . \tag{5.2}$$

The BPR is a metric which evaluates how algorithms utilize available bandwidth. The BPR also provides an indication for the extra resources needed for providing protection as a percentage of the amount of resources required without protection. A low BSR would be considered advantageous, since in general protection bandwidth is considered to be an overhead.

5.1.1.3 Reliability Satisfaction Ratio

The RSR is described by the equation

$$\text{RSR} = \frac{\text{Number of established connections with availability } \geqslant \text{ A}_{\text{req}}}{\text{Total number of established connections}} \, , \tag{5.3}$$

and refers to the ratio of provisioned connections that meet the availability requirements to all provisioned connections. The RSR, therefore considers accepted connections but does not discriminate whether connections use protection or not. The reliability of a connection, which is represented in this study by the availability of that connection, may be increased by having as short a path as possible. Having a protection path also increases the availability of a connection [2]. The RSR measures the ability of a scheme to provision dependable connections. A high RSR indicates that a high percentage of established connections met the promised QoS requirements.

5.1.1.4 Backup Success Ratio

The BSR is the ratio of the total number of reliable connections that were established with protection to the total number of connections that required protection [2, 24]. The BSR measures the efficiency of an algorithm to provision connections that gain reliability with the addition of backup protection paths. Those connections that are routed without a PP are not considered by BSR. Furthermore, the BSR does not discriminate against CSP, which is able to establish protected connections, even though they may not meet their expected QoS requirements. The total number of reliable connections routed with a backup path is used in the calculation of the BSR. [24]. The BSR is described by the equation

$$\text{BSR} = \frac{\text{Total number established connections with protection}}{\text{Total number of connections that required protection}} . \tag{5.4}$$

5.1.1.5 Blocked Connections

Though not a performance metric, the constitution of blocked connections is also analyzed. These results provide another perspective into the performance analysis of RASP and CSP. Analysis of how connections are blocked provide some justification for the results obtained with regards to the other performance parameters. Blocked connections have been classified into three categories, connections blocked due to:

1. Insufficient resources for establishing a working path.
2. Insufficient resources for establishing a working/protection path pair.
3. Inadequate reliability of the working/protection path pair.

It is important to note that these percentages are interdependent since they are calculated as a percentage of the total blocked connections for the specific simulation scenario. They also have a sequential characteristic, since a connection is first blocked if there are no wavelengths available to establish a suitable WP. If a possible working path is discovered and found to be unreliable, a protection path is sought after to improve the availability of the connection.

Secondly, the connection may be blocked due to insufficient resources to find a suitable protection path. Thirdly, if sufficient resources are available to establish a possible protected path pair, then the connection may be blocked due to insufficient reliability of the path pair. In general, if blocking due to low WP resources is high, then it suggests that the network traffic is being dominated by WPs. If the blocking due to low WP/PP resources is high, then it suggests that the available protection bandwidth is insufficient to route suitable shared-backup protection paths. If the blocking due to low reliability is high, then it indicates the the QoS constraints are demanding.

5.1.2 Simulation Environment

The simulation software was written in the $MATLAB^{®}$ [27] programming language. Each simulation considered 10^6 arrival requests [2, 19] resulting in a total of 2 x 10^6 computations (arrivals and terminations). This large number of arrivals are simulated to achieve a narrow confidence interval. On a desktop PC (Pentium 4, 2.4 GHz, 512 MB RAM), the simulation time varied from 4 to 5 hours. The long simulation times prolonged software development since a performance graph requiring performance measurements at 17 different network loads (20 to 100 Erlangs), would take up to 68 hours. A cluster computing facility was therefore used to expedite the development and testing. The cluster comprises 4 nodes, each with a Pentium dual core processor and 2 GB of RAM. Depending on its usage, up to 8 simulations could be run simultaneously. This proved advantageous during the design and test phases by greatly reducing simulation time.

5.2 EXPERIMENT 1: GENERAL COMPARISON

In the first experiment, the availability requirement was varied from 0.99 to 1 [7, 15]. These limits cover the full range of link availabilities as shown in 4.2. The experiment provided the opportunity to determine and understand the trends in performance of both schemes as the network load is varied.

5.2.1 Blocking Probability $(0.99 < a_r < 1)$

Figure 5.1 shows the performance of CSP and RASP with respect to blocking probability.

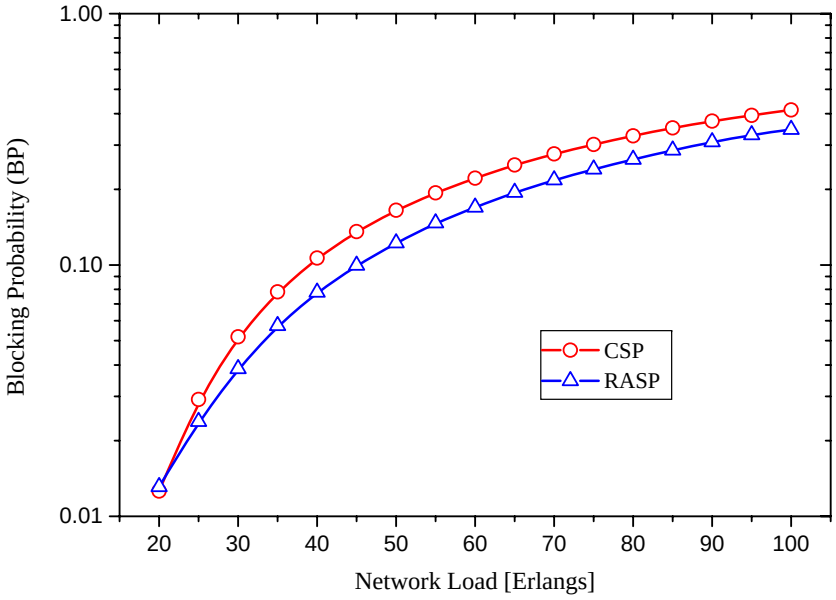

FIGURE 5.1: Blocking Probability versus Network Load ($0.99 < a_r < 1$).

Observation 1. The blocking probability in both cases increased with increasing network load. This is an expected result since at higher networks loads, the increased rate of connection arrivals results in a greater percentage of connections being blocked due to limited network resources.

Observation 2. In comparison, RASP had a lower blocking probability than CSP, with the difference in BP between the two algorithms decreasing at lower loads. In the case of RASP, when the reliability of a working path is greater than the availability requirements,

it is provisioned without a backup path. In addition, reliable partial link disjoint paths are acceptable. This saves resources and frees up bandwidth for subsequent arrival requests to use, resulting in a lower BP. CSP establishes all connections, using a fully link disjoint working/backup path pair, without concern for reliability requirements. This results in a larger percentage of bandwidth being utilized for protection. In the case of CSP, the advantage shared protection was not sufficiently effective to compensate for limited network resources which resulted in a higher BP.

Observation 3. As the load decreased, RASP continued to block connections due to the reliability requirements, and the QoS constraints had a greater impact on blocking compared with the resource constraints experienced by CSP. This caused the respective BP graphs to converge. At a load of 20 Erlangs, CSP had a slightly lower BP than RASP, due to the impact of the reliability requirement constraints on RASP. This trend is expected to continue at loads below 20 Erlangs.

5.2.2 Backup Primary Resource Ratio $(0.99 < a_r < 1)$

Figure 5.2 shows the Backup Primary Ratios for RASP and CSP.

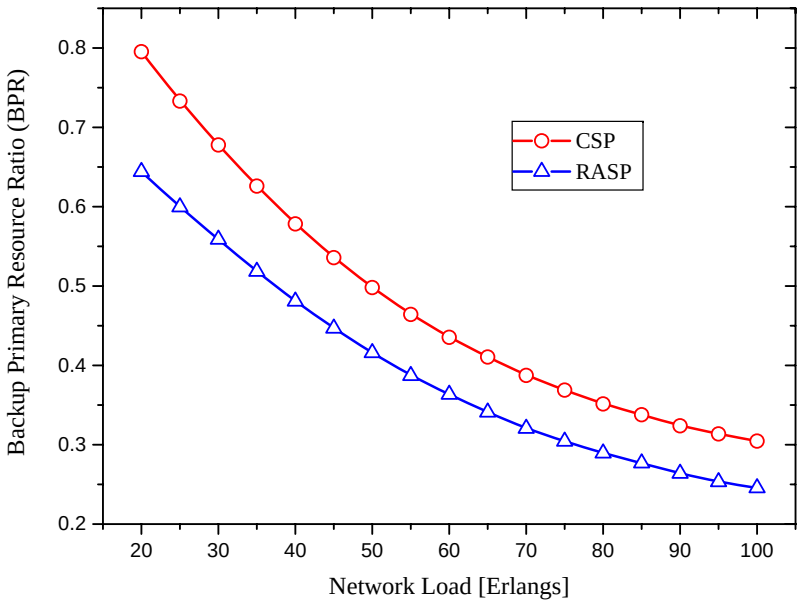

FIGURE 5.2: Backup Primary Resource Ratio versus Network Load $(0.99 < a_r < 1)$.

Observation 1. In both cases, the BPR decreased with increasing network load. As the load increases, more bandwidth resources were used to satisfy and provision more connections. This results in more bandwidth being used for working paths. However, due to the sharing of backup paths, protection bandwidth increases at a lower rate, resulting in a lower BPR.

Observation 2. Under all considered load conditions, RASP presented a lower BPR than CSP. The reasons are the same as discussed earlier. RASP is able to establish reliable working paths without backup protection. The effect is emphasized at lower loads where the difference in the BPR of the respective algorithms is greater. CSP always establishes a pair of fully link disjoint working and protection paths resulting in more bandwidth being used for protection and this results in a higher BPR.

5.2.3 Reliability Satisfaction Ratio $(0.99 < a_r < 1)$

Figure 5.3 presents the performance of RASP and CSP with respect to their Reliability Resource Ratio. The figure shows that RASP was able to ensure that all provisioned connections met the respective reliability requirements. At low loads CSP was able to ensure that 99.5% of provisioned connection were dependable and approximately 99.1% of provisioned connections were dependable at higher loads. At higher loads, since CSP is required to provision every connection with a link disjoint path pair, regardless of its reliability, fewer free resources are available. CSP is therefore constrained and the possibility of establishing longer path pairs with lower reliability is high. This results in more connections violating the QoS requirements, resulting in a lower RSR. With the above argument, it is expected that CSP will perform poorly with respect to RSR when QoS requirements are more demanding.

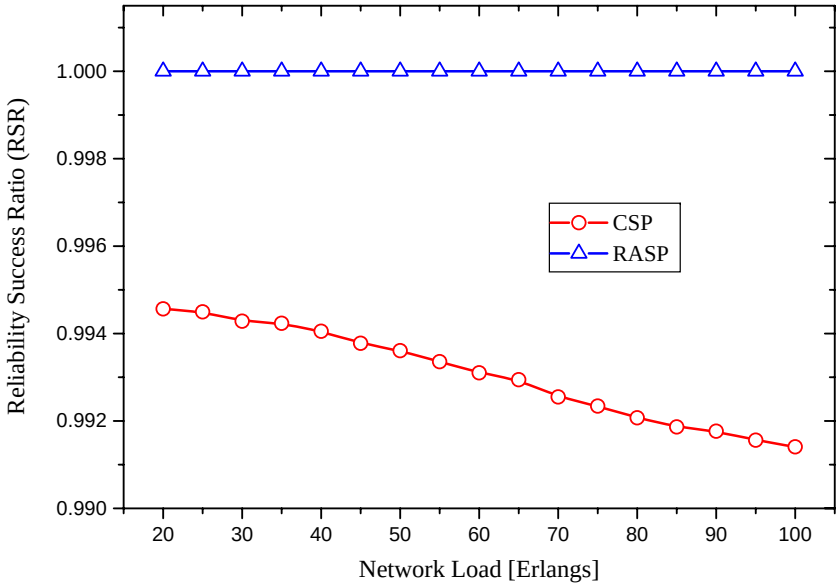

FIGURE 5.3: Reliability Satisfaction Ratio versus Network Load ($0.99 < a_r < 1$).

5.2.4 Backup Success Ratio $(0.99 < a_r < 1)$

Figure 5.4 presents the Backup Success Ratio performance of both CSP and RASP.

Observation 1. The BSR of both CSP and RASP decreased with an increase in network load. The reasons for this trend are the same as discussed under *Observation 1* of the Blocking Probability section. In the case of RASP, the BSR decreased due to insufficient resources and inadequate reliability, resulting in the blocking of more connections that required PPs.

Observation 2. CSP performed better than RASP with respect to the BSR at all loads. This is firstly due to RASP being able to establish reliable connections without backup protection, consuming a larger percentage of working path bandwidth which is not shared. This statement is supported by RASP also having a lower BPR. The remaining protection bandwidth is insufficient to cater for connections requiring backup, resulting in more of these connections being blocked.

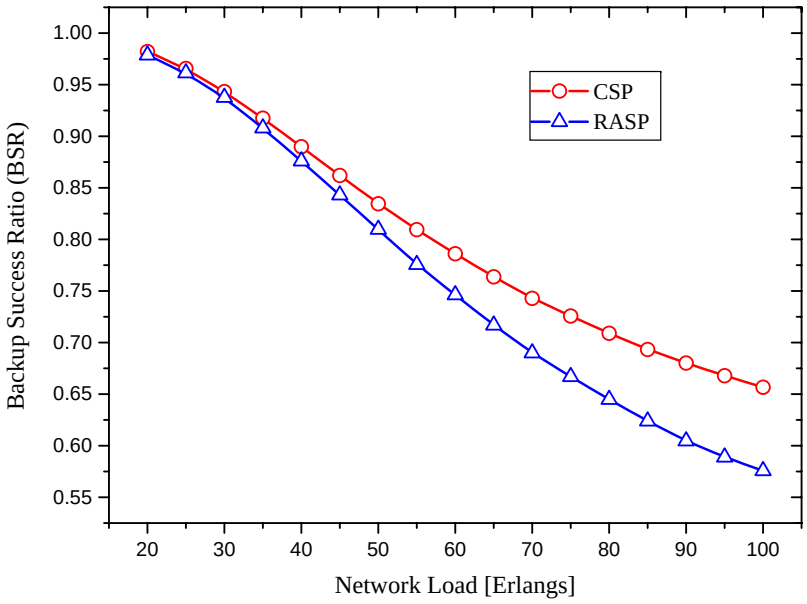

FIGURE 5.4: Backup Success Ratio versus Network Load ($0.99 < a_r < 1$).

Secondly, RASP blocks possible protected connections if their availabilities do not meet the respective requirements even if sufficient resources are available to route these connections. The result also indicates that RASP's ability to use link disjoint protection paths was not effective enough to compensate for resource deficiencies.

Observation 3. As the load increased, CSP displayed increasingly better performance as the graphs diverge. At low loads, there are resources available to establish reliable working/protection path pairs and both CSP and RASP had similar BSR performance. With CSP, all connections require backup protection, therefore a greater percentage of bandwidth is dedicated to protection which can be shared. This point is indicated by the BPR performance in Figure 6.2. Furthermore, CSP does not consider reliability requirements, which is indicated by the lower RSR in Figure 5.3, where fewer protected connections were blocked due to their low availability.

5.2.5 Blocked Connections $(0.99 < a_r < 1)$

The constitution of blocked connections for RASP and CSP are graphically presented in Figure 5.5 and Figure 5.6 respectively. The graphs provide information about the contribution of each cause of blocking to the total blocking.

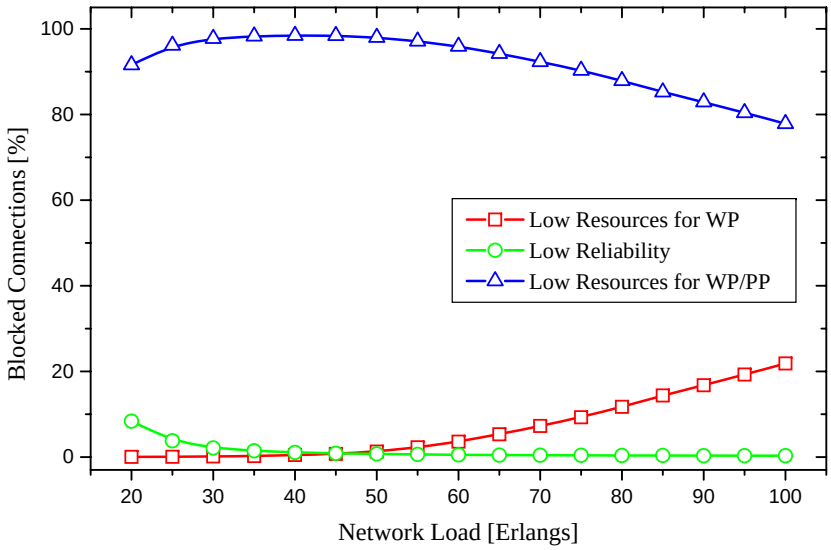

FIGURE 5.5: Blocked connections during RASP simulation ($0.99 < a_r < 1$).

Observation 1: Low Resources for WP. The percentage of connections blocked due to insufficient WP resources increased with increasing network load. At higher network loads, approximately 5-10% more connections were blocked in the case of CSP. This is an expected result and the reasons for this have been discussed above under *Observation 1* of the Blocking Probability section.

