NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

The Effect of Optical Cross Connect Switch Time in Interdomain All-Optical Networks

by

Mark Joseph Francisco
B.Eng., Carleton University, 2001

A thesis submitted to
The Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering

Carleton University
Ottawa, Ontario,
January 16, 2004
© Copyright
2004, Mark Joseph Francisco

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-89836-9

Our file Notre référence

ISBN: 0-612-89836-9

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this dissertation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de ce manuscrit.

While these forms may be included in the document page count, their removal does not represent any loss of content from the dissertation.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

This thesis details the effect of the Optical Cross Connect (OXC) Switch Time upon the establishment of end-to-end lightpaths in all-optical networks. The OXC Switch Times introduces queueing delays to the signalling system, which has not been studied prior to this investigation. The effect of the OXC Switch Time is evaluated with different experimental results under a competitive network. By doing so, it is shown that under various traffic loads, the OXC Switch Time has a large impact on the lightpath setup time, making it take longer for lightpaths to be established.

Experiments show that when the OXC Switch Time is coupled with the Two-Phase Conservative (2PC) signalling algorithm and the Four-Phase Conservative (4PC) signalling algorithm, the overall blocking probability is left unchanged, but the 4PC scheme has an advantage in that the end-to-end establishment of a lightpath requires less time than that of the 2PC scenario. This is a distinct benefit of the 4PC method, unique with the introduction of the OXC Switch Time. Under the scenario studied, the Four-Phase Aggressive (4PA) signalling algorithm underperforms compared to the conservative methods with respect to blocking probability.

These results are tested in the OBGp Protocol which has been described in previous research. OBGp is a new method of managing and controlling resources across the interdomain and is a proposed solution to the interdomain control protocol problem. OBGp is an extension of BGP, making it a lightweight solution that is easily integrated into an existing carrier's network to provide BGP functionality and methods of controlling and managing the underlying all-optical network.

Acknowledgements

I would like to extend my appreciation and gratitude to my two supervising professors Dr. Ioannis Lambadaris and Dr. Changcheng Huang for their constant guidance, encouragement, motivation and support in my research.

This research could never have been completed without the help of my OBGp brothers, Stephen Simpson and Lambros Pezoulas. Thank you, for being part of this grand project and assisting in the realization of OBGp.

My gratitude goes to my friends and colleagues in the Optical Networks Laboratory and the Broadband Networks Laboratory, especially to Ashraf Matrawy, Nisrine Slaymane, Yan Robichaud and Yunhao Li. Their input has been most helpful and I thank them for it.

Special thanks go to my parents, Maximiano and Lilia, my brother and sister-in-law, Max and Hypatia, and all my friends for their patience and understanding through the course of my studies.

Finally, I would like to acknowledge the financial aid of Carleton University, The Natural Science and Engineering Research Council of Canada (NSERC), Communications and Information Technology of Ontario (CITO), Mathematics of Information Technology and Complex Systems (MITACS), Canarie Inc., and the Kaye Family towards the completion of this degree.

Table of Contents

Abstract.....	ii
Acknowledgements	iii
Table of Contents	iv
Table of Figures.....	vii
List of Acronyms	ix
1 Introduction.....	1
1.1 Background.....	1
1.2 Thesis Objectives	3
1.3 Thesis Contribution.....	4
1.4 Publications.....	4
1.4.1 Refereed Conferences	4
1.4.2 Poster Presentations	5
1.5 Thesis Organization	5
2 All-Optical Networks	7
2.1 Introduction.....	7
2.2 Routing and Wavelength Assignment in All-Optical Networks.....	7
2.3 Signalling in All-Optical Networks	10
2.3.1 Centralized Signalling.....	10
2.3.2 Distributed Signalling	11
2.4 All-Optical Network Control Plane Architectures.....	17
2.5 Border Gateway Protocol and Virtual Private Networks.....	21

3	OXC Switch Time Problem and Interdomain Control Protocols in All-Optical Networks	23
3.1	Introduction.....	23
3.2	OXC Switch Time Problem.....	23
3.3	Interdomain Control Protocol for All-Optical Networks.....	28
4	Optical Border Gateway Protocol.....	31
4.1	Introduction.....	31
4.2	OBGP Architecture.....	31
4.2.1	The OBGP Message.....	34
4.2.2	The OBGP Wavelength Table	36
4.3	OBGP Signalling Algorithm.....	38
4.3.1	Conservative Reservation	38
4.3.2	Aggressive Reservation	43
4.3.3	Teardown and Error	45
4.4	The OBGP Router.....	47
5	Simulation Results and Performance Evaluation	49
5.1	Introduction.....	49
5.2	Simulation Model and Scenario.....	49
5.3	Simulation Results	51
5.3.1	Conservative-First-Fit Wavelength Selection Algorithm	52
5.3.2	Conservative-Random Wavelength Selection Algorithm.....	58
5.3.3	Aggressive Scenario Results.....	69
5.3.4	Varying the Number of Wavelengths per Link.....	70

6	Conclusions and Recommendations for Further Study	71
6.1	Conclusion	71
6.1.1	Interdomain Control Protocol in All-Optical Networks	71
6.1.2	OBGP Signalling and OXC Switch Times	72
6.2	Recommendations for Further Study	74
7	References.....	78
A	Appendix A.....	A-1
A.1	Background.....	A-1
A.2	BGP Message Types.....	A-2
B	Appendix B	B-1
B.1	OPNET Simulation Model.....	B-1
B.1.1	BGP Process Models.....	B-1
B.1.2	OBGP Request Generator	B-3
B.1.3	OBGP Message Handler	B-4
B.2	OBGP Configuration	B-5
B.2.1	Wavelength Configuration Files.....	B-6
B.2.2	Nodal Configuration	B-7

Table of Figures

Figure 2-1 - Functional illustration of the Centralized Signalling Approach	11
Figure 2-2 - Functional illustration of the Distributed Signalling Approach.....	12
Figure 2-3 - Optical Control Plane Functional Model	20
Figure 3-1 - Illustration of typical lightpath reservation with no OXC Switch Time...	25
Figure 3-2 - Illustration of typical lightpath setup with OXC Switch Time	26
Figure 3-3 - Illustration of competing lightpath requests	27
Figure 4-1 - Architectural Diagram of an OBGp Gateway Router.....	33
Figure 4-2 - OBGp Message Format	35
Figure 4-3 - λ RI Entry Format	36
Figure 4-4 - OBGp Wavelength Table Fields.....	37
Figure 4-5 - Illustration of the OBGp Conservative Signalling Algorithms	39
Figure 4-6 - Illustration of the OBGp Aggressive Signalling Algorithm.....	43
Figure 5-1 - Simulation Network in OPNET	50
Figure 5-2 - Average Percentage of Blocked Calls with First-Fit Wavelength Selection Algorithm.....	53
Figure 5-3 - Nodal Blocking Summary at 1 call per second with 95% Confidence.....	54
Figure 5-4 - Average Lightpath Setup Time with 95% Confidence	55
Figure 5-5 - Average Wavelength Utilization per link at 1 call per second with First-Fit Wavelength Selection Algorithm.....	56
Figure 5-6 - 2PC Average Percentage of Blocked Calls using Random Wavelength Selection Algorithm with 95% Confidence	59

Figure 5-7 - 4PC Average Percentage of Blocked Calls using Random Wavelength Selection Algorithm with 95% Confidence	59
Figure 5-8- Nodal Blocking Summary at 5 calls per second using Random Wavelength Selection Algorithm.....	61
Figure 5-9 - Percentage of blocked Discovery Phase at 5 calls per second.....	62
Figure 5-10 - Average Lightpath Setup Time with 95% Confidence.....	64
Figure 5-11 - Average SCS Request Queue Size with 95% Confidence.....	65
Figure 5-12 - Average Wavelength Utilization per link at 5 calls per second with Random Wavelength Selection Algorithm	67
Figure 5-13 - Average Percentage of Blocked Messages under different algorithms with 95% Confidence.....	70

List of Acronyms

λ RI:	Lambda Reachability Information
2PC:	Two-Phase Conservative Signalling
4PA:	Four-Phase Aggressive Signalling
4PC:	Four-Phase Conservative Signalling
AS:	Autonomous System
BGP:	Border Gateway Protocol
CU:	Control Unit
CCU:	Central Control Unit
ClaServ:	Classification of Services
CST:	Connection Switch Table
DCU:	Distributed Control Unit
EMS:	Element Management System
E-NNI:	External Network-to-Network Interface
FEC:	Forward Equivalency Class
GMPLS:	Generalized Multi-Protocol Label Switching
IETF:	Internet Engineering Task Force
I-NNI:	Internal Network-to-Network Interface
IP:	Internet Protocol
IS-IS:	Intermediate-System-to-Intermediate-System
ISP:	Internet Service Provider
ITU:	International Telecommunications Union
LDP:	Label Distribution Protocol

LSP: Label Switched Path

MPLS: Multi-Protocol Label Switching

NMS: Network Management System

NNI: Network-to-Network Interface

NLRI: Network Reachability Information

OADM: Optical Add/Drop Multiplexers

OBGP: Optical Border Gateway Protocol

OEO: Optical-to-Electrical-to-Optical

OIF: Optical Internetworking Forum

OOO: Optical-to-Optical-to-Optical

OSPF: Open Shortest Path First

OXC: Optical Cross Connect

RIB: Routing Information Base

RIP: Routing Information Protocol

Rsv: Wavelength Reservation Method

RSVP-TE: Resource Reservation Protocol with Traffic Engineering

RWA: Routing and Wavelength Assignment

SAN: Storage Area Network

SCS: Switch Control Server

SLA: Service Level Agreement

S-D Pair: Source-Destination Pair

TCP: Transmission Control Protocol

Thr: Protecting Threshold Method

UNI: User-to-Network Interface

VPN: Virtual Private Network

WDM: Wavelength Division Multiplexing

1 Introduction

1.1 Background

As Internet usage increases and new applications require new network architectures, the migration from a standard statically configured network to a dynamically allocated network is inevitable. Applications such as high capacity Grid Networks or Storage Area Networks (SAN) require high capacity connections between nodes. A network topology that can be configured and reconfigured based on clients' needs is beneficial. To actively fulfill the demands for high capacity requirements, the current trend is to look towards all-optical networking as a solution.

Wavelength Division Multiplexing (WDM) all-optical networks have become the primary solution in handling the increasing bandwidth demands of the network [Mukherjee 97]. WDM allows for one fibre to carry multiple wavelengths, achieving high bandwidth rates. In traditional networks, signals transmitted from end users had to be converted from electrical signals to optical, and then back to electrical at each node in the network. This conversion is called OEO, or *Optical-to-Electrical-to-Optical* Conversion [Sharma+ 97]. OEO conversion depends on the bit rate and the protocol of the signal. OEO is limiting, since the signal must be converted at every node along the path, slowing its speed. Another name for "OEO networks" is "opaque networks" because the light signal must be stopped and converted electrically [Labourdette+ 03]. The OEO solution is not scalable for future networks.

In a "transparent network", which is known as the all-optical network or *optical-to-optical-to-optical* (OOO) network, the signal remains in the optical domain when it

leaves the source, until the destination. The optical path that the signal traverses is a WDM channel, and can also be referred to as a *lightpath* [Chlamtac+ 92]. Recent developments in *Optical Cross-Connects* (OXCs) and all-optical add/drop multiplexers (OADMs) are making the all-optical network a reality [Chu+ 02].

Scaling the Internet with the all-optical network infrastructure requires novel methods to control the network elements. Separate control domains are required to maintain the interconnections for the all-optical network. Within a control domain, several solutions persist, but the problem of interdomain routing considerations is currently being studied by the IETF [Bernstein+ 02]. Novel methods of setting up end-to-end lightpaths are required.

Traditionally, end-to-end circuit connections have been established via a *Network Management System* (NMS), which issues commands to the various network elements involved in the circuit via an equipment vendor's *Element Management System* (EMS). This is typically the carrier's responsibility and the lightpath setup and configuration have been based on a centralized signalling model. By moving this responsibility from the carrier to the customer and using a distributed approach, customers can better manage their optical wavelengths.

Demands for client initiated lightpaths are increasing, as new applications require high capacity links for only short periods of time [St.Arnaud+ 00]. Grid computing [Foster+ 01], SANs and distance based health and education applications are examples technologies which can benefit from the use of lightpaths. In most cases, the lightpath is required to send a large amount of data. This data may be in the form of large personnel databases requiring backup, MRI images, or particle physics simulation results. The

connection may be semi-permanent, or may last seconds, depending on the application or client requirements.

Previous literature has extensively studied the many different signalling algorithms introduced by all-optical networks [Qiao+ 96] [Mei+ 97] [Zang+ 00] [Zang+ 01] [Zheng+ 01] [Yuan+ 96] [Yuan+ 99] [Pezoulas+ 03]. Few have considered the physical layer properties and its effect on the performance of the establishment of lightpaths in the network.

In this thesis, signalling methods will be studied with the introduction of the OXC Switch Time. This study will be part of a new interdomain routing protocol called the *Optical Border Gateway Protocol* (OBGP), which has been proposed and implemented [Francisco+ 01] [Francisco+ 02] [Pezoulas+ 03] [Pezoulas 03b].

1.2 Thesis Objectives

The purpose of this thesis is to:

1. Develop an interdomain control protocol for managing optical resources through the optical layer by extending the traditional BGP protocol.
2. Demonstrate how routing and signalling can be successfully combined within the OBGP protocol.
3. Evaluate the three signalling algorithms provided by the OBGP protocol.
4. Determine the effect of the Optical Cross Connect Switch Time in the establishment of end-to-end lightpaths.
5. Analyse the performance of the OBGP signalling algorithms through simulation results.

1.3 Thesis Contribution

The following are the main contributions of this thesis:

- Creation of the Optical Border Gateway Protocol as a solution for the interdomain control protocol for all-optical networks.
- Evaluation of distributed backwards signalling algorithms under fixed routing conditions in a heavily congested linear network.
- Empirical evaluation of the impact of the OXC Switch Time on the establishment of lightpaths across interdomain networks.
- Implementation of the OBGp simulation model in OPNET Modeler.

1.4 Publications

1.4.1 Refereed Conferences

- L. Pezoulas, M.J. Francisco, I. Lambadaris, C. Huang, “Performance Analysis of a Backward Reservation Protocol in Networks with Sparse Wavelength Conversion”, The proceedings of the IEEE International Conference on Communications 2003, Anchorage, AK, Volume 2, pages 1468-1473, May 2003.
- M.J. Francisco, L. Pezoulas, C. Huang, I. Lambadaris, “End-to-End signalling and routing for optical IP networks”, The proceedings of the IEEE International Conference on Communications 2002, New York City, NY, Volume 5, pages 2870-2875, May 2002.
- M.J. Francisco, S. Simpson, L. Pezoulas, C. Huang, I. Lambadaris, W. St.Arnaud, “Interdomain Routing in Optical Networks”, The proceedings of SPIE Opticomm 2001, Denver, CO, Volume 4599, pages 120-129, August 2001.

- Y. Li, M.J. Francisco, I. Lambadaris, C. Huang, “Traffic Classification and Service in Wavelength-Routed All-Optical Networks”, The proceedings of IEEE International Conference on Communications 2003, Anchorage, AK, Vol. 2, pages 1375-1380, May 2003.
- Y. Li, M.J. Francisco, I. Lambadaris, C. Huang, “Traffic Classification and Service in Wavelength-Routed All-Optical Networks”, The proceedings of SPIE Asia-Pacific Optical and Wireless Communications Conference 2002, Shanghai, China, Vol. 4910, pages 301-309, October 2002.