Observation 2: Low Resources for WP/PP. The percentage of blocked connections due to low resources to establish a WP/PP path pair was by far the greatest contributor to blocking. There is a decreasing trend with an increase in network load. This is an unexpected result but is however directly linked to *Observation 1* above. This decrease with network load is due to the increase in blocking due to low WP resources. Since an increasing percentage of connections were blocked due to low WP resources, bandwidth becomes available for the establishment of possible WP/PP pairs, which have the required availability to be established. The ability to

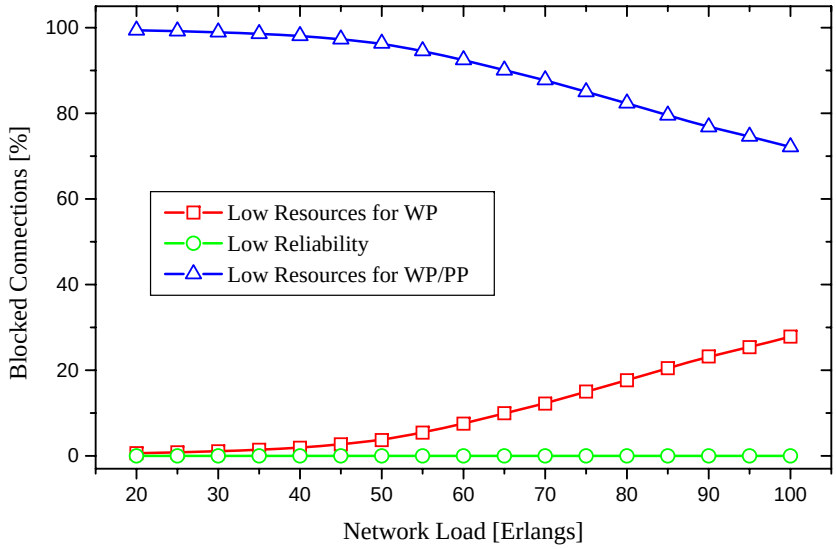

FIGURE 5.6: Blocked connections during CSP simulation ($0.99 < a_r < 1$).

utilize WP links for the partial link disjoint protection paths assists in decreasing the blocking due to low resources for establishing WP/PP pairs. In the case of CSP, approximately 5-10% more connections were blocked due to low resources to establish WP/PP paths. This is due to the requirement that all connections be established with a link disjoint PP. In the case of RASP, it is noted that at very low network loads, the blocking due to low resources to establish WP/PP paths did begin to decrease. This is due to sufficient network resources being available and the ability of RASP to establish connections without PP's or with partial link disjoint PPs.

Observation 3: Low Reliability. Low reliability only affects the blocking of RASP since CSP is not reliability aware. The graph for the CSP simulation indicates this with values of zero for all network loads. In the case of RASP, it is observed that low reliability had a small impact on blocking. The impact was greater at low loads when the effect of the other two causes of blocking were minimal. At low loads, there is sufficient bandwidth resources to establish connection requests and a greater percentage of connections requiring protection paths are blocked due to low reliability. This result is linked to *Observation 3* of the Blocking Probability section. At higher loads, when bandwidth resources are limited, there is a higher probability of connections being blocked due to resource constraints.

5.3 EXPERIMENT 2: LOW AVAILABILITY REQUIREMENT $(0.99 < a_r < 0.9995)$

In this experiment, the effect of low reliability requirements was investigated. The availability requirement was varied from 0.99 to 0.9995. These results are compared with those obtained in Experiment 1, which considered low reliability requirements.

5.3.1 Blocking Probability $(0.99 < a_r < 0.9995)$

Figure 5.7 shows the performance of CSP and RASP with respect to blocking probability under low QoS constraints.

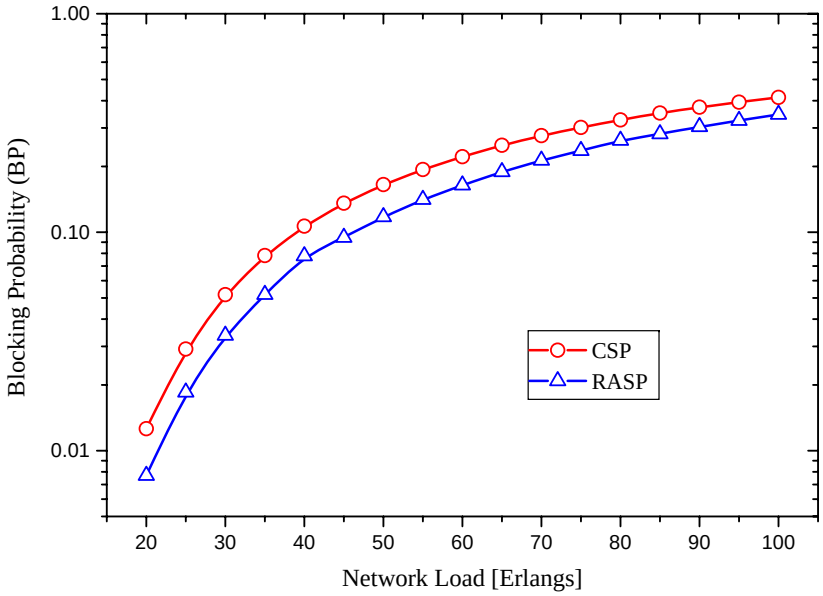

FIGURE 5.7: Blocking Probability versus Network Load ($0.99 < a_r < 0.9995$).

Observation 1. As in Experiment 1, the blocking probability in both cases increased with increasing network load. The reasons for this remain the same as discussed previously.

Observation 2. In comparison, RASP had a lower blocking probability than CSP, with the difference in BP between the two algorithms decreasing at lower loads. The justifications remain the same as discussed previously.

Observation 3. The main difference observed when comparing this result with Figure 5.1, is that BP of RASP and CSP converged at a much lower rate than before, as the load decreased. This is due to the average availability requirement constraint being lower and therefore, it had a lesser impact on RASP than compared with the resource constraints experienced by CSP. Furthermore, the graphs do not intersect at low network loads. It is nor clear whether they may intersect at some lower network load.

5.3.2 Backup Primary Resource Ratio $(0.99 < a_r < 0.9995)$

Figure 5.8 shows the Backup Primary Ratios for RASP and CSP.

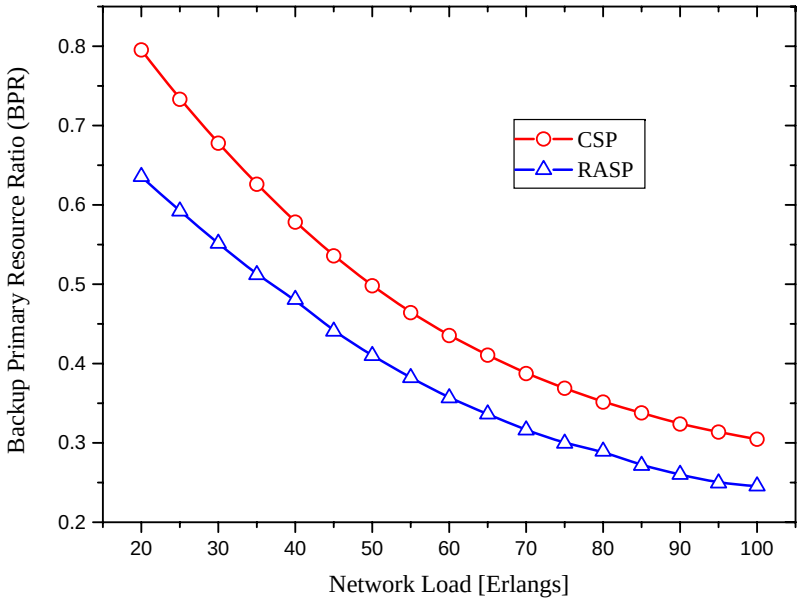

FIGURE 5.8: Backup Primary Resource Ratio versus Network Load $(0.99 < a_r < 0.9995)$.

Observation 1. In both cases, the BPR decreased with increasing network load. The graphs obtained here are similar to those in Experiment 1. The values for the BPR performance of RASP (see Appendix B) are marginally lower (approximately 1%). One would expect a lower BPR since lower availability requirements suggest that more working paths would meet these requirements. Since more connections would be established without PPs, the BPR would decrease. A greater decrease in the BPR performance compared with Experiment 1, would be expected if the QoS constraints were lower.

Observation 2. Under all considered load conditions, RASP exhibited a lower BPR than CSP. This justifications for this observation have been discussed before.

5.3.3 Reliability Satisfaction Ratio $(0.99 < a_r < 0.9995)$

Figure 5.9 presents the performance of RASP and CSP with respect to their Reliability Resource Ratios.

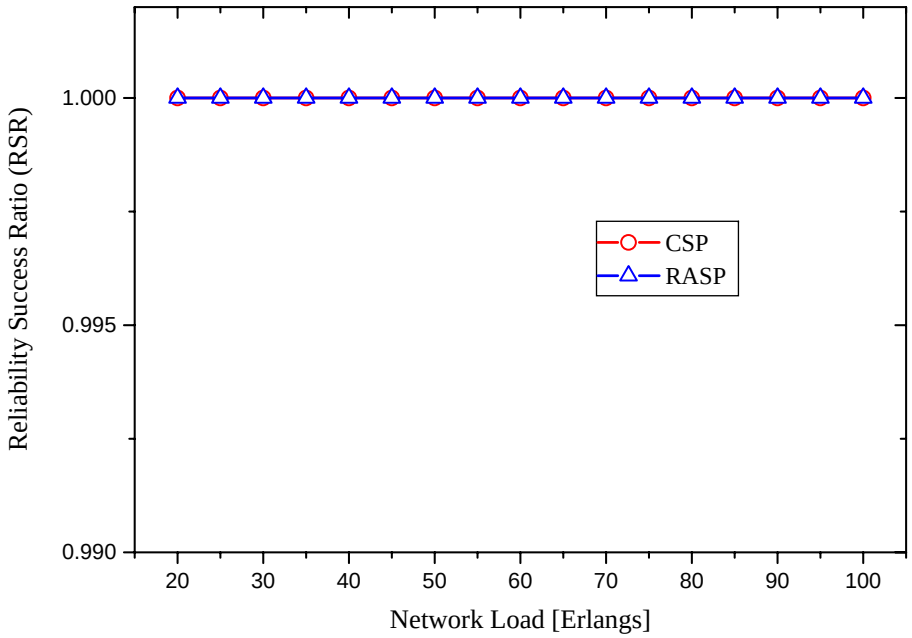

FIGURE 5.9: Reliability Satisfaction Ratio versus Network Load ($0.99 < a_r < 0.9995$).

This result shows that both RASP and CSP were able to ensure that all provisioned connections met the respective reliability requirements by obtaining a value of 1 for the RSR at all simulated network loads. The RSR considers accepted connections and describes the ability of an algorithm to provision connections that meet their reliability requirements. Since the availability requirements are lower, and since the RSR considers accepted connections only, all the connections routed by CSP were able to meet their QoS requirements. The compulsory routing of link disjoint protected path pairs increased the availability of the routed connections. This RSR result suggests that under low QoS constraints, RASP does not exhibit any advantages over CSP.

5.3.4 Backup Success Ratio $(0.99 < a_r < 0.9995)$

Figure 5.10 compares the Backup Success Ratio of CSP and RASP.

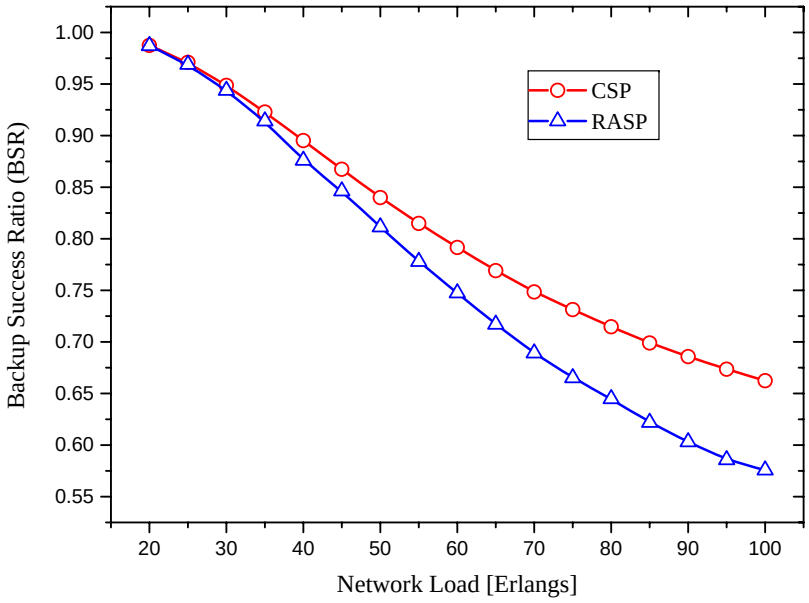

FIGURE 5.10: Backup Success Ratio versus Network Load ($0.99 < a_r < 0.9995$).

Observation 1. The BSR of both CSP and RASP decreased with an increase in network load. The reasons for this trend are the same as discussed previously. Compared with the performance in Experiment 1, results show that both RASP and CSP exhibit better BSR performance under lower availability constraints. This is expected since lower constraints allow more potential protected paths to meet the requirements and be established, thus increasing the BSR. Furthermore, the lower QoS constraints result in a fewer number of connections requiring protection, which contributes to increasing BSR.

Observation 2. CSP performed better than RASP with respect to the BSR at all loads. The lower availability constraints produce the same comparison as observed earlier in Experiment 1. The same justification is applicable.

Observation 3. As the load increased, CSP displayed increasingly better performance as the graphs diverge. This observation is also repeated.

5.3.5 Blocked Connections $(0.99 < a_r < 0.9995)$

Figures 5.11 to 5.15 show how connections have been blocked in Experiment 2. Figure 5.11 and Figure 5.12 consider blocked connections for RASP and CSP respectively and Figures 5.13 to 5.15 look independently at the causes of blocking, comparing RSP and CSP in each case.

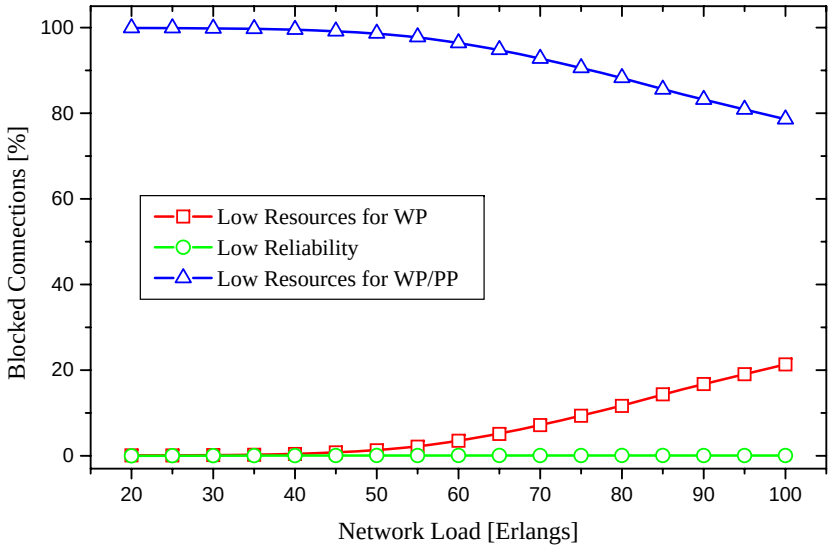

FIGURE 5.11: Blocked connections during RASP simulation ($0.99 < a_r < 0.9995$).

Observation 1: Low Resources for WP. With respect to Figures 5.11 and 5.12, the percentage of connections blocked due to insufficient WP resources increased with increasing network load. With respect to Figure 5.13, it is observed that more connections are blocked in the case of CSP with an approximate 5-7% difference at higher network loads.

Under low QoS constraints, RASP was able to provision more WPs without PPs, while CSP was forced to establish a link disjoint PP for all connections. More PPs result in comparatively fewer resources being available to provision WPs. In comparison with Experiment 1, RASP's performance with respect to low WP resources was marginally higher at very low loads.

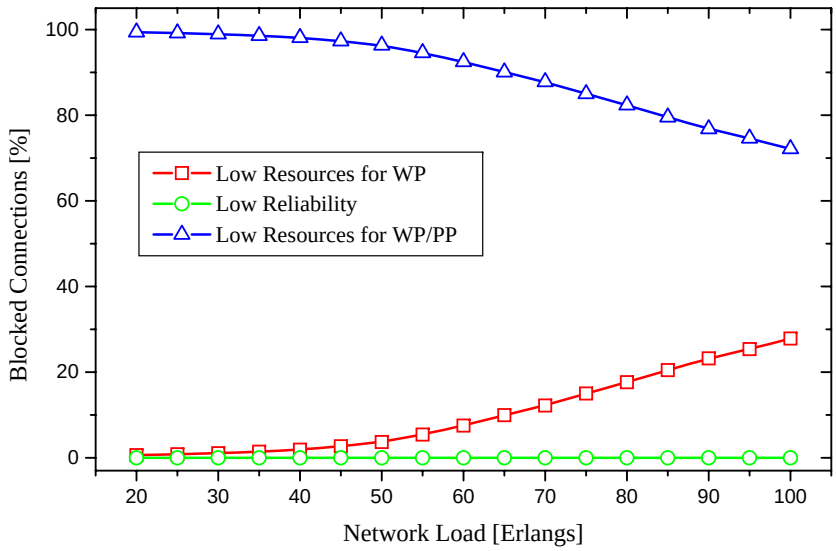

FIGURE 5.12: Blocked connections during CSP simulation ($0.99 < a_r < 0.9995$).

Observation 2: Low Resources for WP/PP. The percentage of blocked connections due to low resources to establish a WP/PP path pair, generally decreased with the increase in network load. This result is contrary to *Observation 1*. This decrease with network load is due to the increase in blocking due to low WP resources. As the network load increases, an increasing percentage of connections are blocked due to low WP resources and bandwidth becomes available for the establishment of possible dependable WP/PP pairs.

With respect to Figure 5.14, approximately 10% more connections were blocked in the case of RASP. This result is contrary to the observation in Experiment 1. In the case of RASP, the low QoS constraints results in a large percentage of connections being established without protection. Therefore a large percentage of unshared bandwidth is dedicated to WP establishment, causing increased blocking of WP/PP pairs.

It is noted that at very low network loads, in the case of RASP, that the blocking of WP/PP pairs did decrease as observed before, but to a lesser degree [see Appendix B]. This is again due to the high availability of network resources at low loads, and the ability of RASP to establish connections without PP's or with partial link disjoint PPs.

Observation 3: Low Reliability. As mentioned earlier, low reliability is a cause for blocking only in the case if RASP. In addition, a connection is blocked only if a WP/PP pair does not meet it availability requirement. In this experiment with low availability requirements, it is observed in Figure 5.15, that no connections were blocked due to low reliability. From this it can be deduced that blocking is completely due to bandwidth limitations. This result justifies the results obtained for the RSR.

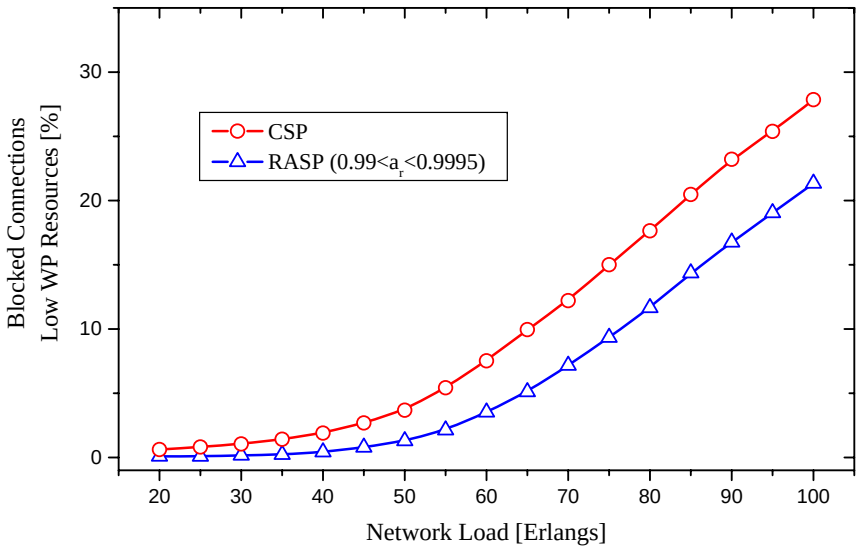

FIGURE 5.13: Percentage of connections blocked due to Low WP Resources ($0.99 < a_r < 0.9995$).

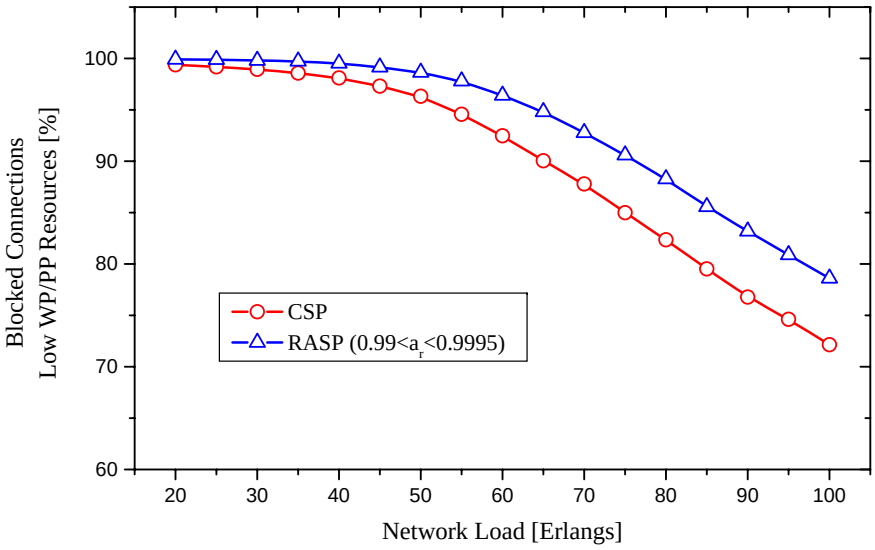

FIGURE 5.14: Percentage of connections blocked due to Low WP/PP Resources ($0.99 < a_r < 0.9995$).

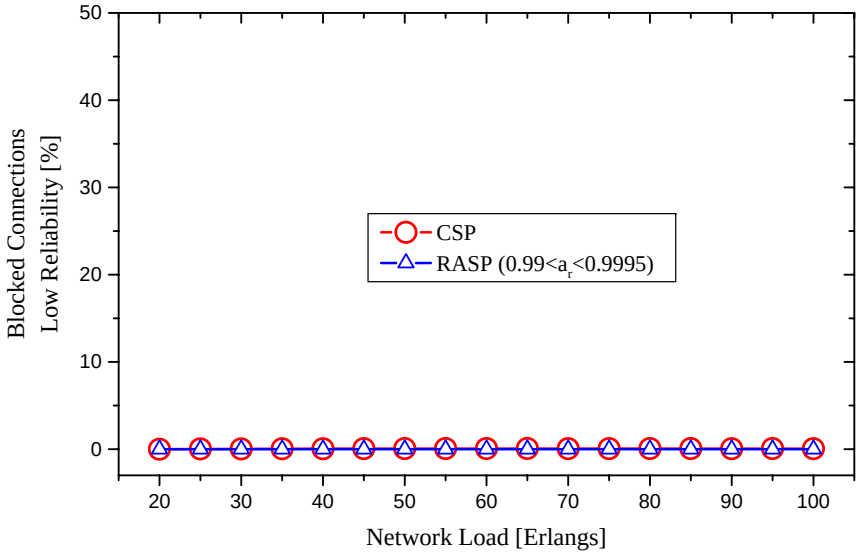

FIGURE 5.15: Percentage of connections blocked due to Low Reliability ($0.99 < a_r < 0.9995$).

5.4 EXPERIMENT 3: MEDIUM AVAILABILITY REQUIREMENT

In this experiment, the effect of medium reliability requirements was considered and the availability requirement was varied from 0.995 to 0.99999. These results are compared with those obtained in Experiment 2, which considered low reliability requirements.

5.4.1 Blocking Probability $(0.995 < a_r < 0.99995)$

The blocking performance of RASP and CSP is shown in Figure 5.16.

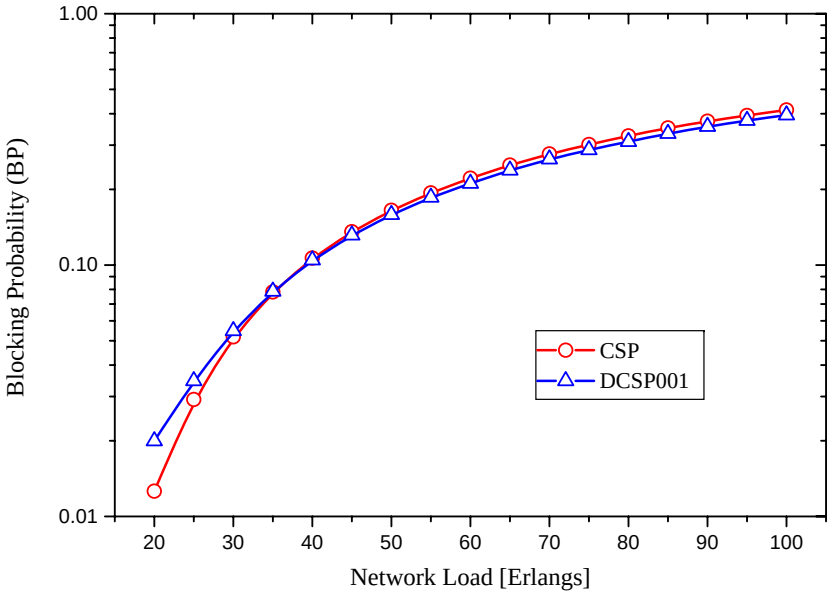

FIGURE 5.16: Blocking Probability versus Network Load ($0.995 < a_r < 0.99999$).

Observation 1. The blocking probability in both cases increased with increasing network load. The reasons for this remain the same as discussed previously.

Observation 2. In comparison, the BP of both RASP and CSP are very similar. This result is in contrast with the result obtained in Experiment 2, where a larger difference in BP was observed with lower availability requirements. Since the blocking performance of CSP is not affected by the changes in availability requirements, one can conclude that the blocking probability of RASP has increased due to an increase in the availability requirements, which is an expected result.

Observation 3. On closer inspection it is observed that there was a divergence at lower and higher network loads. At low network loads RASP had a higher BP and as the load increased, the blocking probability of CSP increased at a higher rate than RASP. This resulted in RASP having better blocking performance at higher network loads. The other example of this behavior was noticed in in Experiment 1 (Figure 5.1), where it was observed that the two curves converged at approximately 20 Erlangs. Here the two graphs converged at approximately 40 Erlangs. In Experiment 1, it was expected that since the BP graphs converge slowly, that they may intersect at a very low network load. In this experiment, the intersection occurs at a higher network load suggesting that the increased QoS constraints increases the network load threshold at which RASP begins to outperform CSP with respect to BP.

5.4.2 Backup Primary Resource Ratio $(0.995 < a_r < 0.99995)$

The BPR performance of RASP and CSP is shown in Figure 5.17.

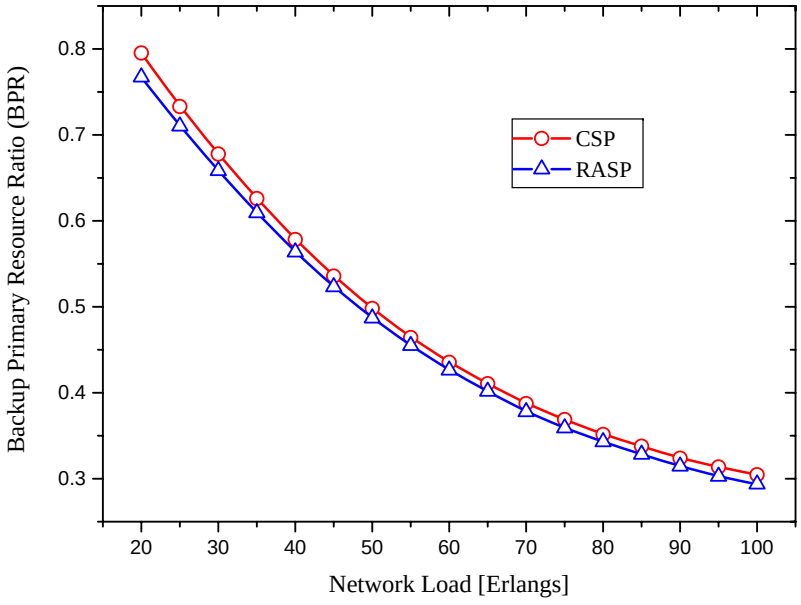

FIGURE 5.17: Backup Primary Resource Ratio versus Network Load ($0.995 < a_r < 0.99999$).

Observation 1. In both cases, the BPR decreased with increasing network load. The graphs obtained here are similar to those obtained in Experiments 1 and 2.

Observation 2. Under all considered load conditions, RASP presented a lower BPR than CSP. This result is repeated for the same reasons as discussed previously. On inspection of the values (see Appendix B), one would notice that the values for the BPR performance of RASP was approximately 5-10% greater than in Experiment 2, causing the difference in BPR performance between CSP and RASP to decrease to approximately 1-3% . This is attributed to greater restrictions in terms of availability requirements. Greater availability requirements result in RASP utilizing more backup resources to route connections and this results in a larger BPR.

5.4.3 Reliability Satisfaction Ratio $(0.995 < a_r < 0.99995)$

The RSR performance of RASP and CSP is shown in Figure 5.18.

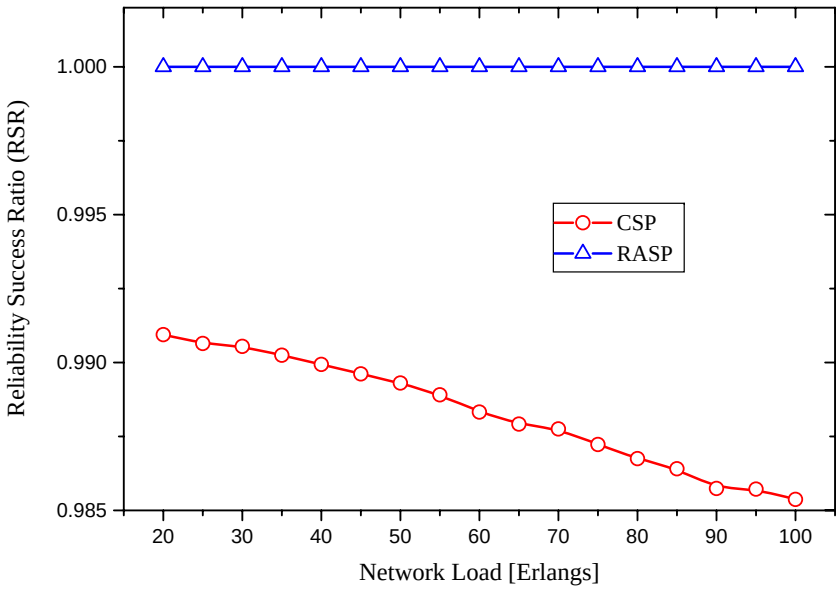

FIGURE 5.18: Reliability Satisfaction Ratio versus Network Load ($0.995 < a_r < 0.99999$).

This result indicates a similar trend to that obtained in Experiment 1. In comparison with Experiment 2, it is observed that the RSR performance of CSP has decreased by

approximately 1-2% . CSP is able to establish 99% of connections at low loads and approximately 98% of connections at high loads. It is expected that the RSR performance of CSP would decrease with increasing network load. Since the availability requirements are higher compared with Experiment 2, CSP was not able ensure that all provisioned connections met the required availability, while RASP was consistently able to achieve this.

5.4.4 Backup Success Ratio $(0.995 < a_r < 0.99995)$

The BSR performance of RASP and CSP is shown in Figure 5.19.

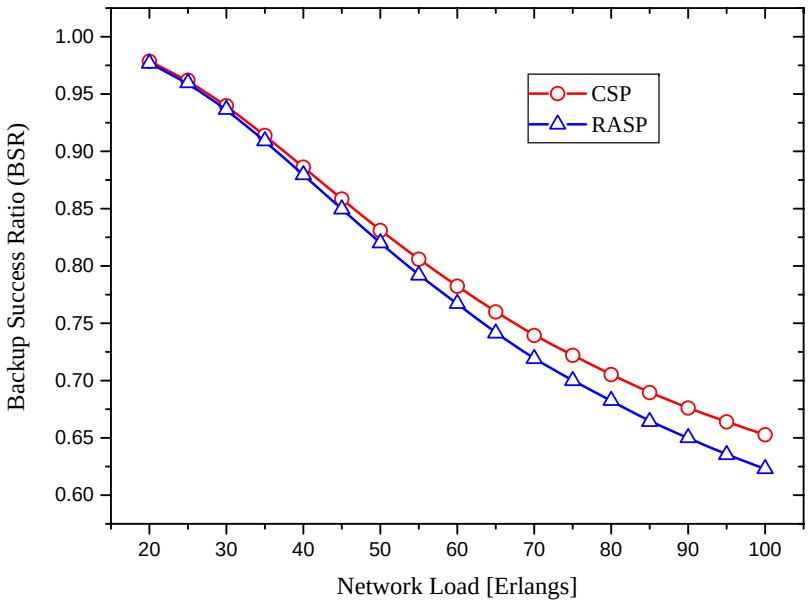

FIGURE 5.19: Backup Success Ratio versus Network Load ($0.995 < a_r < 0.99999$).

Observation 1. The BSR, of both CSP and RASP, decreased with an increase in network load. The reasons for this trend are the same as discussed previously.

Compared with the performance in Experiment 2, results show that CSP exhibited worse BSR performance in this experiment. This is due to the increase in the availability requirements, which superficially increases the number of CSP connections requiring protection. RASP exhibited better BSR performance for most network loads when comparing with Experiment 2. This can be attributed to backup sharing and the establishment of partial link disjoint PPs.

At low loads, the BSR performance of RASP decreased, due to insufficient protection resources.

Observation 2. CSP performed better than RASP with respect to the BSR at all loads. The higher average availability constraints produce the same comparison with the same justifications as mentioned earlier.

Observation 3. As the load increased, CSP displayed increasingly better performance as the graphs diverge. This observation is also repeated. It is observed that the two graphs diverge to a lesser degree and that the performance difference between CSP and RASP is lower. This is due to RASP having superior performance under higher QoS constraints, the reasons for which are described in Observation 1.

5.4.5 Blocked Connections $(0.995 < a_r < 0.99995)$

Figures 5.20 to 5.24 show how connections have been blocked in Experiment 3. Figure 5.20 and Figure 5.21 consider blocked connections for RASP and CSP respectively and Figures 5.22 to 5.24 look independently at the causes of blocking, comparing RSP and CSP in each case.

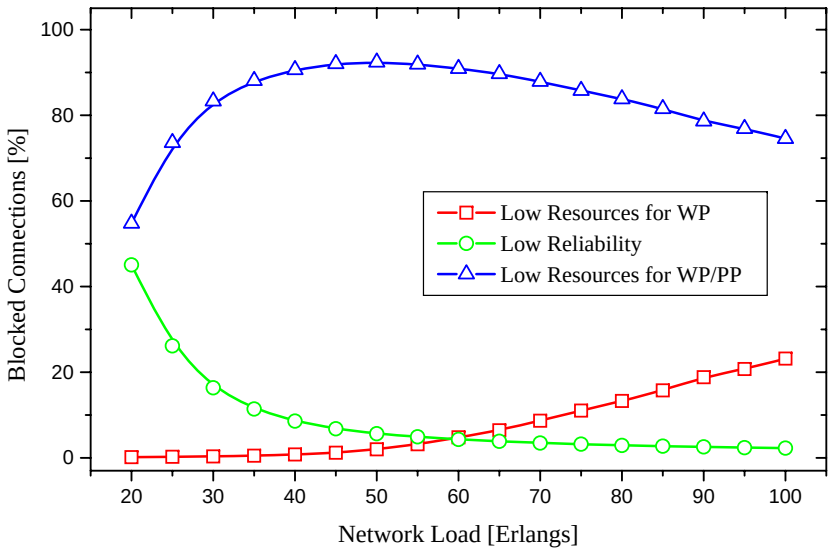

FIGURE 5.20: Blocked connections during RASP simulation ($0.995 < a_r < 0.99999$).