1.4.2 Poster Presentations

- M.J. Francisco, I. Lambadaris, C. Huang, “End-to-End Signalling and Routing for Optical IP Networks”, a poster in MITACS AGM 2002, University of British Columbia, Vancouver, BC, May 2002.
- M.J. Francisco, I. Lambadaris, C. Huang, “Interdomain Routing in Optical Networks”, a poster in CITO Knowledge Network Conference, Ottawa, ON, October 2001.

1.5 Thesis Organization

This thesis is organized as follows: Chapter 2 gives background information about All-Optical Networks. The problem of the Optical Cross Connect Setup Time is introduced in Chapter 3, as is the problem of interdomain control protocols for all-optical networks. Chapter 4 introduces our design of the Optical Border Gateway Protocol (OBGP) as a solution to the interdomain control protocol and discusses its architecture and signalling algorithms. Major contributions are presented in Chapter 5, with a

performance analysis of the system under different signalling schemes. Finally, conclusions are drawn out in Chapter 6.

2 All-Optical Networks

2.1 Introduction

All-Optical networks are networks where the signal is transmitted completely in the optical domain, avoiding the need for OEO conversions. The major benefit of all-optical networks is reducing the cost and increasing the scalability of the network, since transceivers are very expensive. The introduction of WDM harboured the ability to send multiple signals through different wavelengths in the same fibre. Since a wavelength is agnostic of the protocol and bit rate, a single fibre multiplexing several wavelengths can transmit terabits of data, depending only on the connected equipment. Controlling the all-optical network has been in discussion in the OIF, ITU and IETF [Banerjee+ 00] [Bernstein+ 02]. This chapter is broken into 4 distinct sections: In the first section the Routing and Wavelength Assignment in all-optical networks is discussed. The second section describes signalling in all-optical networks. Thirdly, architectures of controlling all-optical networks in the interdomain are described. Lastly, how it parallels issues faced in the IP world.

2.2 Routing and Wavelength Assignment in All-Optical Networks

Establishment of lightpaths between source and destination pairs within an all-optical network can be decomposed into two sub-problems:

1. Route determination for the lightpath
2. Wavelength selection and assignment for the lightpath

The combined problem is better known as the *Routing and Wavelength Assignment* (RWA) problem [Ramaswami+ 95]. The objective of RWA is to route lightpaths and assign wavelengths in a manner that minimizes the amount of consumed network resources, while ensuring that lightpaths do not share the same wavelength along the same fibre link. The absence of wavelength converters along the lightpath restricts the establishment of the lightpaths to use the same wavelength, which is also known as the *Wavelength Continuity Constraint* (WCC) [Banerjee+ 00] [Mei+ 97].

Two basic types of WDM networks exist: static WDM networks and dynamic WDM networks. In a static WDM network, lightpath requests are known in advance and the network can be planned from the outset. In a dynamic WDM network, requests for connections arrive to the network dynamically and remain for an amount of time before the connection is torn down. If there are enough resources in the network, then the lightpath can be assigned. If a lightpath cannot be assigned, because of insufficient resources making the destination unreachable, then the request is said to be *blocked*.

Different routing algorithms exist for determination of a source-destination pair, such as Fixed Routing (FR), Fixed Alternate Routing (FAR) [Harai+ 97] [Ramamurthy+ 98], and Adaptive Routing [Mokhtar+ 98]. In Fixed Routing (FR), a single path is determined for a source-destination pair. These predetermined paths can be discovered through different heuristics, such as shortest path routing. In FAR, a set of source-destination pairs are predetermined within a network. As a connection request arrives, a single route is chosen from the set of available routes and a connection is established along that path. In Adaptive Routing, the network state is considered when

determining least cost path for a specific lightpath request by assessing the current state of the network.

Once a path is chosen for a specific lightpath the wavelength must be selected to fully establish the end-to-end connection from a set of feasible wavelengths between the source and destination. The second part of the RWA problem is the wavelength assignment, or wavelength selection. Several different algorithms have been proposed for the wavelength selection process, the simplest algorithms being First-Fit, Random, First-Available, Most-Used and Least-Used [Chlamtac+ 92] [Birman 95] [Subramaniam+ 97] [Zang+ 01]. In the First-Fit and Random algorithms, the wavelengths are indexed in a scheme that is uniform across the entire network. In the First-Fit algorithm, the wavelength with the lowest index in the set of available wavelengths will be selected as the candidate wavelength. The Random algorithm is similar, except that the wavelength is selected randomly with an even distribution. In the Least-Used algorithm, each wavelength maintains a usage counter, which is incremented as the wavelength is repeatedly chosen. The wavelength with the lowest counter value is selected as the candidate wavelength. This method gives preference to shorter lightpaths and is known to under perform compared to the Random and First-Fit algorithms [Rouskas 02]. The Most-Used algorithm is similar to Least-Used, except that the wavelength which is used the most is selected as the candidate wavelength. Most-Used outperforms Least-Used, Random and First-Fit in blocking probability. The drawback of the Least-Used and Most-Used algorithms is that global information about the network is necessary and is not feasible in a distributed control environment. The First-Available

Algorithm was presented and was shown to outperform First-Fit and Random selection algorithms [Pezoulas+ 03].

2.3 Signalling in All-Optical Networks

All-optical networks come in a variety of flavours. Connection oriented all-optical networks require lightpaths to be established between end nodes, whereas connectionless networks, such as optical packet switched networks, require new technologies to route and forward packets. Connection oriented all-optical networks, are circuit switched in nature and require a signalling algorithm to reserve resources along the path and construct the lightpath [Banerjee+ 01]. There are two types of control for signalling: centralized and distributed [Mei+ 97].

2.3.1 Centralized Signalling

In the Centralized Signalling scheme, a central node maintains knowledge of the entire network. The Central Control Unit (CCU) contains a database of the status of each node, the state of each link and the connections that are currently active. The CCU is connected to each client within the network in a control network. Those clients are also connected through their optical interfaces, to the data network.

When a lightpath is required from the source to the destination, the requesting client node sends a request message to the CCU, informing it that a lightpath is needed to a given destination client. Upon receiving the request, the CCU determines the path through its knowledge of all resources in the network, which is kept current through resource updates within the network. If enough resources exist to establish a lightpath, all OXCs along the path are instructed to establish the lightpath by reconfiguring their

switch fabric. Once the lightpath has been established, a signal is sent to the client node informing it that the setup is complete. Figure 2-1 illustrates the architecture of the centralized signalling approach.

The advantage of Centralized Signalling is that there is one centralized server to appropriately assign lightpaths across the network, thus having a consistent view of the network. The disadvantage of centralized signalling is that the CCU becomes overburdened with a high degree of route computation for each request. The network architecture doesn't scale well. In addition, it becomes a single point of failure for the entire network.

Figure 2-1 - Functional illustration of the Centralized Signalling Approach

2.3.2 Distributed Signalling

The alternative to the centralized signalling scheme is a decentralized approach [Mei+ 97] [Zheng+ 01]. If no central node exists, then signalling is performed in a

distributed manner. Figure 2-2 illustrates the distributed signalling architecture, which consists of a control network and a data network. The control network is connected via a separate in-band or out-of-band channel, denoted by the dashed line. The data network consists of optical switches and data transmission channels. These channels form the lightpaths. The hosts contain transmitters and receivers that transmit signals optically, while the OXCs switch the signal from one channel to another. In the case of All-Optical networks, these OXCs switch wavelengths. Each Distributed Control Unit (DCU) in the network maintains a table which contains the status of each wavelength on each port, plus the current connections that the node maintains.

Figure 2-2 - Functional illustration of the Distributed Signalling Approach

A wavelength can maintain three states:

- *Available* – The wavelength is currently unused, and may be reserved for a lightpath.