Observation 1: Low Resources for WP. With respect to Figure 5.20 and Figure 5.21, the percentage of connections blocked due to insufficient WP resources increased with increasing network load. There is an approximate 4-6% difference between CSP and RASP at higher network loads.

In comparison with Experiment 2, the blocking due to low WP resources, in the case of RASP, was marginally higher at all loads. This is due to the increase in the average availability requirement. With respect to Figure 5.22, it is observed that more connections were blocked in the case of CSP. This result is repeated.

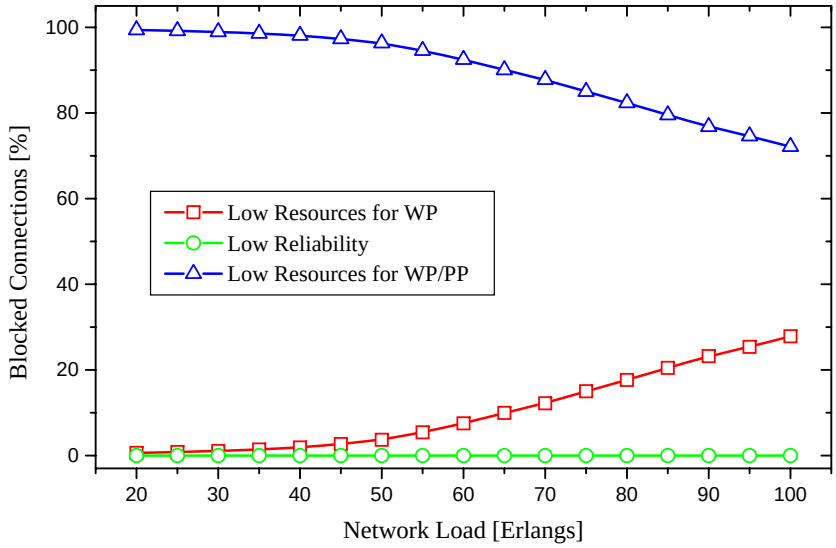

FIGURE 5.21: Blocked connections during CSP simulation ($0.995 < a_r < 0.99999$).

Observation 2: Low Resources for WP/PP. The percentage of blocked connections due to low resources to establish a WP/PP path pair, in the case of CSP, generally decreased with the increase in network load.

When compared with Experiment 2, the results obtained for RASP were different with very low blocking at low loads, an increasing trend from 20 to 45 Erlangs and finally a decreasing trend till 100 Erlangs. The distinct feature of the RASP curve is the initial increasing trend, which is due to the increased availability requirements. This suggests that possible WP/PP path pairs were blocked due to low reliability. With respect to Figure 5.23, it is observed that RASP outperformed CSP at loads lower than 70 Erlangs.

Observation 3: Low Reliability. As expected, low reliability is a cause for blocking only in the case if RASP. In addition, a connection is blocked only if a WP/PP pair does not meet it availability requirement. In this experiment with medium availability requirements, it is observed in Figure 5.24, that more connections were blocked due to low reliability at low network loads. This was correctly suspected in Observation 2 above. There was a decreasing trend for all loads from approximately 45% to 3% .

In comparison with Experiment 2, where no connections were blocked due to low reliability, the result indicates that the increase in availability requirements has had an impact.

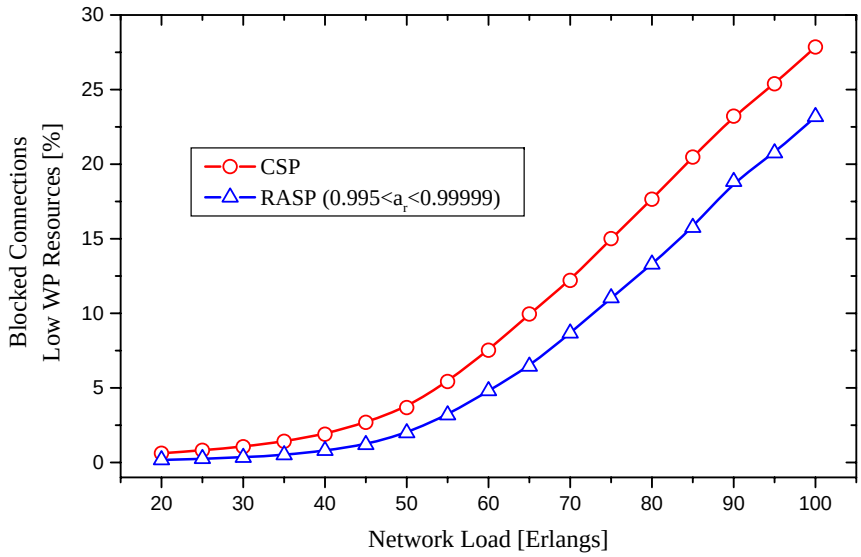

FIGURE 5.22: Percentage of connections blocked due to Low WP Resources ($0.995 < a_r < 0.99999$).

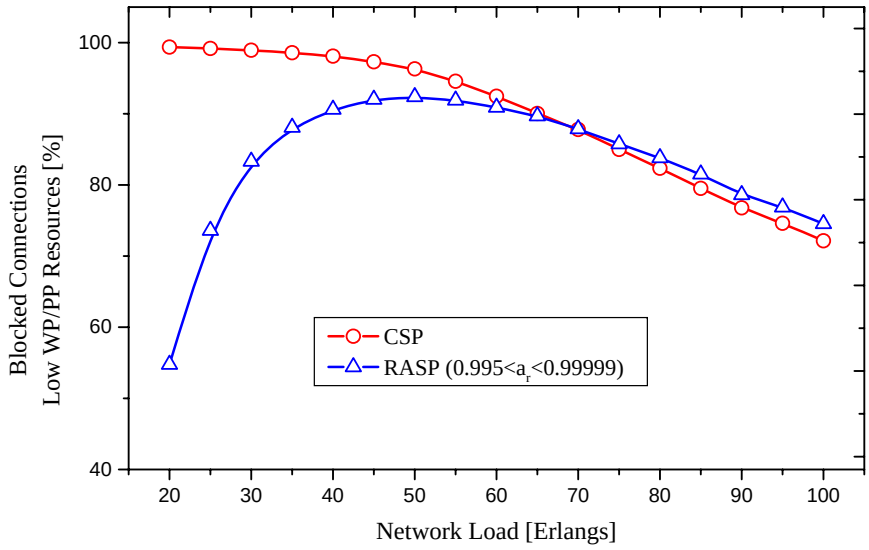

FIGURE 5.23: Percentage of connections blocked due to Low PP Resources ($0.995 < a_r < 0.99999$).

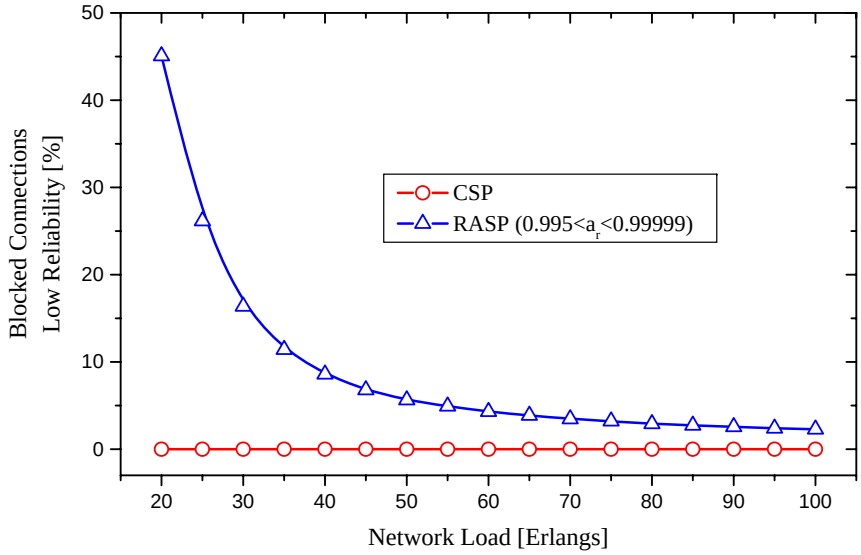

FIGURE 5.24: Percentage of connections blocked due to Low Reliability ($0.995 < a_r < 0.99999$).

5.5 EXPERIMENT 4: HIGH AVAILABILITY REQUIREMENT

5.5.1 Blocking Probability $(0.9995 < a_r < 1)$

The blocking performance of RASP and CSP is shown in Figure 5.25.

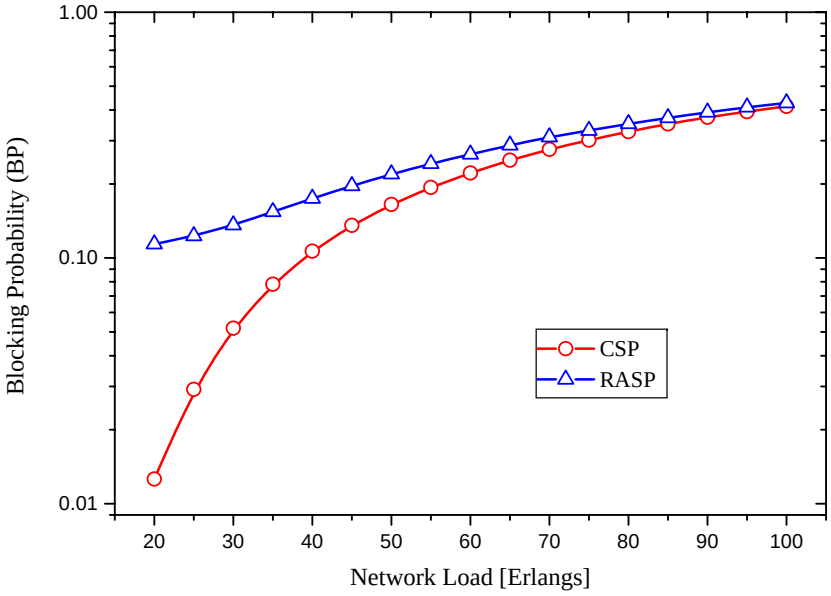

FIGURE 5.25: Blocking Probability versus Network Load $(0.9995 < a_r < 1)$.

Observation 1. The blocking probability in both cases increased with increasing network load. The reasons for this remain the same as discussed previously.

Observation 2. In comparison with results obtained in Experiments 2 and 3, it is observed that RASP had a higher blocking blocking probability than CSP. This is the first instance of this case. Since the blocking performance of CSP is not affected by the changes in availability requirements, one can conclude that the blocking probability of RASP has increased due to an increase in the availability requirements, which was also observed in Experiment 2.

Observation 3. The graph shows a divergence between the BP as the network load decreases. This divergence was noticeable in Experiment 1 and 3. In this case however, the divergence is exaggerated. This is again due to the availability awareness of RASP which, even at low loads caused more connections to be blocked due to non compliance with the

availability requirements. The exaggeration is due to the high availability requirements used in this experiment. It is also observed that there was no convergence or crossing over of curves for the considered network range loads.

5.5.2 Backup Primary Resource Ratio $(0.9995 < a_r < 1)$

The BPR performance of RASP and CSP is shown in Figure 5.26.

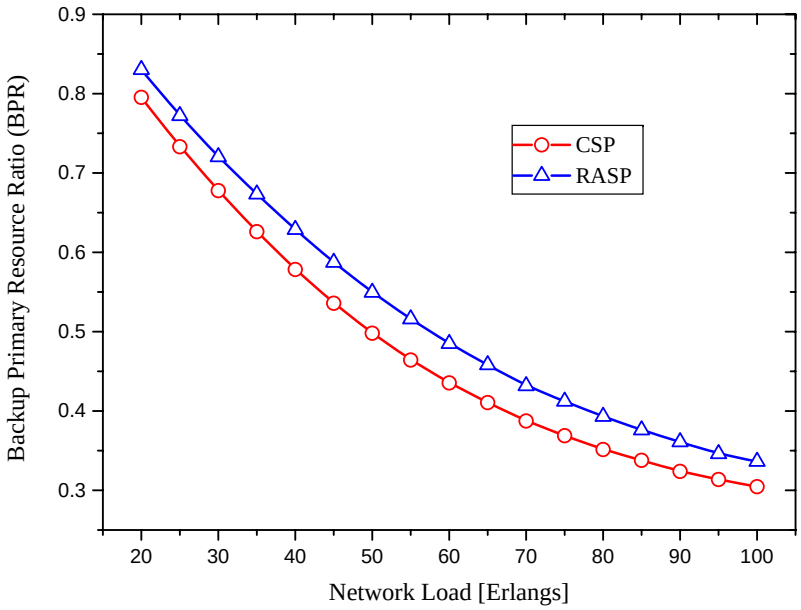

FIGURE 5.26: Backup Primary Resource Ratio versus Network Load $(0.9995 < a_r < 1)$.

Observation 1. In both cases, the BPR decreased with increasing network load. The obtained graphs display a similar trend to those in previous experiments. On closer inspection (see Appendix B), one would notice that the values for the BPR performance of RASP has increased by approximately 5% for all network loads in comparison with Experiment 3. One would expect a higher BPR, since higher availability requirements suggest that more connections would be established with PPs, resulting in the increase of the BPR.

Observation 2. Under all considered load conditions, RASP presented a higher BPR than CSP. This is the first instance of this observation. The result is due to more efficient usage of protection resources and the ability of RASP to establish link disjoint PPs.

5.5.3 Reliability Satisfaction Ratio $(0.9995 < a_r < 1)$

The RSR performance of RASP and CSP is shown in Figure 5.27.

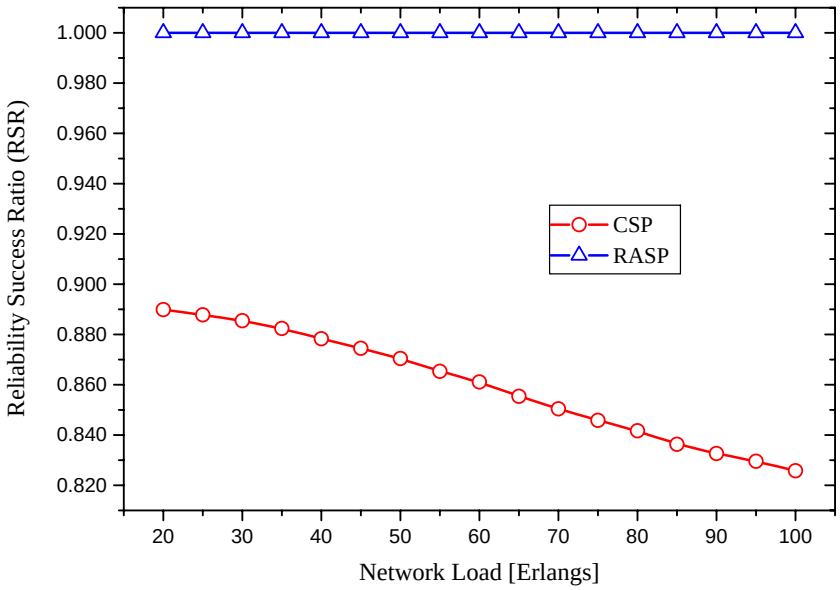

FIGURE 5.27: Reliability Satisfaction Ratio versus Network Load ($0.9995 < a_r < 1$).

This result shows a similar trend to that obtained in earlier experiments. In contrast to Experiment 3, it is observed that the RSR of CSP has decreased greatly by approximately 10-15% . CSP was able to establish 89% of connections at low loads and approximately 83% of connections at high loads. In this experiment, since the availability requirements are high, CSP was not able to provision a large percentage of connections with guaranteed availability. As before, RASP was able to consistently ensure that all provisioned connections complied with the required availability.

5.5.4 Backup Success Ratio $(0.9995 < a_r < 1)$

The BSR performance of RASP and CSP is shown in Figure 5.28.

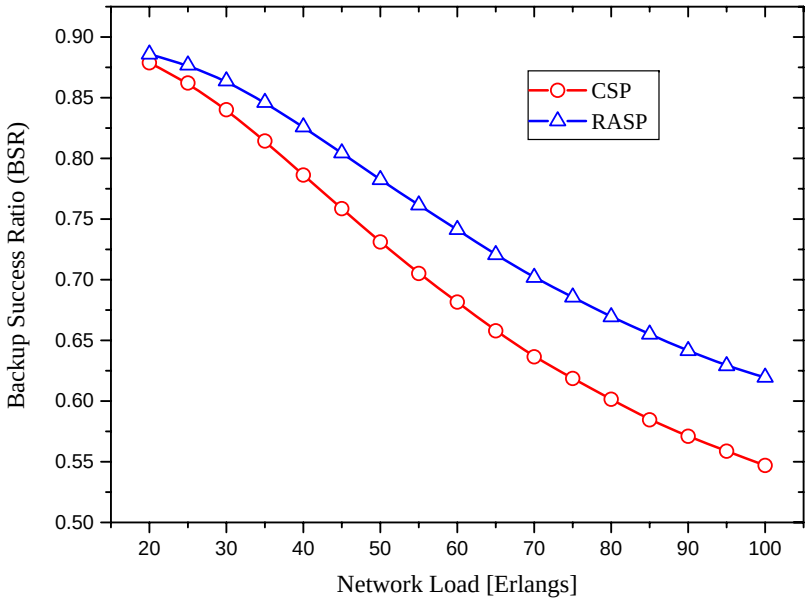

FIGURE 5.28: Backup Success Ratio versus Network Load ($0.9995 < a_r < 1$).

Observation 1. The BSR of both CSP and RASP decreased with an increase in network load. The reasons for this trend are the same as discussed previously under BPR (Observation 1). Compared with the performance in Experiment 3, results here show that both RASP and CSP exhibit worse BSR performance. This is due to the increase in the availability requirements which is expected, since higher constraints result in more connections requiring protection. Furthermore, higher QoS constraints result in a lower percentage of potential protected connections being able to meet the requirements, thus decreasing BSR.

Observation 2. RASP performed better than CSP with respect to BSR at all loads. The higher availability constraints produce this result for the time. In comparison with Experiment 3, the BSR performance of RASP was higher for all network loads. The result indicates that RASP's ability to use link disjoint protection paths proved effective enough to compensate for resource deficiencies. The result also highlights the inability of CSP to establish connections that require protection when the availability requirement constraints are high.

Observation 3. As the load increased, RASP displayed increasingly better performance as the graphs diverge. This observation is also unique to this experiment. The sharing of backup resources along with ability to provision partial link disjoint paths allow RASP to efficiently provision a PP when required. This result shows that an advantage of the RASP algorithm lies in its ability to provide superior BSR performance under high network loads and high availability requirement constraints.

5.5.5 Blocked Connections $(0.9995 < a_r < 1)$

Figures 5.29 to 5.33 show how connections have been blocked in Experiment 3. Figure 5.29 and Figure 5.30 give consider blocked connections for RASP and CSP respectively and Figures 5.31 to 5.33 look independently at the causes of blocking, comparing RSP and CSP in each case.

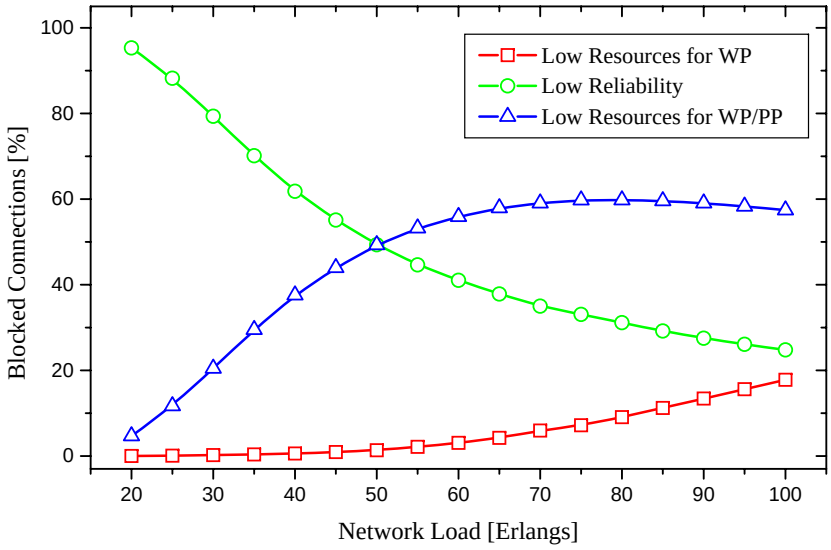

FIGURE 5.29: Blocked connections during RASP simulation ($0.9995 < a_r < 1$).

Observation 1: Low Resources for WP. With respect to Figure 5.29 and Figure 5.30, the percentage of connections blocked due to insufficient WP resources increased with increasing network load. There is an approximate 10% difference between CSP and RASP at higher network loads. In comparison with Experiment 3, RASP blocking performance with respect to low WP resources was marginally lower at all loads.

With respect to Figure 5.31, it is observed that more connections were blocked in the case of CSP. This result is repeated.

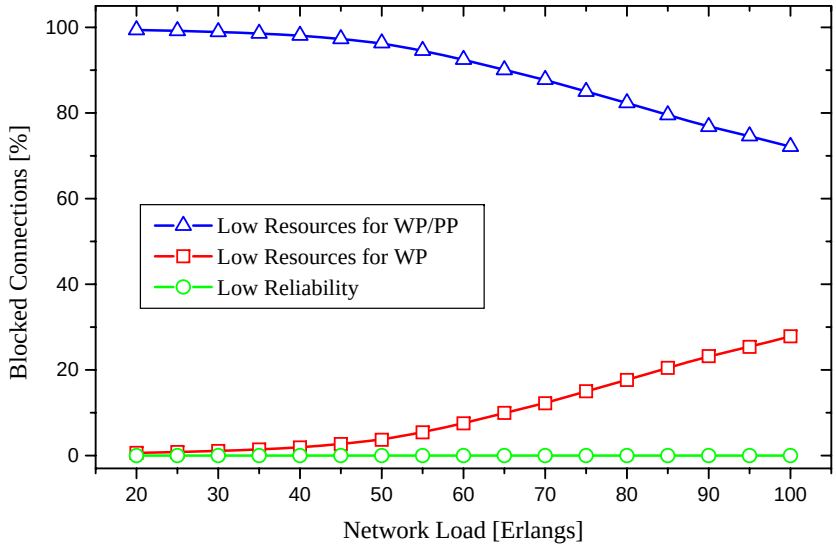

FIGURE 5.30: Blocked connections during CSP simulation ($0.9995 < a_r < 1$).

Observation 2: Low Resources for WP/PP. In the case of CSP, the percentage of blocked connections due to low resources to establish a WP/PP path pair, generally decreased with increasing network load. The result for RASP has a similar trend to that observed in Experiment 3. There was an increasing trend from 20 to 75 Erlangs and then a decreasing trend from 75 to 100 Erlangs.

For all network loads, RASP performed better that CSP with respect to blocking due to low WP/PP resources. Compared with the result in Experiment 3, the blocking has decreased. This proves that RASP was able to establish a higher percentage of WP/PP path pairs by utilizing bandwidth more efficiently. This can be attributed to the establishment of partial link disjoint PPs. This result justifies the superior BSR performance, that RASP exhibited.

Observation 3: Low Reliability. In this experiment with high availability requirements, it is observed in Figure 5.33, that more connections were blocked due to low reliability at low network loads. There was a decreasing trend for all loads from approximately 95% to 22% . A dramatic increase is observed in the percentage of blocked connections due to low reliability, when compared with Experiment 3, where the blocking varied from 45% to 3% . This indicates that the increase in availability requirements had a large impact. Since the blocking due to low resources has decreased, this confirms that the high availability requirements had a large impact on the blocking probability results.

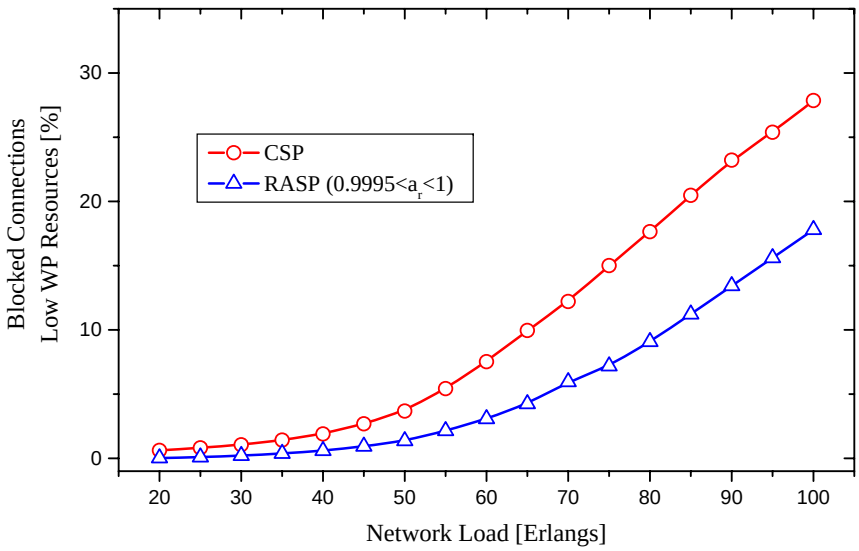

FIGURE 5.31: Percentage of connections blocked due to Low WP Resources ($0.9995 < a_r < 1$).

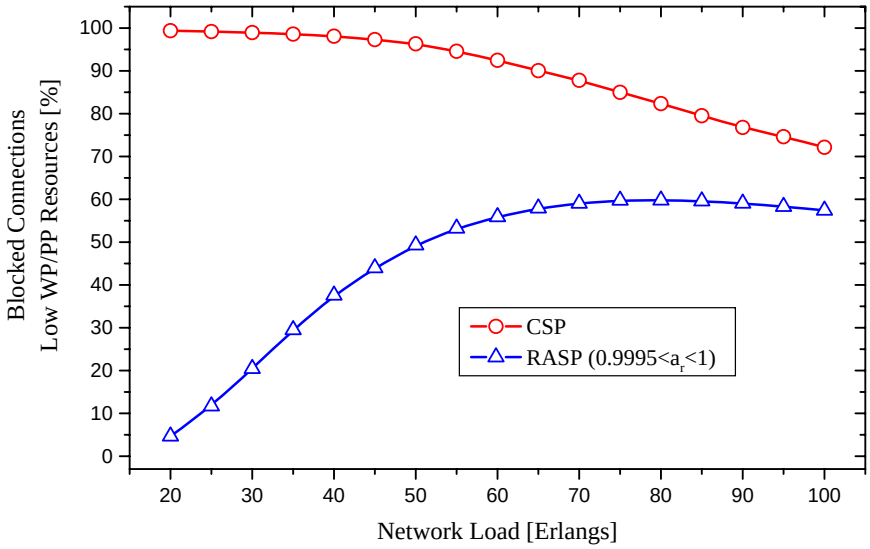

FIGURE 5.32: Percentage of connections blocked due to Low PP Resources ($0.9995 < a_r < 1$).

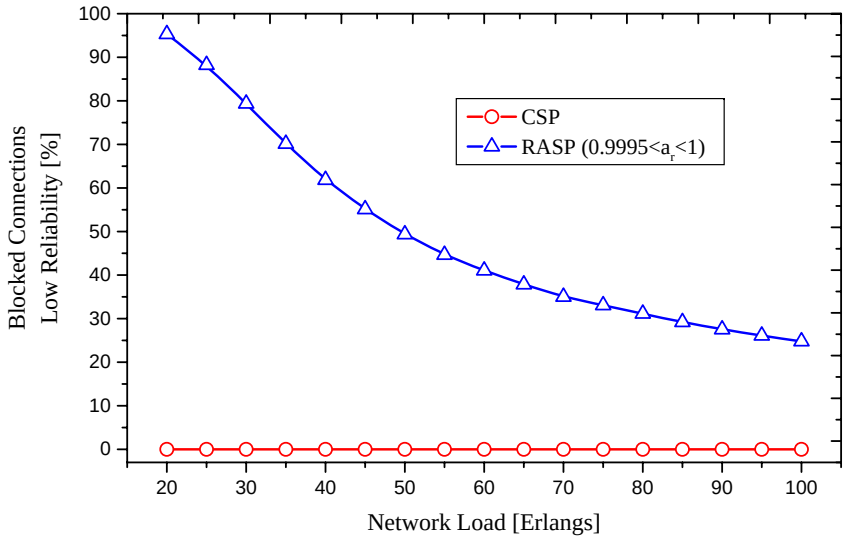

FIGURE 5.33: Percentage of connections blocked due to Low Reliability ($0.9995 < a_r < 1$).

CHAPTER SIX
CONCLUSION

6.1 INTRODUCTION

The problem addressed in this study was to investigate the performance of dynamic shared-path protection schemes under different network loads and in scenarios with different Quality of Service (QoS) requirements. This was done using a simulation approach that implemented two centralized, shared-backup, path protection schemes, referred to as Reliability Aware Shared-backup Protection (RASP) and Conventional Shared-backup Protection (CSP). CSP is distinguished by the lack of reliability awareness and, if resources are available, its ability to exclusively establish link disjoint path pairs. In contrast, RASP has the advantage of being reliability aware, allowing it to provide differentiated services by considering Service Level Agreement (SLA) requirements. Furthermore, with the use of a link disjoint parameter (ξ), RASP is able to include working path (WP) links in the search for a suitable protection path (PP), resulting in partial link disjoint connections [18]. This ability increases the probability of finding a suitable PP. CSP was investigated with the intention of comparing its performance with that of RASP. The comparison was used to highlight possible advantages and disadvantages and to observe the effect of RASP's distinguishing attributes on performance.

The study resulted in a simulation platform being designed using the $MATLAB^{®}$ [27] programming language. The simulator uses Poisson traffic and incorporates dynamic routing and wavelength assignment. Suitable link cost models along with Dijkstra's least cost algorithm were used to compute the working and backup routes. A physically dimensioned South African WDM topology was used as the sample network. The network consists of 19 nodes, that are situated at major metropolitan areas.

Each link is considered to be a bi-directional fibre with 8 wavelengths. The following assumptions were made in the analysis of connection availability [2, 6, 15]:

1. Links are the only components that may fail.
2. Only one link may be in a failed state at any time.
3. The link MTBF and MTTR are independent, memoryless processes.

Since connections are carried by a single path, the connection availability is equal to the path availability which may comprise a WP or WP/PP path pair. Availability was used as a QoS parameter [21] and as a measure of reliability [10].

The performance parameters were evaluated as a function of network load which was varied from 20 to 100 Erlangs with increments of 5 Erlangs. The four performance metrics evaluated were:

1. Blocking Probability (BP), which is used to determine the percentage of connection requests that were blocked,
2. Backup Primary Resource-Ratio (BPR), which is used to evaluate the ratio of the total protection bandwidth to primary path bandwidth,
3. Reliability Satisfaction Ratio (RSR) which is used to determine the percentage of established connections that complied with the availability requirements,
4. and Backup Success Ratio (BSR) which determines the ratio of connections that that were established with protection paths to the total number of connections that required protection.

An analysis of the constitution of blocked connections was used to provide further insight into the performance of RASP and CSP.

6.2 DISCUSSION OF RESULTS

6.2.1 Experiment 1

Experiment 1 considers a general comparison of performance between RASP and CSP. The availability requirement is uniformly distributed in a large range from 0.99 to 1 $[7, 15]$. The objective of this investigation is to obtain a general evaluation of the performances of both algorithms.

6.2.1.1 Blocking Probability

Three observations were highlighted (with reference to Figure 5.1):

1. Both schemes exhibited an increasing BP trend as the network load increased.
2. RASP exhibited a lower blocking probability than CSP, with the comparative difference in BP decreasing as the network load decreased.
3. At 20 Erlangs, CSP has a slightly lower BP than RASP.

The results suggest that both schemes are affected negatively by increasing network load. RASP made more efficient use of resources to reduce the number of blocked connections. The divergence, as network load increased, suggests that RASP is relatively more effective in minimizing the BP as the network load increases. Conversely, as the network load decreases, RASP performs poorly due to the severe impact of reliability requirement constraints on connection establishment.

6.2.1.2 Backup Primary Resource Ratio

Two observations were highlighted (with reference to Figure 5.2):

1. Both schemes displayed a decreasing trend with respect to BPR as the network load increased.
2. RASP presented a lower BPR than CSP under all load conditions.

The observations suggest that, as the network load increased, backup sharing caused the percentage of wavelength links used for protection to increase at a lower rate resulting in a

decreasing BPR. In comparison, however, RASP's superior BPR and BP performance indicate that it utilized less protection bandwidth to obtain a lower blocking probability. The lower BPR of RASP indicates optimized allocation and utilization of backup resources [2].

6.2.1.3 Reliability Satisfaction Ratio

Two observations were highlighted (with reference to Figure 5.3):

1. RASP exhibited an RSR of 1 throughout.
2. CSP's RSR performance exhibited a decreasing trend with increasing network load.

The RSR is evaluated to show that RASP is able to provision dependable connections only. This was confirmed. The observation of the RSR performance of CSP may appear meaningless since it does not consider reliability requirements. The results do however give an indication of what percentage of connections, provisioned by CSP, were dependable. The decreasing trend in the RSR performance of CSP is caused by CSP becoming constrained due to the lack of free resources. This results in longer path pairs (which have lower availabilities) being required in order to establish connections. This establishment of an increasing number of connections having a lower than required reliability, resulted in a decreasing RSR. It is also noteworthy from the results that, although CSP is not reliability aware, a high percentage of reliable connections were provisioned. The RSR performance of CSP is expected to get increasingly worse as the QoS guarantees get more demanding.

6.2.1.4 Backup Success Ratio

Three observations were highlighted (with reference to Figure 5.4):

1. The BSR of both CSP and RASP decreased with an increase in network load.
2. CSP outperformed RASP with respect to BSR at all loads.
3. As the load increased, CSP displayed increasingly better performance as the graphs diverge.

Observation 1 suggests that, as the network load increased, the schemes were negatively affected by both insufficient resources and QoS constraints. RASP, however, is affected by both QoS constraints as well as insufficient resource, while CSP is affected by insufficient resources only.

This explains Observation 2. Furthermore, RASP, which exhibited a lower BPR, consumed a larger percentage of bandwidth for WPs resulting in less protection bandwidth being available for sharing. The lower BSR indicates that the bandwidth utilization did not allow RASP to cater PPs for connections requiring backup. Under these QoS constraints, the ability of RASP to provision partial link disjoint connections proved ineffective.

6.2.1.5 Blocked Connections

Three observations were highlighted (with reference to Figures 5.5 and 5.6):

1. The percentage of connections blocked due to insufficient WP resources increases with increasing network load.
2. CSP is unaffected by low reliability.
3. The percentage of blocked connections due to insufficient WP/PP resources was the greatest contributor to blocking.