- *Reserved* – The wavelength is being held for a connection. While in this state, the wavelength cannot fulfill any other lightpath requests. If a wavelength is reserved, there is no guarantee that the lightpath will be successfully established.
- *Unavailable* – The wavelength request was fulfilled and the wavelength is carrying data traffic.

The advantage of Distributed Signalling is that it scales with the size of the network. There is no single point of failure in the network and each node maintains a visibility to neighbouring nodes, lessening the route computation. The disadvantage is that routing and wavelength status information must be distributed, requiring the need for updates that are sent to all participating nodes in the network, increasing the control traffic. An inconsistent view of the system state results when updates are distributed infrequently.

The performance of various distributed control protocols have been studied in [Qiao+ 96] [Yuan+ 96]. Performance analysis was conducted under different signalling schemes. The first scheme was *Path Multiplexing* (PM), which required that the same wavelength be used all along the lightpath and is also known as WCC. The second scheme, known as *Link Multiplexing* (LM), allowed for different wavelengths to be used along the lightpath. This required that each node in the network have a wavelength converter.

Two methods of signalling were introduced for distributed signalling: *forward reservation* and *backward reservation* [Qiao+ 96]. In the forward reservation scheme, a control message propagates from the source to the destination, reserving one or more wavelengths along the forward path on a hop-by-hop basis. The source node initially

selects a wavelength as its candidate wavelength for the establishment of a lightpath. As the control message passes through an intermediate node along the path, the node evaluates the available wavelengths and reserves one or more wavelengths found on the following link. The request is blocked if no wavelengths are available to the next link. When blocked, a failure message is returned to the source along the same path, unreserving the previously reserved wavelengths returning all wavelengths to their original state. If the control message successfully reaches its destination and is never blocked, then the lightpath is established. An ACK message is returned back to the source, along the same path.

The primary problem with forward reservation is its inability to foresee the network state. For every attempt that forward reservation makes, there is no guarantee of the availability of the selected wavelength on the following link. If a request is blocked, the request must be made again, which can result in several requests before a lightpath is successfully established.

The other option is a backward reservation scheme. In this scheme, the wavelength is reserved at the destination. A control message is propagated from the source to the destination first. As the message is received along any intermediate node, it collects wavelength information along the path, finding a set of common available wavelengths from the source to the destination. When the request message reaches the destination, the message contains the set of available wavelengths that can be used to setup a lightpath all along the path. The destination then selects a wavelength and sends a message back to the source along the same path, reserving the selected wavelength. By reserving wavelengths in the reverse direction, wavelengths are idle for less time than if

wavelengths are reserved in the forward direction. The disadvantage of backward reservation schemes is that the request message that “discovers” the available wavelengths may have out-of-date information when the source receives the set of available wavelengths. This problem is known as the *stale link state problem* and is encouraged by long round trip times between the S-D pair and the possibility of competition if several lightpath requests share the same path.

There are two different types of reservation methods which can be coupled with forward and backward reservation schemes [Yuan+ 96]. As request messages propagate between nodes, one or more wavelengths can be reserved. Under the Aggressive Reservation Method, all available wavelengths are reserved for the lightpath. This method was thought to increase the possibility of establishment of a lightpath, as all resources between a source-destination pair are reserved. The Conservative Reservation allows for only one wavelength to be reserved at a time. This reservation scheme only reserves necessary wavelengths, leaving remaining wavelengths for other competing lightpath requests. The drawback of conservative reservation is that if the competition is intense, then the Stale Link State problem becomes apparent.

The different combinations of Forward and Backward Reservation gives us four different flavours of signalling, described as follows:

1. *Aggressive-Forward Reservation*: The source initiates a request message and reserves all wavelengths along the path to the destination. At the destination, one candidate wavelength is selected and a message is returned to the source, releasing all unused wavelengths. This signalling scheme requires two signalling phases: The forward propagating message where a reservation message travels

from the source to the destination, and a backward propagating acknowledgement message that travels from the destination to the source.

2. *Conservative-Forward Reservation:* The source selects one candidate wavelength and sends a request message to the destination. At each intermediate node, the request message reserves that one wavelength, and is blocked if it is unavailable. At the destination, an acknowledgement is sent along the path to confirm the establishment of a lightpath. As in the Aggressive-Forward Reservation scheme, there are two phases involved: The forward propagating reservation message and the backward propagating acknowledgement message.
3. *Aggressive-Backward Reservation:* The source initiates the reservation by sending a request to the destination. At each intermediate node, the request message is updated by maintaining a set of available wavelengths, which are common to the source-destination pair. At the destination, a message is returned to the source, reserving all wavelengths within that set. When the source receives this message, it selects a candidate wavelength and transmits an acknowledgement message to the destination, which also unreserves wavelengths along the path. The Aggressive-Backward Reservation message requires three phases to establish a lightpath: The forward propagating message discovers all common and available wavelengths between the source and the destination. A message travels from the destination to the source reserving the set of available wavelengths. At the source, a wavelength is selected and all other wavelengths along the path must be unreserved to complete the creation of the lightpath.

4. *Conservative – Backward Reservation:* In Conservative Backward Reservation, the source discovers all common wavelengths from the source to the destination. At the destination, a wavelength is selected as the candidate wavelength for the lightpath and a message is sent back to the source. There are two phases involved in Conservative-Backward Reservation: The forward propagating message that discovers all available wavelengths and the backward propagating message that reserves the one wavelength along the path.

It has been found that backward reservation schemes outperform forward reservation schemes with respect to the blocking probability [Yuan+ 96] [Mei+ 97]. Backward reservation schemes reduce bandwidth waste during the reservation period and their usefulness is proportional to the propagation delay. Aggressive schemes also suffer from a higher blocking probability. More wavelengths are held for a longer period, causing other requests to be blocked even though the network may have the bandwidth for more lightpath requests.

Considerations of propagation delay are made in [Mei+ 97] [Zang+ 01], without the OXC Switch Time and the effect it has on the typically two-phase lightpath establishment in WDM Networks. This issue is further elaborated upon in Chapter 3.

2.4 All-Optical Network Control Plane Architectures

Connection oriented All-Optical Networks are circuit switched networks, due to their high capacity and slow switching speed. The Internet, on the other hand, is a connectionless packet-switched network. The integration of the IP layer with the optical network has been a widely discussed issue in [Rajagopalan+ 00] [Li+ 02] [Rajagopalan+ 04]. The functions of the control plane are to discover the neighbours and

routes as well as signal and manage resources. Some of the possible interconnection models for IP over Optical networks are: the overlay model, the peer model and the interdomain model [Rajagopalan+ 00].

The *overlay model* is similar to the classical IP over ATM model. Under this model, optical networks are treated as a transparent sub-layer and the IP layer is independent of the corresponding protocols in the optical networks. This approach has been widely studied and is being standardized in OIF (Optical Internetworking Forum) and IETF (Internet Engineering Task Force). While this approach may lead to early deployment of optical networks, it is a loosely coupled solution. With the development of optical switches, it is foreseeable that optical packet networks may be possible in the future.

Under the *peer model*, optical cross-connects (OXC's) are treated just like any other router in terms of the control plane. This will allow a seamless interconnection of IP and optical networks. As pointed out in [Rajagopalan+ 00], the drawback of this approach is that it requires routing information specific to optical networks to be known to routers and may eventually overburden router networks.

The *interdomain model* allows for separate instances of the IP and optical domains. Information from optical networks is selectively passed to the router networks to allow internetworking through interdomain routing protocols. Overburdening the router network can be avoided by selectively choosing the information to be passed. This model is further discussed in Section 3.3.

Research work thus far has been focused on the overlay model [Saha+ 03]. With this model, networks are classified into either optical networks or client networks. Each

optical network is a small and manageable domain typically owned and controlled by one carrier. Each client is a higher layer network such as an IP network. Figure 2-3 shows a diagram of the interconnections of the different networks. There are several different interfaces that interconnect the networks with one another.