The great contribution of insufficient WP/PP resources to blocking suggests that the networks physical capacity was being heavily loaded. It is also evident that the higher blocking probability exhibited by CSP which was noticed earlier, resulted from the comparatively higher percentage of connections being blocked due to the lack of resources to provision a WP. RASP was not severely affected by reliability constraints, and blocking was only noticeable at low loads. This result was unexpected but confirms why the BP graphs converged at lower loads eventually resulting in CSP performing better at 20 Erlangs.

6.2.2 Experiments 2, 3 and 4

In Experiments 2,3 and 4, three availability requirement ranges were considered to simulate three classes of service. Within each class of service, the performance of both schemes was evaluated. In the high availability class, the availability requirement was uniformly distributed between 0.9995 and 1. In the medium class, the availability requirement was uniformly distributed between 0.995 and 0.99995. In the low class, the availability requirement was uniformly distributed between 0.99 and 0.9995.

6.2.2.1 Blocking Probability

A summary of the blocking performance of RASP and CSP is given in Figure 6.1. The figure shows only one graph for CSP, since it is not affected by availability requirement constraints.

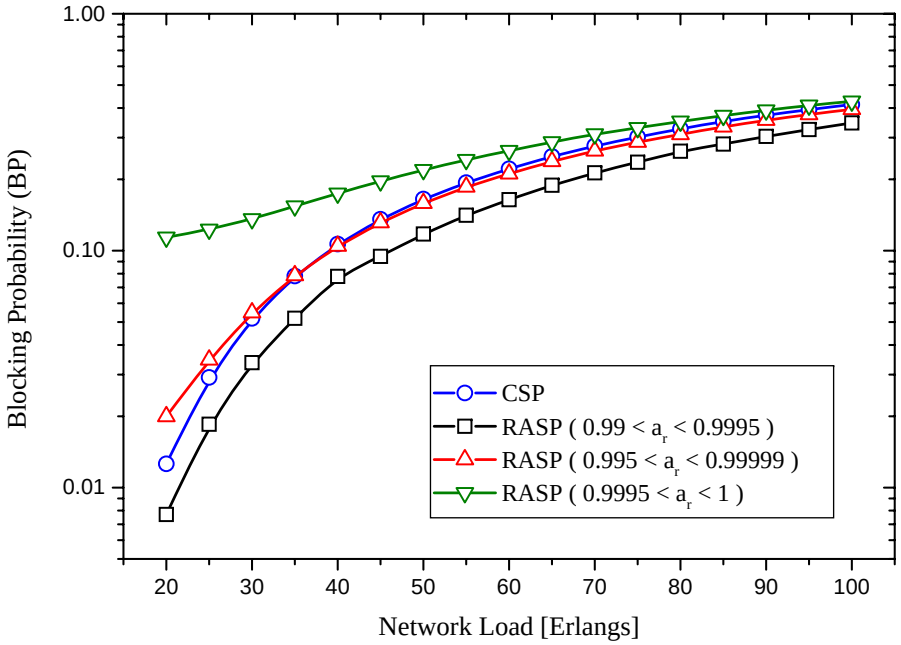

FIGURE 6.1: The effect of required availability constraints on Blocking Probability.

Is is observed that RASP was increasing affected by reliability awareness as the QoS constraints are increased. The results confirm expected behavior, with the BP performance of RASP being the highest when the average availability requirements were high and the lowest when the average availability requirements were low. An intermediate result was obtained when the average availability requirements were medium. Under high availability requirement constraints and at low network loads, the divergence of the RASP's BP performance from the trend was higher than expected. Under medium availability requirement constraints, the BP performance of RASP and CSP is similar with RASP performing marginally better as the network load increases. The result shows that CSP is only able to provide superior blocking performance when QoS constraints are high.

6.2.2.2 Backup Primary Resource Ratio

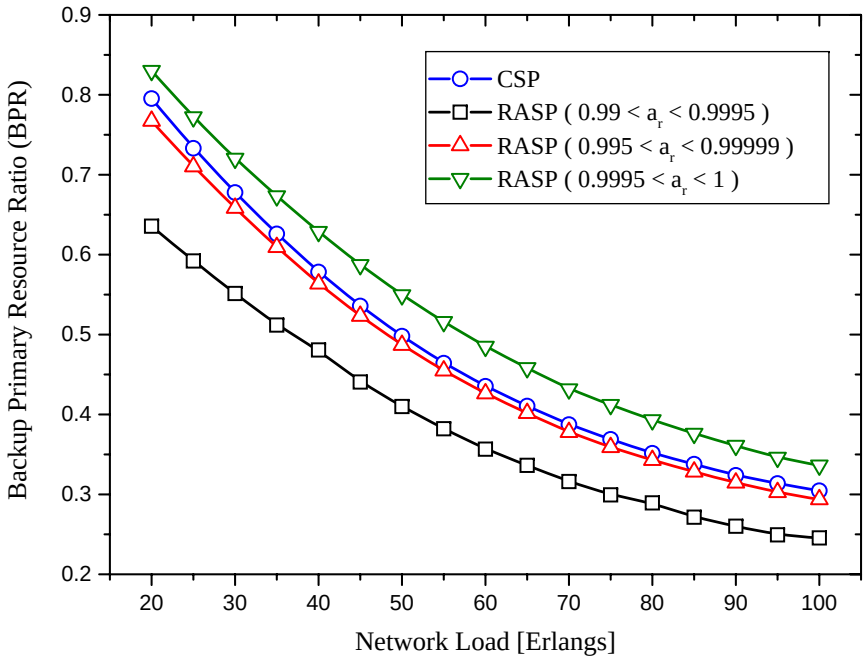

FIGURE 6.2: The effect of required availability constraints on Backup Primary Resource Ratio.

Figure 6.2 shows the BPR performance curves of RASP and CSP under the different classes of service. It is observed that the trends in BPR performance curves are similar with a small divergence at low network loads. The BPR of RASP was the highest when the average availability requirements were high and the lowest when the average availability requirements were low. An intermediate result was obtained when the average availability requirements were medium. Therefore, the BPR performance of RASP improved as the availability requirement constraints were decreased. The increasing availability requirement constraints resulted in RASP routing more protected connections in order for them to meet the required QoS. Is is assumed that the link disjoint PPs increased the percentage of bandwidth utilized for protection. This result confirms that CSP is only able to obtain superior BPR performance when the QoS constraints are high.

6.2.2.3 Reliability Satisfaction Ratio

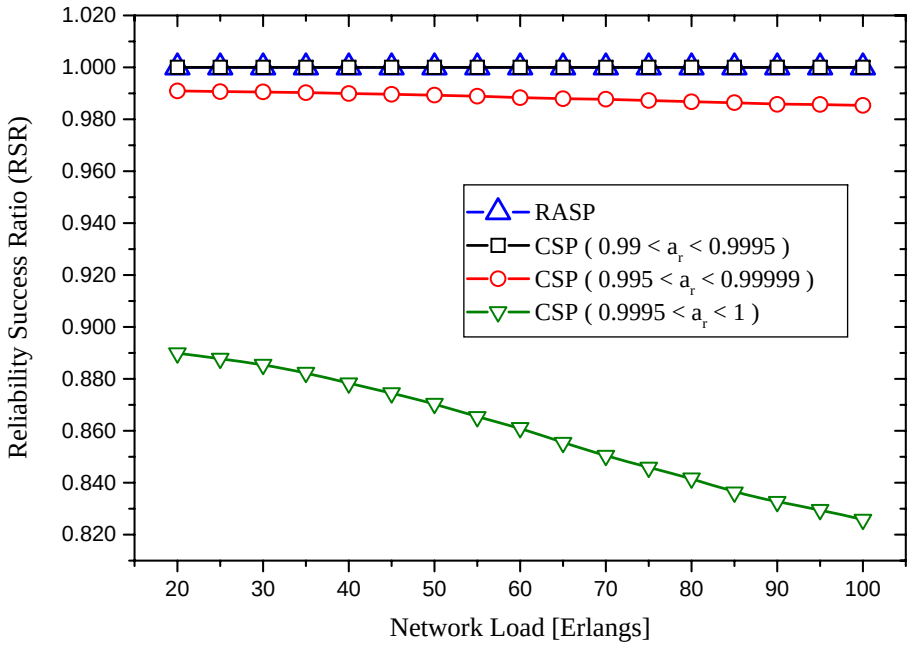

FIGURE 6.3: The effect of required availability constraints on Reliability Satisfaction Ratio.

The effect of required availability constraints on the Reliability Satisfaction Ratio, is described by Figure 6.3. Since RASP will only provision connections that are dependable, i.e. those that meet the availability requirements, it will always, by definition, obtain a RSR of 1. The tradeoff, however, as observed above, is that the BP and BPR performance of RASP is negatively affected by its reliability awareness. When considering RSR performance, it is CSP that is affected by the variation of availability requirements. When the average availability requirements were low, CSP was able to match the performance of RASP by obtaining an RSR of 1 at all network loads. When the average availability requirements were medium, the RSR performance of CSP decreased minimally, as the network load increased. When the average availability requirements were high, the RSR performance of CSP performance decreased to a larger extent and the decreasing trend in the curve was comparatively exaggerated. These results show that since CSP is unaware of required availability, that it will continue to establish connections regardless of whether they are dependable or not. Therefore, CSP will be unsuitable for networks with stringent QoS guarantees.

6.2.2.4 Backup Success Ratio

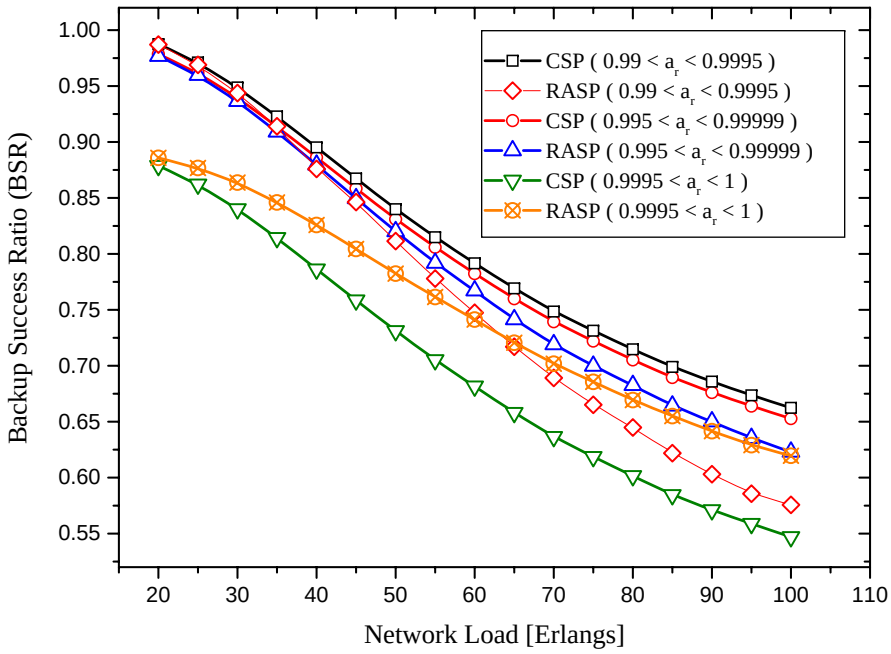

FIGURE 6.4: The effect of required availability constraints on Backup Success Ratio.

The BSR is the only performance parameter that affects both CSP and RASP as the QoS is varied. The BSR performance is shown in Figure 6.4.

With respect to RASP, as the QoS constraints were increased, the BSR behavior was not uniformly dynamic at all network loads. As the QoS service constraints increased from low to medium, the BSR from 40 to 100 Erlangs increased. This signifies that RASP was able to provision a higher percentage of protected connections. This cannot be attributed to more available network resources since the network load is increasing. It also cannot be attributed to lower QoS constraints, since the average availability requirement increased. This observation therefore suggests that RASP's unique ability to establish link disjoint protection paths had a positive impact on BSR, by allowing WP links to be included in the search for a suitable PP. Furthermore, if more PPs are established, more protection bandwidth becomes available to be shared. This further contributes positively to the increase of BSR at higher network loads. At network loads less than 40 Erlangs, the BSR did decrease marginally, which can be attributed to higher QoS constraints. As the QoS constraints increased from medium to high, the BSR decreased. The decrease can be attributed to the higher QoS constraints which severely decrease the ability of RASP to establish protected connections.

With respect to CSP, the BSR performance curve is as expected and is uniformly dynamic as the QoS constraints were increased. When the availability requirements were low, the BSR was high. A marginal decrease occurred when the QoS constraints increased from low to medium. A comparatively larger decrease was evident when the QoS constraints were increased from medium to high. Since CSP blocks connections based on the lack of available resources only, its blocking performance does not change with an increase in QoS requirements. As the QoS constraints increase, the number of connections that require protection increases, and by definition, this adversely affects the BSR performance of CSP.

6.2.2.5 Blocked Connections

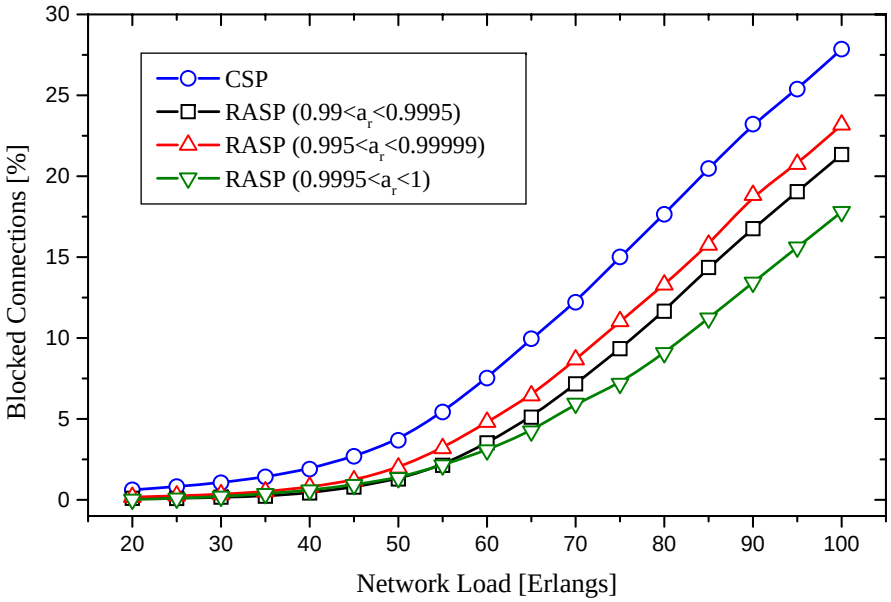

FIGURE 6.5: Percentage of connections blocked due to Low WP Resources.

Figure 6.5 shows the effect of QoS constraints on blocking due to the low WP resources. From the figure, it is observed that CSP blocked the highest percentage of connections. As the QoS constraints increased from low to medium, the percentage of blocked connections increased. This is due to the unavailability of resources as more connections are routed with protection paths. When the QoS constraints are increased from medium to high, the percentage of blocked connections due to low WP resources decreased dramatically. This is possibly due to the high percentage of connections being blocked due to inadequate reliability.

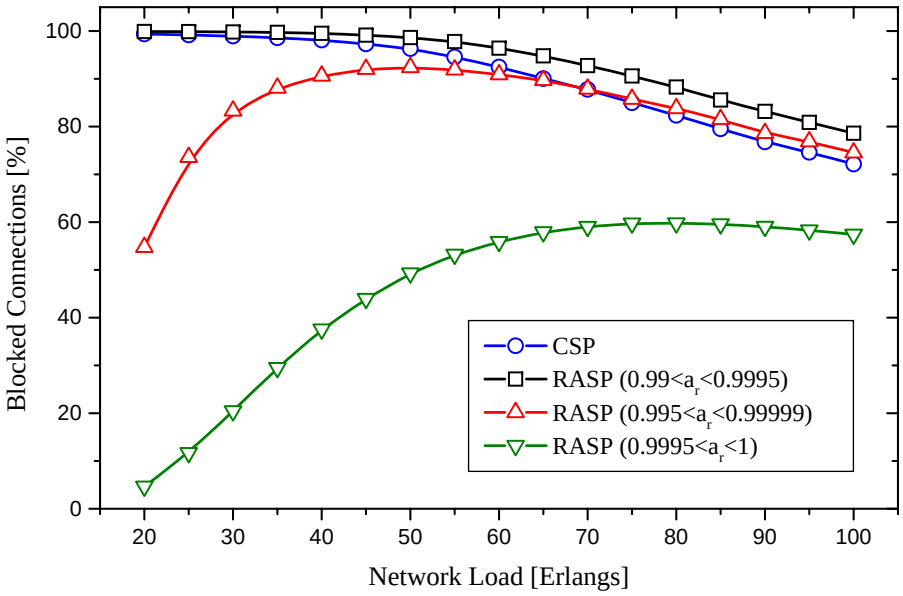

FIGURE 6.6: Percentage of connections blocked due to Low WP/PP Resources.

Figure 6.6 shows the effect of QoS constraints on blocking due to the low WP/PP resources. The percentage of connections blocked by CSP decreased gradually as the network load increased. This decrease is due to the sharp increase in blocked connections due to a lack of working path resources. With respect to RASP, as the QoS constraints increase, the blocking due to insufficient WP/PP resources decreased. This is due to the large percentage of connections that were blocked due to inadequate reliability. This decrease also provides a possible justification for why the BPR of RASP increased as the QoS constraints increased, since it indicates that progressively lower percentages of connections were blocked due to insufficient resources to establish a WP/PP pair.

Figure 6.7 shows the effect of QoS constraints on blocking due to the low reliability of connections. The results in this figure are as expected. CSP exhibited zero blocking, since it is not affected by low reliability. The increase in QoS constraint, resulted in RASP blocking more connections due to low reliability. When the availability requirement constraints were medium, a sharp increasing trend is observed at loads below 40 Erlangs. This suggests that the decrease in BSR at these network loads, may be attributed to reliability constraints.