UNI and NNI refer to the control interface types between client networks and provider networks.

- UNI – The User-Network Interface is the interface type between a client network and a node in the optical network.
- I-NNI – The Internal Network-Network Interface is the interface type between two nodes within the same control domain.
- E-NNI – The External Network-Network Interface is the interface type between two nodes in different control domains.

Figure 2-3 - Optical Control Plane Functional Model

The optical network, as stated above, is broken into smaller, more manageable domains called Control Domains. These domains interact using the I-NNI. Requests for lightpaths come from Client Networks which use the UNI. The focus of the study is the E-NNI, the interaction of the control domains of the Optical Networks. The E-NNI control protocol must manage routing information on the network, and perform signalling to establish end-to-end lightpaths.

The Generalized Multi-Protocol Label Switching (GMPLS) Protocol has been introduced as a method for managing and controlling network elements for the establishment of Label Switched Paths (LSPs) in carrier networks [Banerjee+ 01]. While GMPLS is a new technology, efforts have been made to push its predecessor, MPLS, as

the control plane technology for traditional IP Networks and is described in the following section.

2.5 Border Gateway Protocol and Virtual Private Networks

The Border Gateway Protocol is the de-facto standard for IP routing in the interdomain [RFC 1771]. The purpose of BGP is to determine the reachability of networks across different domains. A more detailed description of BGP version 4 is available in Appendix A. BGP and the Multi-Protocol Label Switching (MPLS) Protocol have been coupled to maintain end-to-end network connectivity in traditional IP datagram networks, acting as an interdomain control protocol for IP networks. This technology has allowed carriers to provide private network services across different domains. This service is known as “Virtual Private Networks” and guarantees a level of control over an IP network that spans other provider networks which are MPLS enabled. This unique end-to-end service parallels the purpose of an all-optical network control plane architecture.

MPLS is a protocol that creates Label Switched Paths (LSPs), which simplify the forwarding of packets in a network [RFC 3031]. MPLS enhances source routing and introduces circuit switching in a typically packet switched network. This results in controlled Quality of Service (QoS), Traffic Granularity and Traffic Engineering aspects. MPLS uses a signalling protocol, the Resource Reservation Protocol with Traffic Engineering (RSVP-TE) [RFC 3209] to bind MPLS labels to Forward Equivalency Classes (FECs) and distribute them to MPLS Peers.

Current trends in the MPLS VPN area have been pushing for standardization of the method to establish the LSPs between the different source and destination pairs. One

recent trend being pushed is to use Label Distribution Protocols (LDP), such as RSVP-TE, to distribute the label amongst MPLS peers across the interdomain. In the interdomain, this solution does not scale as different service providers (domains) have diverse protocols implemented within the intradomain. Carriers are not forthcoming with sharing their own network information with other service providers. To address this issue, BGP has been extended as a method to distribute the labels, in a proposed BGP/MPLS VPN solution [RFC 2547]. The major benefit of adding signalling functionality to BGP is that it is a lightweight solution. The opposition states that BGP is overloaded and a separate LDP would be better suited, since specific functionality can be put into the LDP protocol.

BGP is well suited for signalling since it is deployed in all carriers' networks and is the de-facto standard for routing. Extending BGP to support label distribution would facilitate the upgrade of the networks and provide an easy solution for interdomain VPN service. The other argument is that by providing multi-services, separate protocols versus one overburdened protocol require the same amount of resources and computing power [Reardon 03].

3 OXC Switch Time Problem and Interdomain Control Protocols in All-Optical Networks

3.1 Introduction

Signalling methods for all-optical networks have been extensively studied in previous literature. The methods and differences are well-noted, as described in the previous chapter. Previous literature has also detailed the switching fabrics of Optical Cross-Connects (OXC), implementing new switch architectures which decrease switching time, but increase performance [Ma+ 03]. The focus of the first part of this chapter is the inherent problem overlooked by the previous literature: the effect of the OXC Switch Time on all-optical signalling algorithms. A method of implementing control domain protocols is the focus of the second part of this chapter.

3.2 OXC Switch Time Problem

Optical Cross-Connects are typically Micro-Electro Mechanical (MEM) based switches, which direct signals of light from one input port to an output port. They are physical layer devices which are independent of wavelength, bit rate, protocol, modulation format, fibre mode and conversion capabilities. They can be as small as a 1x1 switch (1 input port and 1 output port) and can be as large as a 1000x1000 switch. An OXC with W wavelengths, N input and N output ports can be described as W independent $N \times N$ OXCs [Rouskas 02].

An OXC requires a fixed amount of time to reconfigure the switch fabric completing the setup of an input port to an output port. This time may vary between

picoseconds to seconds, based on the switch fabric technology, as mentioned in [Ma+ 03]. A signal cannot be transmitted to the input port of an OXC immediately after a request for a switch to the OXC occurs. If a signal enters the OXC input port before the completion of the OXC switch, the signal integrity cannot be guaranteed for the corresponding output port and can possibly affect the existing connections maintained by the switch. Once the switch is completed, then the signal can safely be transmitted along the path.

In previous literature, the OXC Switch Time was never considered when measuring the performance of a network with dynamic lightpath creation [Pezoulas+ 03] [Barry+ 95]. Two problems arise from the exclusion of the OXC Switch Time for signalling. Firstly, signalling requests must be queued at each node, as the OXC Switch Time occurs. This involves creating network elements with a queueing system, effectively changing the behaviour of the signalling algorithm. The second problem that arises can be traced to the inconsistency of the system state, when the OXC Switch Time is introduced to signalling systems. Figure 3-1 and Figure 3-2 illustrate how the OXC Switch Time affects the time required for establishing a lightpath. The path consists of three types of nodes: a source (S), several intermediate nodes (I) and a destination (D). Between the source and the destination, there are n intermediate nodes. Linking all nodes along the path are $n+1$ links with a propagation delay of d_{prop} . A processing delay is required at each intermediate node, which is denoted as d_{proc} . A special processing delay made at the destination node is denoted as d_{sel} , and represents the time required for selecting a wavelength. A queueing delay is introduced at each intermediate node and is denoted as d_{qi} , where i is the intermediate node index.

Figure 3-1 - Illustration of typical lightpath reservation with no OXC Switch Time

Figure 3-1 illustrates the scenario where the OXC Switch Time is not considered in signalling. It should be noted that this is the typical signalling method that is studied in previous literature. When $t = 0$, a request is generated from S , discovering all available wavelengths linking S to D . At each node, a small processing delay is experienced, eliminating uncommon wavelengths from the set. As the request reaches D , a processing delay is involved for the wavelength selection and the message is sent back to S . The setup delay of a lightpath without the OXC Switch Time is found in equation (3.1).

$$t_{setup} = 2((n+1)d_{prop} + nd_{proc}) + d_{sel} \quad (3.1)$$

Figure 3-2 - Illustration of typical lightpath setup with OXC Switch Time

Figure 3-2 illustrates the addition of the OXC delay time, d_{oxc} , at each intermediate node. We consider here that the OXC Switch Time includes the processing delay (d_{proc}). This delay affects the overall t_{setup} , with respect to the number of intermediate nodes the path contains. The setup time for one request is found in equation (3.2).

$$t_{setup} = 2(n+1)d_{prop} + n(d_{proc} + d_{oxc}) + \sum d_{qi} + d_{sel} \quad (3.2)$$

As OXC Switch Time increases, the lightpath setup time increases linearly with the number of intermediate nodes. The increase in setup time raises the possibility that

the link state changes, before the message has completed the lightpath request. By introducing competing S-D pairs, the status of a wavelength may change from the time that the availability was discovered during the discovery phase, to the time that the setup is completed. Figure 3-3 attempts to illustrate this point.

Figure 3-3 - Illustration of competing lightpath requests

The network in Figure 3-3 shows two S-D Pairs ($S1$ to $D1$ and $S2$ to $D2$) with shared links between $I1$, $I2$ and $I3$. The links contain fibres which maintain only one wavelength, λ , for illustrative purposes. Initially, the network does not have any lightpaths setup, and the wavelength, λ , is available for usage. At time t , $S1$ requests a lightpath to $D1$ and discovers that wavelength λ , is available to its destination, which is illustrated by the arrow labeled (1), in Figure 3-3. Shortly after the request leaves $S1$, another request is made by $S2$, which also discovers that wavelength λ , is available to its destination, illustrated as (2). Since the propagation delay for $S2$ - $D2$ is shorter than $S1$ - $D1$, $D2$ selects λ and tries to reserve it along the $D2$ - $I2$ link, illustrated as (3). At $I2$, the request is halted, while the OXC is set up for the connection. As $S2$'s request is being served at $I2$, $D1$ receives $S1$'s request, selects λ and reserves the wavelength along the $D1$ - $I3$ link, labeled as (4). After a given switch time, $S2$'s request completes the setup at $I2$ and continues back to the source. Unfortunately for $S1$'s request, the status of λ has

changed, and as the request propagates from $I3$ to $I2$, $I2$ cannot fulfill the request since the wavelength is now *Unavailable*. This is illustrated as (5) in Figure 3-3.

This problem is known as the *Fairness Problem* [Barry+ 95], wherein the shorter lightpaths (those with shorter hop-counts or smaller propagation times) will reserve resources quicker than longer lightpaths. Previous studies have suggested solutions to the Fairness Problem, such as the Protecting Threshold Method (*Thr*), the Wavelength Reservation Method (*Rsv*) in [Birman 95] and the Classification of Services (*ClaServ*) method in [LiFrancisco+ 03]. The focus of those studies was the effect of the lightpath setup and blocking caused by interference with competing requests.

Previous works have not studied how the OXC Switch Time, if varied, affects the lightpath setup time and blocking probability of the system.

3.3 Interdomain Control Protocol for All-Optical Networks

The Internet is an amalgamation of several interconnected domains appearing to be a unified network. Each domain may be owned or operated by different carriers, each maintaining their own routing protocols and standards, which may or may not mesh with a neighbour's control domain. The role of the interdomain control protocol is to allow a new and novel method for interconnecting different networks to function as a single entity. This is also known as the E-NNI [Rajagopalan+ 04].

The difficulty arises when unifying the network for the common goal of establishing end-to-end lightpaths in the network. A centralized controller, which would work well in the intradomain, would never be accepted amongst the carriers for the interdomain, since many would not agree on the centralized party to manage the system.

There is also the difficulty of hiding or sharing information from one domain to another domain. Thus, a distributed protocol that uses signalling is necessary in the interdomain.

Routing resource information cannot be exchanged easily from one neighbour to another in the interdomain since information can become out of date quickly, leaving the source with an inappropriate view of the status of the network. A signalling protocol must be implemented to “*discover*” the available resources along each link from end-to-end to acquire the most recent status of the network resources.

Two different approaches can be considered to address the interdomain control protocol for optical networks. Since this area can be considered a green field, one can either develop a brand new protocol, or extend existing and proven interdomain protocols.

The benefit of building a new protocol is to build a solution that specifically addresses the problem, with the scalability to answer future demands. The drawback with this approach is the difficulty of creating a robust protocol that is easily accepted by the networking community. Since the area of the interdomain is very sensitive between carriers, installing a new protocol is difficult. Like the problem of the BGP/MPLS VPNs, carriers argue that using previously installed protocols has an advantage over new and untested protocols.

The second approach has an advantage over the first since it extends existing and proven protocols. In the interdomain, BGP is the de-facto standard for interdomain routing in IP networks, making it a viable candidate for a new interdomain control protocol for all-optical networks. Extensions are necessary for BGP to be aware of the

optical properties such as wavelength routing and signalling for end-to-end lightpath connections.

In the following chapter, extensions are added to BGP, which can then be used as an interdomain control protocol for optical networks. The protocol enables client networks to signal to other domains, establishing end-to-end lightpaths, using a backward reservation algorithm, coupled with the protocol's four-phase reservation approach, which has never been studied.

4 Optical Border Gateway Protocol

4.1 Introduction

The Optical Border Gateway Protocol (OBGP) is a protocol providing edge customers a control method for establishing lightpaths across different autonomous domains through an optical network [Francisco+ 01][Francisco+ 02][Pezoulas+ 03]. This solution is proposed as an External NNI solution, for interconnectivity of external control domains. It was decided early on to extend BGP as an interdomain control protocol for optical networks, to leverage the BGP's acceptance in the interdomain. By extending the Border Gateway protocol, a method to establish lightpaths across several domains is introduced, using the existing IP network as the control plane. In this chapter, OBGP is introduced by outlining the requirements, design and benefits of using OBGP to set up and manage lightpaths.

4.2 OBGP Architecture

The Optical Border Gateway protocol is designed as an extension of the Border Gateway Protocol. BGP is an interdomain routing protocol for IP networks, which is further described in Appendix A. BGP is the de-facto standard, for routing in the interdomain. Routing protocols in the interdomain are designed with different features in mind as opposed to intradomain routing protocols. The reason why a separate routing protocol is required for the interdomain is to interconnect different networks together. These ISPs or carriers may compete with one another and thus, disclosure of internal

network configurations and paths may not be a viable option for business purposes. Thus, BGP divides networks into several Autonomous Systems (ASs), also called domains.

BGP was originally designed for the IP data network and does not contain any functionality for optical attributes, but some of BGP's attributes can be leveraged for OBG. BGP carries the full path information, which is useful in preventing routing loops. OBG benefits from BGP's full path visibility, using the *AS_PATH* as a fixed route for a lightpath.

The initial OBG design, presented in [Francisco+ 01], showed a protocol of OBG which combines both routing and signalling into a lightweight solution. It is now known that BGP routing can be used to determine source and destination pairs with a decoupled signalling protocol such as RSVP-TE. This is similar to the method that is being pursued in BGP/MPLS VPNs, where BGP is used as the signalling protocol to distribute labels. In the same fashion, wavelength information can also be distributed using a separate LDP in OBG.

The basic high-level requirements for OBG to work as a lightpath setup and control protocol are as follows:

- Provide a set of attributes to maintain lightpath information.
- Carry lightpath reservation requests and responses between OBG based devices.
- Propagate up-to-date information on the status of resources in the network.

Making minimum changes to BGP while meeting these requirements is a very important consideration. Leveraging current BGP properties is key to meeting these requirements.

Figure 4-1 - Architectural Diagram of an OBG Gateway Router

A typical OBG Router consists of three components, illustrated in Figure 4-1 [St.Arnaud+ 00]. The first component is a typical Border Gateway Router, which is found at the gateway of an AS. This router performs BGP peering and acts as a node in the control plane for OBG. The second component, the Switch Control Server (SCS), is a control interface between the router and the OXC. This interface maintains the connection state for the OBG process in the router. The third component is the OXC. In this study, the OXC has no wavelength conversion capability and can switch any wavelength from one port to another [Ramaswami+ 97b]. OBG with wavelength conversion capability was studied in [Pezoulas+ 03]. A more detailed description of the OBG Router is found in Section 4.4.

Initially, the UPDATE message was considered for carrying the lightpath setup information, in addition to normal reachability information [Francisco+ 01]. This would have required several extensions to be made to the normal UPDATE message. A decision was made to separate the optical information and create a new OBGp message. This message can be introduced to the network as a BGP Capability [RFC 2842] for flexibility. By using a separate message, it is easier to make changes to the protocol. Changes which are specific to the establishment of lightpaths are separated from the routing protocol itself, thus allowing traditional BGP routers and OBGp capable routers to inter-work seamlessly. Another reason to not piggy-back optical information with the UPDATE, is to decouple the routing and the signalling, allowing the use of separate signalling protocols.