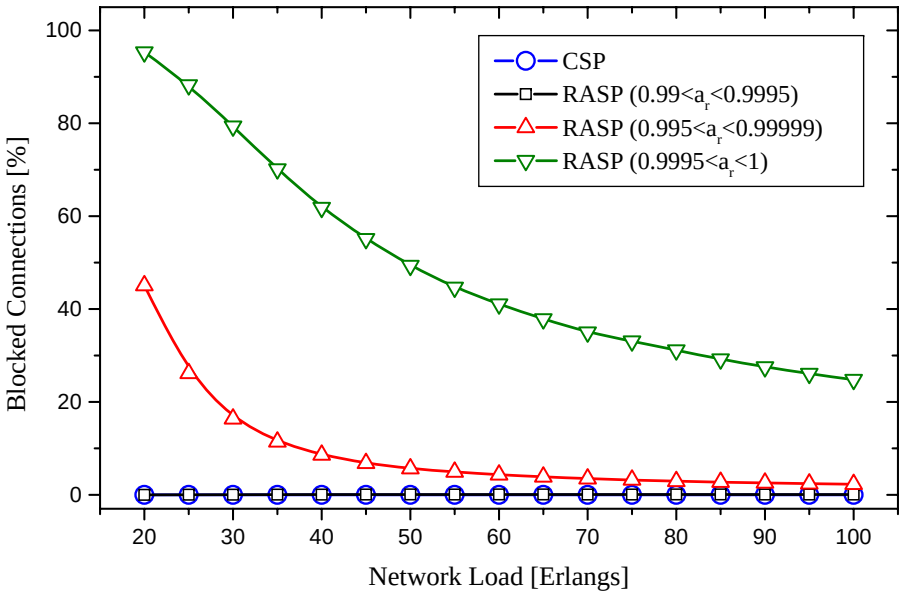

FIGURE 6.7: Percentage of connections blocked due to Low Reliability.

6.2.3 Summary

Increasing network load, which is related to an increased rate of arrival of connection requests, generally has a negative impact on performance. As the network load increases, more resources are utilized to provision connections. The BP, RSR and BSR performance, all indicated a decrease in performance due to insufficient resources when the network load was increased.

The BPR performance decreased with increasing network load. This was attributed to a higher blocking and the sharing of backup paths.

Increasing QoS constraints had a negative impact on the performance of RASP due to its availability awareness. This was clearly indicated by the results of the BP and BPR evaluations. The RSR performance of CSP also suffered due to increased QoS constraints. Due to reliability awareness, RASP will maintain a RSR of 1, indicating its ability to route dependable connections.

Under both medium and low QoS constraints, RASP was able to exhibit superior BP and BPR performance, when compared with CSP, indicating that it is able to use less protection bandwidth to establish more dependable connections. When QoS constraints were high, RASP performed relatively poorly and although it utilized more protection bandwidth, indicated by a higher BPR, it was unable to sufficiently reduce its BP in order to outperform CSP.

With respect to the BSR, RASP did indicate superior performance when the QoS constraints were high. This can be attributed to RASP being able to provision partial link disjoint WP/PP pairs by using WP links in the search for PPs, thus increasing the probability of finding one. This was confirmed by lower blocking due to insufficient WP/PP resources.

The increasing importance of QoS and differentiated services makes dynamic reliability aware protection schemes, such as RASP, desirable [3, 9, 40]. Due to reliability awareness, RASP outperforms CSP, under low and medium QoS constraints, with respect to BP, RSR and BPR performance metrics. Due to its ability to provision partial link disjoint protection paths, at high network loads RASP also outperforms CSP with respect to the BSR performance metric. It can be concluded from the simulation study that to simultaneously incorporate reliability awareness into a shared-path protection schemes like RASP and obtain high BP and BPR performance under demanding QoS constraints without increasing the network capacity, will be a challenging proposition. Improvements to RASP may involve more comprehensive cost modelling and more efficient routing and wavelength assignment. Waiting queues may also be incorporated to reduce the BP by possibly reducing the number of blocked connections.

6.3 RECOMMENDATIONS FOR FUTURE WORK

The outcomes of this study could be extended:

- by evaluating the percentage of dependable connections as a percentage of the total number of connection arrival requests. This would have given more significance to the availability awareness of RASP.
- with a stronger investigation of the contribution of ξ, by evaluating the percentage of protection paths that were partially link disjoint, to ascertain how often they were used.

In addition, there are a number of other interesting issues that are related to the present study and to simulation studies involving dynamic survivable WDM networks, which may be investigated in future studies. Some of these are listed below:

- Investigate waiting queues which would allow reliable connections, that could not be established immediately due to resource constraints, to join a queue for a specific time period, during which they can be established pending the availability of resources,

- Investigate the empirical definition of ξ and its effect on performance independently, i.e by allowing CSP to also be reliability aware and comparing the performance of two reliability aware schemes,
- Incorporate priority awareness so that RASP may be able to determine the availability of a connection based on its route as well as its priority,
- The use of different network topologies with a variation in capacity,
- The use of different dynamic routing algorithms with different routing cost models [40],
- Different approaches to evaluating connection availability [31], e.g. take into account the sharing of backup paths. [45],
- Considering dual or multiple failure scenarios and possible queuing methods to determine the priority of connections waiting in a backup queue [13, 36],
- Investigate the effect of recovery time on performance by considering restoration as an alternative to protection,
- Investigate and compare the performance of p-cycles (which is another type of protection scheme used for the dynamic establishment of recoverable connections) with the classical shared-path protection approaches used herein [49].

REFERENCES

[1] B. Xiang, H. Yu, S. Wang, and L. Li, "A differentiated shared protection algorithm supporting traffic grooming in WDM mesh networks," in *2004 IEEE International Conference on Communications, Circuits and Systems(ICCCAS 2004)*, 27-29 Jun. 2004, pp. 628–632.

[2] W. Fawaz, K. Chen, and G. Pujolle, "Priority-enabled optical shared protection: An online efficiency evaluation study," *Computer Communications*, vol. 30, no. 18, pp. 3690–3697, Dec. 2007.

[3] J. Zhang and B. Mukherjee, "A Review of Fault Management in WDM mesh networks: Basic Concepts and Research Challenges," *IEEE Network*, vol. 18, no. 2, pp. 41–48, Apr. 2004.

[4] R. Keralapura, C. N. Chuah, G. Iannaccone, and S. Bhattacharyya, "Service availability: a new approach to characterize IP backbone topologies," in *2004 Twelfth IEEE International Workshop on Quality of Service(IWQOS 2004)*, 7-9 June. 2004, pp. 232–241.

[5] L. Zhou and W. D. Grover, "A Theory for Setting the 'Safety Margin' on Availability Guarantees in an SLA," in *5th International Workshop on Design of Reliable Communication Networks (DRCN '05)*, 6-19 Oct. 2005, pp. 403–409.

[6] J. Zhang, K. Zhu, H. Zang, and B. Mukherjee, "A new provisioning framework to provide availability-guaranteed service in WDM mesh networks," in *Proc. IEEE International Conference on Communications*, 11-15 May. 2003, pp. 1484–1488.

[7] R. He, B. Lin, L. Li, and C. Gu, "Dynamic Shared Path Protection Algorithm in WDM Mesh Networks Under Service Level Agreements Contraints," in *Proc. 2005 IEEE Sixth Internation Conference on Parallel and Distributed Computing, Applications and Technologies*, 5-8 Dec. 2005, pp. 205–209.

[8] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM Mesh Networks," *Journal of Lightwave Technology*, vol. 21, no. 4, pp. 870–883, Apr. 2003.

[9] W. Fawaz, F. Martignon, K. Chen, and G. Pujolle, "A novel protection scheme for quality of service aware WDM networks," in *2005 IEEE International Conference on Communications (ICC 2005)*, vol. 3, May 2005, pp. 1720–1725.

[10] J. Vasseur, M. Pickavet, and P. Demeester, *Network Recovery*, 1st ed. San Fransisco, USA: Morgan Kaufmann Publishers, 2004.

[11] D. Elie-Dit-Cosaque, M. Ali, and L. Tancevski, "Informed dynamic shared path protection," in *Proc. Optical Fiber Communication Conference and Exhibit (OFC 2002)*, 17-22 Mar. 2002, pp. 492–493.

[12] G. Shen and W. D. Grover, "Survey and performance comparison of dynamic provisioning methods for optical shared backup path protection," in *2005 2nd International Conference on Broadband Networks*, vol. 2, 3-7 Oct. 2005, pp. 1310–1319.

[13] L. Guo, H. Yu, and L. Li, "Path-based protection in WDM mesh networks subject to double-link failures," *AEU - International Journal of Electronics and Communications*, vol. 60, no. 6, pp. 497–470, Jun. 2006.

[14] L. Guo, J. Cao, H. Yu, and L. Li, "Path-based routing provisioning with mixed shared protection in WDM mesh networks," *Journal of Lightwave Technology*, vol. 24, no. 3, pp. 1129–1141, Mar. 2006.

[15] R. He, L. Lin, and L. Li, "Dynamic service-level-agreement aware shared-path protection in WDM mesh networks," *Journal of Network and Computer Applications*, vol. 30, pp. 429–444, Apr. 2007.

[16] G. Xue, W. Zhang, J. Tang, and K. Thulasiraman, "Establishment of survivable connections in WDM networks using partial path protection," in *2005 IEEE International Conference on Communications(ICC 2005)*, vol. 3, 16-20 May. 2005, pp. 1756–1760.

[17] C. Saradhi and C. Murthy, "Dynamic establishment of differentiated survivable lightpaths in WDM mesh networks," *Computer Communications*, vol. 27, no. 3, pp. 273–294, Feb. 2004.

[18] R. Lin, S. Wang, L. Li, and L. Guo, "A New Network Availability Algorithm for WDM Optical Networks," in *Proc. 2005 IEEE Fifth International Conference on Computer and Information Technology*, 21-23 Sep. 2005, pp. 480–484.

[19] L. Guo, H. Yu, and L. Li, *Networking - ICN 2005*, ser. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, April 2005, vol. 3420/2005, ch. A New Path Protection Algorithm for Meshed Survivable Wavelength-Division-Multiplexing Networks, pp. 68–75.

[20] A. K. Somani, *Survivability and Traffic Grooming in WDM Optical Networks*, 1st ed. Cambridge, UK: Cambridge University Press, 2006.

[21] C. Saradhi, M. Gurusamy, and L. Zhou, "Differentiated QoS for survivable WDM optical networks," *IEEE Communications Magazine*, vol. 42, no. 5, pp. 8–14, May. 2004.

[22] S. Rao and C. Murthy, "Distributed dynamic QoS-aware routing in WDM optical networks," *Computer Networks*, vol. 48, no. 4, pp. 585–604, Jul. 2005.

[23] (2008, Jan.). [Online]. Available: http://www.neotel.co.za/neotel/view/neotel/en/page302

[24] Y. Huang, W. Wen, J. Zhang, J. P. Heritage, and B. Mukherjee, "A New Link-State Availability Model for Reliable Protection in Optical WDM Networks," in *2004 IEEE International Conference on Communcations*, 20-24 Jun. 2004, pp. 1649–1653.

[25] G. Conte, M. Listanti, M. Settembre, and R. Sabella, "Protection and Restoration Strategies in WDM Mesh Networks," in *Proceedings of ONDM 2002 Conference*, June 2002, pp. 1–15.

[26] C. Ou, J. Zhang, H. Zang, L. Sahasrabuddhe, and B. Mukherjee, "New and Improved Approaches for shared-Path Protection in WDM Mesh Networks," *Journal of Lightwave Technology*, vol. 22, no. 5, pp. 1223–1232, May. 2004.

[27] (2008, Jan.). [Online]. Available: http://www.mathworks.com/

[28] W. Grover, *Mesh-Based Survivable Networks Options and Strategies for Optical, MPLS, SONET and ATM Networking*, 1st ed. New Jersey, USA: Prentice Hall PTR, 2004.

[29] R. Ramaswami and K. N. Sivarajan, *Optical Networks A Practical Perspective*, 2nd ed. San Fransisco, USA: Morgan Kaufmann Publishers, 2002.

[30] T. E. Stern and K. Bala, *Multiwavelength Optical Networks*, 1st ed. New Jersey, USA: Prentice Hall PTR, 2000.

[31] D. Arci, G. Maier, A. Pattavina, G. Petecchi, and M. Tornatore, " Availability models for protection techniques in WDM networks," in *Proceedings. Fourth International Workshop on Design of Reliable Communication Networks (DRCN 2003)*, Oct 2003, pp. 158–166.

[32] N. Skorin-Kapov and M. Kos, "Static Routing and Wavelength Assignment in Wavelength Routed WDM Networks," in *Proceedings of IEEE MELECON 2006*, May 2006, pp. 16–19.

[33] B. Yao and B. Ramamurthy, "Survivable traffic grooming with path protection at the connection level in WDM mesh networks," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 2846–2853, Oct. 2005.

[34] R. Shenai and K. Sivalingam, "Hybrid survivability approaches for optical WDM mesh networks," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3046–3055, Oct. 2005.

[35] C. Saradhi, L. Zhou, M. Gurusamy, and C. Murthy, "Distributed network control for establishing reliability-constrained least-cost lightpaths in WDM mesh networks," in *Proc. Eighth IEEE International Symposium on Computers and Communication(ISCC 2003)*, vol. 1, 30 Jun. - 3 Jul. 2003, pp. 678–683.

[36] S. Thiagarajan, R. Ranganathan, L. Blair, and J. Berthold, "Economical evolution to high availability networks," in *Fourth International Workshop on Design of Reliable Communication Networks, 2003. (DRCN 2003)*, 19-22 Oct. 2003, pp. 311–316.

[37] R. He, H. Wen, L. Li, and G. Wang, "Shared Sub-Path Protection Algorithm in Traffic-Grooming WDM Mesh Networks," *Photonic Network Communications*, vol. 8, no. 3, pp. 239–249, Nov. 2004.

[38] P.-H. Ho and H. Mouftah, "Shared protection in mesh WDM networks," *IEEE Communications Magazine*, vol. 42, no. 1, pp. 70–76, Jan. 2004.

[39] M. Keshtgary, F. Al-Zahrani, and A. Jayasumana, "Network survivability performance evaluation with applications in WDM networks with wavelength conversion," in *2004 29th Annual IEEE International Conference on Local Computer Networks*, 16-18 Nov. 2004, pp. 344–351.

[40] H. Lou, H. Yu, and L. Li, "Static Routing and Wavelength Assignment in Wavelength Routed WDM Networks," in *Proceedings. 2005 International Conference on Communications, Circuits and Systems, 2005*, May 2005, pp. 580–584.

[41] D. Menasce, "QoS issues in Web services," *IEEE Internet Computing*, vol. 1, no. 6, pp. 72–75, Nov. 2002.

[42] J. E. Flood, Ed., *Telecommunication Networks*, 2nd ed. Stevenage, UK, UK: Institution of Electrical Engineers, 1996.

[43] M. Clouqueur and W. D. Grover, "Availability analysis of span-restorable mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 4, pp. 810–821, May. 2002.

[44] H. Cankaya, A. Lardies, and G. Ester, "Availability aware cost modeling of mesh architectures for long-haul networks," in *Proceedings. ISCC 2004. Ninth International Symposium on Computers and Communications*, vol. 2, 28 June-1 July 2004, pp. 766–771.

[45] S. Verbrugge, D. Colle, P. Demeester, R. Huelsermann, and M. Jaeger, "General availability model for multilayer transport networks," in *Proc. IEEE 5th International Workshop on Design of Reliable Communication Networks(DRCN 2005)*, 16-19 Oct. 2005, pp. 85–92.

[46] H. Zang, J. Jue, L. Sahasrabuddhe, R. Ramamurthy, and B. Mukherjee, "Dynamic lightpath establishment in wavelength routed WDM networks," *IEEE Communications Magazine*, vol. 39, no. 9, pp. 100–108, Sep. 2001.

[47] S. Maesschalck, D. Colle, I. Lievens, M. Pickavet, P. Demeester, C. Mauz, M. Jaeger, R. Inkret, B. Mikac, and J. Derkacz, " Pan-European Optical Transport Networks: An Availability-based Comparison," *Photonic Network Communications*, vol. 5, no. 3, pp. 203–225, May. 2003.

[48] M. Vaughn and R. Wagner, "Metropolitan network traffic demand study," *13th Annual Meeting. IEEE Lasers and Electro-Optics Society (LEOS 2000)*, vol. 1, pp. 102–103, 2000.

[49] R. Lu, T. Fangcheng, and L. Chang, "Dynamic establishment of restorable connections using p-cycle protection in WDM networks," *Optical Switching and Networking*, vol. 3, no. 3-4, pp. 191–201, Dec. 2006.

[50] R. G. Chamberlain. (1996, Oct.) GIS FAQ Q5.1: Great circle distance between 2 points. [Online]. Available: http://www.movable-type.co.uk/scripts/GIS-FAQ-5.1.html

LIST OF FIGURES

LIST OF TABLES

APPENDIX A
HAVERSINE FORMULA

The Haversine Formula [50] is used to calculate the airline distance (the great circle distance which ignores elevation differences) between two points on the Earth's surface. It requires the longitude and latitude of both points as well as the equatorial and polar radii of the Earth.

Inputs:
Lat1 - latitude of the first point.
Long1 - longitude of the first point.
Lat2 - latitude of the second point.