The new message is called the OBGp message. The format of this message is described in Section 4.2.1. The concept of wavelength tables is introduced in Section 4.2.2. These tables are necessary for the SCS to maintain and store the status of wavelengths the router may own.

4.2.1 The OBGp Message

The OBGp Message is the primary message used for managing and configuring the optical cross connects and to update the wavelength tables at each OBGp device. The OBGp Message header is the same as the BGP Message header, as described in [RFC 1771] and Appendix A, containing the *Marker*, *Length*, and *Type* fields. The OBGp Message is introduced as a new “Type” of message and is identified by the type-code of 5.

4 bits	4 bits	4 bits	16 bits	Variable	Variable
Phase Id	Sub-code	Op. Mode	Attr. Length	Attributes	λ RI

Figure 4-2 - OBGp Message Format

In addition BGP's fixed-sized header, the OBGp message contains the following fields, also illustrated in Figure 4-2:

Phase ID: This 4-bit field indicates the phase of the lightpath setup process. The possible phases which are described in Section 4.3 are i) Discovery, ii) Reservation, iii) Setup, iv) Confirmation, v) Aggressive-Discovery, vi) Aggressive-Reservation, vii) Setup-Unreserve, viii) Teardown, ix) Error.

Sub-Code ID: This 4-bit field is used by an OBGp message in the Confirmation or Error Phase. For Confirmation Messages, it allows to distinguish between Setup-Confirmation and Teardown-Confirmation Messages. For Error Messages, it allows to indicate in what phase the error occurred.

Operational Mode: This 4-bit field is used to indicate the mode of operation of OBGp: Two-Phase Conservative, Four-Phase Conservative or Four-Phase Aggressive.

Attributes Length: This 16-bit unsigned integer indicates the length of the Attributes field.

Attributes: The supported OBGp attributes are as follows:

AS_PATH: This is the AS Path from the source to destination for which a lightpath is being setup. This path information is acquired from the BGP *Routing Information Base* (RIB) and is a list of nodes that the path must traverse. As an

OBGP message travels from one router to the other along this path, this attribute does not change.

NEXT_HOP: The NEXT_HOP attribute defines the IP address of the interface of the border router that should be used as the next hop for the OBGP Message.

CONN_ID: The CONN_ID attribute is used to identify lightpath connections, which are unique to the entire network. The connection identifier consists of a unique identifier plus the AS_PATH that corresponds to the specific lightpath.

λ RI: The Lambda Reachability Information (λ RI) attribute contains a list of allowable wavelengths that can be used for the lightpath connection, each of which is formatted as shown in Figure 4-3.

Lambda_num	Src_port_ID	Dst_port_ID
------------	-------------	-------------

Figure 4-3 - λ RI Entry Format

The fields in the λ RI are as follows:

Lambda_Num: The index of the wavelength.

Src_port_ID: The port on the OXC of the source end point.

Dst_port_ID: The port on the OXC of the destination end point.

4.2.2 The OBGP Wavelength Table

Each OBGP router must maintain the current status of the wavelengths in a wavelength table, which is similar to the *Connection Switch Table* (CST) in [Ramaswami+ 97a]. Using the wavelength table, OBGP can easily determine which

wavelengths are available to be used for lightpath setup. The fields of the wavelength table are illustrated in Figure 4-4.

AS Number	Lambda ID	Connection ID	Pref. Level	Setup Complete
-----------	-----------	---------------	-------------	----------------

Figure 4-4 - OBGW Wavelength Table Fields

AS Number: The AS Number of the neighbour border router to which the specific wavelength connects.

Lambda ID: This consists of the wavelength number, source port number and destination port number, as illustrated in Figure 4-3.

Connection ID: This consists of an identifier plus the AS_PATH that corresponds to the lightpath. If this field is NULL, then it means that the wavelength is available for a connection. If this field is assigned a *CONN_ID*, then the wavelength has been assigned to a connection and is unavailable. The numeric identifier plus the AS_PATH create a unique connection identifier to the entire network. The numeric identifier within the connection ID is not unique by itself across the network. It is only unique with respect to the same Source and same AS Path from that source.

Preference Level: This field is included for future use, to allow for assigning different preference levels to specific wavelengths.

Setup Complete: This Boolean field indicates if the connection for the specific wavelength has been completed by the OXC. If there is a *CONN_ID* in the *Connection ID field* and the *Setup Complete Field* is set to *TRUE*, then the wavelength is considered to be reserved.

4.3 OBGP Signalling Algorithm

OBGP employs several reservation techniques, for creating end-to-end lightpaths. OBGP implements a backward reservation protocol in Two-Phase Conservative (2PC), Four-Phase Conservative (4PC) and Four-Phase Aggressive (4PA) Modes. The decision to choose a backward reservation protocol for OBGP versus a forward reservation protocol was made because it was proven that the backward reservation protocols have a lower average blocking probability than forwards reservation protocols [Yuan+ 99]. The following sections describe the backward reservation signalling methods of the OBGP protocol.

Each OBGP node must be aware of its position within the *AS_PATH* in order to process the OBGP messages correctly. A node's classification will change according to the relative position in the lightpath segment that is being reserved. If a node is listed at the head of a lightpath reservation segment, then it is classified as the Source. If it is listed at the tail, then it is the Destination. Otherwise, the node is considered to be an Intermediate node.

The OBGP protocol can operate in two conservative modes and one aggressive mode. The following sections will explain the two conservative modes, which have been presented in [Francisco+ 02] and why only one aggressive mode is required.

4.3.1 Conservative Reservation

OBGP typically operates under two conservative reservation modes: Two-Phase Conservative (2PC) and Four-Phase Conservative (4PC). The 2PC scheme is similar to previous literature's signalling schemes, as reviewed in Section 2.3.2. The 4PC scheme is dissimilar to previous signalling algorithms, simply because it takes into account the

OXC Switch Time in its design. Figure 4-5 (a) and (b) illustrate the difference between 2PC and 4PC. In 4PC, all resources have to be reserved first before the setup of the OXCs can take place thus requiring extra phases. This ensures that during the Setup Phase, resources are reserved and guaranteed for the requested lightpath. On the other hand, in 2PC, Setup takes place immediately after having determined if there is a wavelength available for use. 2PC was thought to be a more efficient and faster approach, since a lightpath can be created within a round-trip time. As the number of users increases, lightpaths may overlap and compete for the same wavelengths. 4PC can be more efficient, since the user is notified of the wavelength reservation sooner, at the expense of taking longer to set up the lightpath.

Figure 4-5 - Illustration of the OBGP Conservative Signalling Algorithms

The phases are described below:

Discovery Phase: An OBG Message with the Phase ID field set to Discovery is created by the source node when a lightpath is required to be established. The source node, S , fills the λ RI field with all available wavelengths connected to the next node along the AS_PATH . The AS_PATH and the Connection ID are stored in the Attributes field of the message. The Discovery OBG Message traverses all nodes along the AS_PATH probing for the available wavelengths that can potentially be used for establishing the lightpath connection. If the OBG process determines that a specific wavelength is in use (by comparing the availability or existence of a wavelength in its wavelength table), then it eliminates the wavelength from the list of available wavelengths. When the destination node D receives the Discovery OBG Message, the λ RI contains all the common wavelengths that can be used for establishing a lightpath from the source to the destination. The destination node, D , can then select a wavelength using any wavelength selection algorithm and the Discovery Phase ends.