Long2 - latitude of the second point.
A - the Earth's equatorial radius, 6378 km.
B - the Earth's polar radius, 6357km.
Output:
D_{air} - the airline distance between the two points.

$$L(mean\_latitude) = \frac{(Lat1 + Lat2)}{2}. \tag{A.1}$$

$$dLat = |Lat1 - Lat2|. \tag{A.2}$$

$$dLong = |Long - Long2|. \tag{A.3}$$

$$E(eccentricity\_of\_ellipsiod) = \sqrt{\left(1 - \left(\frac{B^2}{A^2}\right)\right)}. \tag{A.4}$$

$$R1 = \frac{A \times (1 - E^2)}{\left(1 - E^2 \sin^2 L\right)^{3/2}}. \tag{A.5}$$

$$R2 = \frac{A}{\sqrt{1 - E^2 \sin^2 L}}. \tag{A.6}$$

$$R(\text{Earth's mean radius}) = R1\left(\frac{dLat}{dLat + dLong}\right) + R2\left(\frac{dLong}{dLat + dLong}\right). \tag{A.7}$$

$$a = \sin^2\left(\frac{dLat}{2}\right) + \cos(L1)\cos(L2)\sin^2\left(\frac{dLong}{2}\right). \tag{A.8}$$

if a < 1, (A.9)

(A.10)

$$c = 2\sin^{-1}\sqrt{a}, \tag{A.11}$$

else, (A.12)

(A.13)

$$c = 2\sin^{-1} 1. \tag{A.14}$$

$$D_{air} = R \times c.$$

TABLE A.1: Node Latitudes and Longitudes

Node	Node Full Name	Latitude	Longitude
CT	Cape Town	33.56 S	18.29E
MB	Mossel Bay	34.08 S	22.08 E
PE	Port Elizabeth	33.58S	25.40E
EL	East London	33.00S	27.55E
Mta	Mthatha	31.36S	28.49E
Dbn	Durban	29.53 S	30.53 E
Emi	Empangeni	28.50S	31.52E
Upt	Upington	28.25S	21.15E
Kby	Kimberly	28.43S	24.46E
Bfn	Bloemfontein	29.06S	26.07E
Mkg	Mafikeng	25.50S	25.38E
Kdp	Klerksdorp	26.53S	26.38E
Jhb	Johannesburg	26.11 S	28.3 E
Hbt	Hartebeespoort	28.45S	20.32E
Pta	Pretoria	25.45 S	28.14 E
Nst	Nelspruit	25.29S	30.59E
Pke	Polokwane	23.53S	29.26E
Hst	Hoedspruit	24.22S	31.02E
Msa	Musina	22.20S	30.05E

TABLE A.2: Intercity Airline Distance (km)

	CT	Dbn	EL	Emi	Jhb	Kby	Kdp	Mfg	MB	Mta	Msa	Nst	Pke	PE	Pta	Upt
Bfn	884.3	408.8	319.9	X	474.4	263.9	383.6	X	613.9	235.6	X	X	X	421.5	X	X
CT	X	X	X	X	X	833.5	X	X	343.4	X	X	X	X	X	X	664.5
Dbn	X	X	X	149.2	497.3	X	X	X	X	287.5	X	X	X	X	X	X
EL	X	X	X	X	X	X	X	X	X	180.5	X	X	X	239.9	X	X
Emi	X	X	X	X	471.1	X	X	X	X	X	X	376.3	X	X	X	X
HES	X	X	X	X	56.9	X	172.5	220.6	X	X	X	X	259.9	X	31.8	X
Hdt	X	X	X	X	X	X	X	X	X	X	241.2	124.1	161.8	X	X	X
Jhb	X	X	X	X	X	X	159.1	X	X	X	X	X	X	X	50.7	X
Kby	X	X	X	X	X	X	279.1	330.2	X	X	X	X	X	X	X	344.6
Kdp	X	X	X	X	X	X	X	150.3	X	X	X	X	X	X	X	X
Mfg	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	521.2
MB	X	X	X	X	X	X	X	X	X	X	X	X	X	318.5	X	X
Msa	X	X	X	X	X	X	X	X	X	X	X	X	182.8	X	X	X
Nst	X	X	X	X	299.8	X	X	X	X	X	X	X	231.9	X	282.4	X
Pke	X	X	X	X	X	X	X	X	X	X	X	X	X	X	242.9	X

TABLE A.3: Intercity Fibre Distance (km)

	CT	Dbn	EL	Emi	Jhb	Kby	Kdp	Mfg	MB	Mta	Msa	Nst	Pke	PE	Pta	Upt
Bfn	1326.4	613.2	479.7	X	711.7	395.8	575.4	X	920.8	353.4	X	X	X	632.2	X	X
CT	X	X	X	X	X	1250.2	X	X	515.1	X	X	X	X	X	X	996.8
Dbn	X	X	X	223.8	746	X	X	X	X	431.2	X	X	X	X	X	X
EL	X	X	X	X	X	X	X	X	X	270.8	X	X	X	359.9	X	X
Emi	X	X	X	X	706.6	X	X	X	X	X	X	564.5	X	X	X	X
HES	X	X	X	X	85.3	X	258.8	330.9	X	X	X	X	389.8	X	47.7	X
Hdt	X	X	X	X	X	X	X	X	X	X	361.7	186.2	242.7	X	X	X
Jhb	X	X	X	X	X	X	238.7	X	X	X	X	X	X	X	76.1	X
Kby	X	X	X	X	X	X	418.6	495.3	X	X	X	X	X	X	X	516.9
Kdp	X	X	X	X	X	X	X	225.5	X	X	X	X	X	X	X	X
Mfg	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	781.8
MB	X	X	X	X	X	X	X	X	X	X	X	X	X	477.7	X	X
Msa	X	X	X	X	X	X	X	X	X	X	X	X	274.3	X	X	X
Nst	X	X	X	X	449.6	X	X	X	X	X	X	X	347.9	X	423.5	X
Pke	X	X	X	X	X	X	X	X	X	X	X	X	X	X	364.4	X

APPENDIX B
TABLES OF RESULTS

B.1 EXPERIMENT 1

TABLE B.1: RASP Performance with $0.99 < a_r < 1$

Load	BP	BPR	RSR	BSR
20	0.01311	0.6441	1	0.97867
25	0.02382	0.59957	1	0.96138
30	0.03863	0.55873	1	0.93763
35	0.05735	0.51835	1	0.90797
40	0.07786	0.48072	1	0.87594
45	0.09961	0.44667	1	0.84301
50	0.12231	0.41585	1	0.8097
55	0.14685	0.387	1	0.77558
60	0.16961	0.36334	1	0.74623
65	0.19434	0.34087	1	0.7168
70	0.2179	0.32044	1	0.68989
75	0.24049	0.30424	1	0.66688
80	0.263	0.28939	1	0.64482
85	0.28613	0.27688	1	0.6238
90	0.30883	0.26383	1	0.60455
95	0.32936	0.25325	1	0.58885
100	0.34601	0.24553	1	0.57569

TABLE B.2: CSP Performance with $0.99 < a_r < 1$

Load	BP	BPR	RSR	BSR
20	0.01261	0.7953	0.99456	0.9821
25	0.02918	0.73308	0.9945	0.9657
30	0.0518	0.67777	0.99428	0.94329
35	0.07819	0.62597	0.99423	0.91751
40	0.10665	0.57829	0.99405	0.88984
45	0.13579	0.53575	0.99378	0.86198
50	0.16519	0.49805	0.99361	0.83455
55	0.19371	0.46419	0.99336	0.80945
60	0.22167	0.43534	0.9931	0.78609
65	0.25002	0.41054	0.99294	0.76371
70	0.27686	0.38736	0.99255	0.74287
75	0.3018	0.36883	0.99234	0.72574
80	0.32654	0.35143	0.99207	0.709
85	0.35134	0.33783	0.99186	0.69327
90	0.37373	0.32381	0.99177	0.68014
95	0.39383	0.31363	0.99156	0.66784
100	0.41404	0.30456	0.99141	0.65668

TABLE B.3: CSP's Percentage blocked connections with $0.99 < a_r < 1$

	CSP		
Load	Low Reliability	Low WP Resources	Low PP Resources
20	0	0.6188	99.3812
25	0	0.81549	99.18451
30	0	1.05395	98.94605
35	0	1.41834	98.58166
40	0	1.89599	98.10401
45	0	2.68801	97.31199
50	0	3.67931	96.32069
55	0	5.42779	94.57221
60	0	7.52667	92.47333
65	0	9.95484	90.04516
70	0	12.20815	87.79185
75	0	15.00848	84.99152
80	0	17.64798	82.35202
85	0	20.47647	79.52353
90	0	23.21682	76.78318
95	0	25.38525	74.61475
100	0	27.85583	72.14417

TABLE B.4: RASP's Percentage blocked connections with $0.99 < a_r < 1$

	RASP		
Load	Low Reliability	Low WP Resources	Low PP Resources
20	8.37501	0.04653	91.57846
25	3.80171	0.05674	96.14155
30	2.12274	0.15307	97.72419
35	1.47307	0.23614	98.29079
40	1.0767	0.48002	98.44328
45	0.85702	0.74261	98.40037
50	0.73939	1.27265	97.98797
55	0.62323	2.27403	97.10274
60	0.51716	3.58849	95.89435
65	0.45613	5.31329	94.23059
70	0.43113	7.2584	92.31047
75	0.41778	9.29056	90.29165
80	0.37875	11.7324	87.88885
85	0.38981	14.37964	85.23055
90	0.32094	16.76713	82.91193
95	0.32029	19.27982	80.39989
100	0.3088	21.85739	77.83381

B.2 EXPERIMENT 2, 3 AND 4

B.2.1 Blocking Probability (BP)

TABLE B.5: BP Performance with varying Availability Requirement

Load	CSP	RASP $0.99 < a_r < 0.9995$	RASP $0.995 < a_r < 0.99999$	RASP $0.9995 < a_r < 1$
20	0.01261	0.007697	0.01997	0.11385
25	0.02918	0.018506	0.03462	0.12296
30	0.0518	0.033692	0.05476	0.13612
35	0.07819	0.051842	0.0786	0.15406
40	0.10665	0.077856	0.10468	0.17458
45	0.13579	0.094602	0.13148	0.19651
50	0.16519	0.117573	0.15864	0.21941
55	0.19371	0.141117	0.18596	0.2419
60	0.22167	0.164162	0.21117	0.26414
65	0.25002	0.188960	0.23837	0.28758
70	0.27686	0.213168	0.26401	0.3105
75	0.3018	0.236196	0.28741	0.32997
80	0.32654	0.263005	0.30961	0.35135
85	0.35134	0.281763	0.33347	0.3716
90	0.37373	0.303758	0.35579	0.39158
95	0.39383	0.324521	0.37586	0.41063
100	0.41404	0.346008	0.39508	0.42689

B.2.2 Backup Primary Resource Ratio (BPR)

TABLE B.6: BPR Performance with varying Availability Requirement

Load	CSP	RASP $0.99 < a_r < 0.9995$	RASP $0.995 < a_r < 0.99999$	RASP $0.9995 < a_r < 1$
20	0.7953	0.635573	0.76721	0.83013
25	0.73308	0.592023	0.71017	0.77208
30	0.67777	0.551395	0.65827	0.72027
35	0.62597	0.511913	0.6093	0.67328
40	0.57829	0.480724	0.56359	0.62862
45	0.53575	0.440594	0.52316	0.58718
50	0.49805	0.409977	0.48678	0.54943
55	0.46419	0.382135	0.4548	0.51593
60	0.43534	0.356617	0.42614	0.48491
65	0.41054	0.336309	0.40159	0.45816
70	0.38736	0.316106	0.37781	0.43167
75	0.36883	0.299372	0.35896	0.41212
80	0.35143	0.289385	0.34284	0.39304
85	0.33783	0.271502	0.32825	0.376
90	0.32381	0.260176	0.31447	0.36099
95	0.31363	0.249328	0.30262	0.34603
100	0.30456	0.245533	0.29364	0.33631

B.2.3 Reliability Success Ratio (RSR)

TABLE B.7: RSR Performance with varying Availability Requirement

Load	RASP	CSP $0.99 < a_r < 0.9995$	CSP $0.995 < a_r < 0.99999$	CSP $0.9995 < a_r < 1$
20	1	1	0.990943	0.889900
25	1	1	0.990645	0.887826
30	1	1	0.990547	0.885493
35	1	1	0.990249	0.882400
40	1	1	0.989937	0.878315
45	1	1	0.989617	0.874516
50	1	1	0.989307	0.870456
55	1	1	0.988909	0.865350
60	1	1	0.988326	0.861120
65	1	1	0.987919	0.855399
70	1	1	0.987756	0.850426
75	1	1	0.987229	0.845852
80	1	1	0.986750	0.841700
85	1	1	0.986408	0.836295
90	1	1	0.985742	0.832654
95	1	1	0.985721	0.829560
100	1	1	0.985372	0.825792

B.2.4 Backup Success Ratio (BSR)

TABLE B.8: CSP BSR Performance with varying Availability Requirement

Load	$0.99 < a_r < 0.9995$	$0.995 < a_r < 0.99999$	$0.9995 < a_r < 1$
20	0.98747	0.97854	0.878752
25	0.97105	0.96197	0.862120
30	0.94871	0.93976	0.840079
35	0.92283	0.91385	0.814310
40	0.89517	0.88618	0.786238
45	0.86738	0.8584	0.758539
50	0.83991	0.83095	0.731109
55	0.81487	0.80585	0.705144
60	0.79155	0.78233	0.681617
65	0.76913	0.75987	0.657915
70	0.74845	0.73932	0.636502
75	0.73134	0.72203	0.618605
80	0.71466	0.70522	0.601530
85	0.69896	0.68949	0.584538
90	0.68579	0.67604	0.571023
95	0.67352	0.66394	0.558726
100	0.66238	0.65273	0.546984

TABLE B.9: RASP BSR Performance with varying Availability requirement

Load	$0.99 < a_r < 0.9995$	$0.995 < a_r < 0.99999$	$0.9995 < a_r < 1$
20	0.987066	0.976737	0.88582
25	0.968980	0.959758	0.87679
30	0.943790	0.936463	0.86374
35	0.914041	0.909086	0.84598
40	0.875941	0.879496	0.82579
45	0.846276	0.849507	0.80442
50	0.811391	0.820048	0.78235
55	0.777907	0.791820	0.76147
60	0.747392	0.767123	0.74126
65	0.716978	0.741329	0.72056
70	0.689141	0.719052	0.70182
75	0.665112	0.699826	0.6856
80	0.644820	0.682559	0.66934
85	0.621890	0.664165	0.65501
90	0.603037	0.650076	0.6415
95	0.585646	0.635309	0.62903
100	0.575693	0.623018	0.61952

TABLE B.10: Percentage of connections blocked due to Low WP Resources

Load	CSP	RASP $0.99 < a_r < 0.9995$	RASP $0.995 < a_r < 0.99999$	RASP $0.9995 < a_r < 1$
20	0.6188	0.09094	0.17029	0.01845
25	0.81549	0.08105	0.25421	0.0854
30	1.05395	0.16621	0.33236	0.20497
35	1.41834	0.22183	0.50508	0.36998
40	1.89599	0.42002	0.7967	0.5923
45	2.68801	0.78857	1.19103	0.92567
50	3.67931	1.28686	1.9819	1.35454
55	5.42779	2.13086	3.18071	2.14139
60	7.52667	3.51908	4.80412	3.06844
65	9.95484	5.10743	6.44219	4.23634
70	12.20815	7.16477	8.66659	5.95131
75	15.00848	9.33843	11.02896	7.18013
80	17.64798	11.65381	13.28915	9.07714
85	20.47647	14.35746	15.7518	11.21654
90	23.21682	16.75643	18.82817	13.42988
95	25.38525	19.04499	20.74666	15.60534
100	27.85583	21.34013	23.17961	17.79651

TABLE B.11: Percentage of connections blocked due to Low WP/PP Resources

Load	CSP	RASP $0.99 < a_r < 0.9995$	RASP $0.995 < a_r < 0.99999$	RASP $0.9995 < a_r < 1$
20	99.3812	99.90906	54.75809	4.67026
25	99.18451	99.89193	73.5939	11.69361
30	98.94605	99.80411	83.30168	20.45005
35	98.58166	99.72995	88.08427	29.50222
40	98.10401	99.53088	90.6173	37.57697
45	97.31199	99.14695	92.02102	43.9727
50	96.32069	98.64169	92.38445	49.3061
55	94.57221	97.7997	91.9017	53.22263
60	92.47333	96.41086	90.89798	55.89363
65	90.04516	94.81054	89.6898	57.92658
70	87.79185	92.7719	87.87754	59.04645
75	84.99152	90.59595	85.76305	59.75404
80	82.35202	88.27581	83.80748	59.78398
85	79.52353	85.57156	81.51934	59.58106
90	76.78318	83.18497	78.61214	59.05379
95	74.61475	80.88629	76.86843	58.30117
100	72.14417	78.60555	74.53731	57.42035

B.2.5 Percentage Blocked Connections

TABLE B.12: Percentage of connections blocked due to Low Reliability

Load	CSP	RASP $0.99 < a_r < 0.9995$	RASP $0.995 < a_r < 0.99999$	RASP $0.9995 < a_r < 1$
20	0	0	45.07162	95.31129
25	0	0.02702	26.15189	88.22099
30	0	0.02968	16.36596	79.34498
35	0	0.04822	11.41065	70.1278
40	0	0.04909	8.586	61.83073
45	0	0.06448	6.78795	55.10163
50	0	0.07144	5.63365	49.33937
55	0	0.06945	4.91759	44.63598
60	0	0.07005	4.2979	41.03793
65	0	0.08203	3.868	37.83708
70	0	0.06333	3.45588	35.00224
75	0	0.06562	3.20799	33.06583
80	0	0.07038	2.90337	31.13888
85	0	0.07098	2.72886	29.2024
90	0	0.0586	2.55968	27.51633
95	0	0.06872	2.38491	26.09349
100	0	0.05432	2.28308	24.78314