Reservation Phase: The Reservation Phase is specific to 4PC. The purpose of this phase is to reserve the selected wavelength along the AS_PATH . This Phase follows the Discovery Phase after the destination node selects a wavelength. Once a suitable wavelength is chosen by the Destination, the candidate wavelength is inserted into an emptied λ RI and placed into a new Reservation OBG Message, which is an OBG message with the Phase ID set to Reservation. The Connection ID field is set to be the AS_PATH and the unique identifier chosen by the source node. This message propagates from the destination to the source along the same AS_PATH , but in the reverse direction. When an Intermediate node receives the Reservation OBG Message, the node will find

the candidate wavelength in its wavelength table and sets the Connection ID field in the wavelength table to the Connection ID. By setting the Connection ID field in the wavelength table, the wavelength in that wavelength table is currently being reserved by that connection, thus, no other connection can use that specific wavelength. Any further Discovery OBGMP Messages that pass through this Intermediate node will eliminate this wavelength because it is found to be unavailable. When the Reservation OBGMP Message reaches the Source node, and all the nodes along the *AS_PATH* have updated their wavelength tables, then the Reservation Phase has completed.

Setup Phase: In the 4PC mode, the Setup Phase follows the Reservation Phase at the Source Node. If the Reservation OBGMP Message arrives at the Source node, then the candidate wavelength is reserved all along the *AS_PATH*, but the Intermediate nodes have yet to make any hardware connections, so data cannot flow through the lightpath. The purpose of the Setup Phase is to send a signal to all nodes along the *AS_PATH* configuring the switch fabric of each OXC in the path. The Source node creates an OBGMP Message with the Phase ID set to “Setup”, and propagates that along the *AS_PATH* in the forward direction. The λ RI contains the Connection ID of the candidate wavelength. When an Intermediate node receives the Setup OBGMP Message, it checks the Connection ID and searches through its wavelength table for the Connection ID. If a match is found, then a request is made to the SCS to configure the switch. The SCS confirms the configuration once it is completed and the *Setup Complete* field in the Wavelength Table is set to *TRUE*. The Intermediate node then forwards the message to the following node in the *AS_PATH*.

Reservation/Setup Phase: The 2PC mode eliminates the separate Setup and Reservation Phases of 4PC. The Reservation/Setup Phase follows the common Discovery Phase and is initiated by the destination, *D*. Once a candidate wavelength is selected by the Destination, a new OBGp Message is created with the Phase ID set to Reservation/Setup and the candidate wavelength information is placed in the λ RI of the Reservation/Setup OBGp Message. This message is propagated back to the Source, *S*, along the *AS_PATH*. When an Intermediate node receives the Reservation/Setup OBGp Message, the node performs both the Reservation and Setup Phases of 4PC. First, the wavelength is found in the Wavelength Table and checked for its availability. If available, then the wavelength is reserved. A request is made to the SCS, requesting that the optical switch be configured for the specific connection. Once the switch is completed, then the Setup Complete field is updated in the Wavelength Table and the message is forwarded along the *AS_PATH* to the Source node. When the Reservation/Setup OBGp Message arrives at the Source node, the lightpath is established and confirms that there is a successful setup. Traffic can then be sent over the newly established lightpath.

Confirmation Phase: In 4PC, when the Setup Phase is completed, the lightpath is ready for traffic to be sent from the Source to the Destination, but until the Source receives some kind of acknowledgement, it cannot start sending data over the lightpath. Thus, the destination node *D* needs to send a Confirmation OBGp Message back to the source along the *AS_PATH*. A Confirmation OBGp Message is an OBGp message with its phase field set to Confirmation. There are two types of Confirmation Messages: Lightpath Setup and Lightpath Teardown. To distinguish between the two, the sub-code

field of the message is set to Setup or Teardown. Once the Source Node receives the Confirmation OBGP Message with the sub-code set to Setup, then it can proceed to send data along the lightpath.

4.3.2 Aggressive Reservation

OBGP also has an optional mode of Four-Phase Aggressive (4PA) Reservation. Aggressive Reservation differs from Conservative Reservation in that all available wavelengths are reserved before actually being used. The phases in 4PA differ slightly from those found in 2PC and 4PC, which is shown in Figure 4-6.

Figure 4-6 - Illustration of the OBGP Aggressive Signalling Algorithm

Aggressive Discovery Phase: If the Source Node uses the 4PA algorithm for signalling a lightpath to be setup, an Aggressive Discovery OBGP Message is created with the Phase ID set to Aggressive Discovery and is propagated along the *AS_PATH* towards the destination, as a Conservative Discovery OBGP Message would propagate.

The primary difference between the Aggressive Discovery and the Conservative Discovery Message is at the Intermediate Nodes. Upon receiving an Aggressive Discovery OBGp Message, the Intermediate Node will collect the common existing wavelengths, regardless of their availability. At the Destination Node, D , the Aggressive Discovery OBGp Message will contain all common wavelengths between the source and the destination in its λRI .

An Error would occur if there are no common wavelengths that exist from the source to the destination.

Aggressive-Reservation Phase: The purpose of the Aggressive-Reservation Phase is to reserve all available wavelengths that are common between the source and the destination along the AS_PATH . Thus, the Destination Node, D , creates an Aggressive-Reservation OBGp Message with the λRI that was discovered through the Aggressive-Discovery Phase. This λRI is a set of common wavelengths along the AS_PATH , meaning that they are all candidate wavelengths to establish a lightpath. As the Aggressive-Reservation OBGp Message propagates to the Source Node, S , each Intermediate Node reserves all the wavelengths in its wavelength table that are unused for a connection or unreserved. This means that the Connection ID field in its wavelength table is set to NIL. The λRI is updated to reflect all common wavelengths which have been reserved along the path. Upon reaching the Source Node, S , a wavelength can be selected from λRI , which now contains wavelengths which have been reserved along the path. Any of these wavelengths is a valid candidate wavelength.

If, along the path, there are no common wavelengths that can be reserved, an Error Message is sent to the Source.

Setup-Unreserve Phase: The unused resources now must be relinquished and each Intermediate Node must configure its optical switch for the selected wavelength. The Source Node, S, creates a Setup-Unreserve OBG Message with the Connection ID and AS_PATH in the Attributes, and the λ RI which contains the selected wavelength. This message is propagated to the Destination along the AS_PATH. When an Intermediate Node receives the Setup-Unreserve OBG Message, it unreserves any entry in its wavelength table which contains the Connection ID, except for that of the selected wavelength, found in the λ RI. Then it proceeds to request to the SCS to configure the switch. When the switch is complete, then the Setup-Unreserve OBG Message can be propagated along the AS_PATH towards the Destination.

Confirmation Phase: As the Setup-Unreserve OBG Message reaches the Destination, the lightpath is completed, but a notification must be sent back to the Source indicating that traffic can now be sent. This Confirmation OBG Message has the Phase ID set as Confirmation, with the sub-code as Setup, similar to the Conservative modes.

A Two-Phase Aggressive Backward Reservation scheme cannot exist, since the wavelengths that are reserved must be released after the Aggressive Reservation/Setup Phase. To release the wavelengths, another two phases are required, and thus, 2PA will break down into a four-phase scheme when the OXC Setup Time is taken into consideration.

4.3.3 Teardown and Error

Dynamic allocation of lightpaths requires that the lightpaths be torn down after the lightpath is no longer needed, or if an error occurs.

Teardown Phase: Teardown is always a two-phase process for 4PA, 2PC and 4PC. The source node requests that the lightpath be torn down by creating a Teardown OBGp Message. That message will contain the Connection ID and *AS_PATH* of the given connection and is sent towards the destination. When a node receives the Teardown OBGp Message, it signals to the SCS to release the given connection and updates the Wavelength Table by finding the given Connection ID and setting it to *NIL*, to indicate that the wavelength is now available. Once the Teardown OBGp Message reaches the Destination, the lightpath has been torn down and a Confirmation OBGp Message is sent back to the Source. This Confirmation OBGp Message has a sub-code ID of Teardown.

Error Phase: Whenever an error is detected, an Error OBGp Message is sent back to the Source Node. An Error OBGp Message is an OBGp message with its Phase ID set to Error and the sub-code field is set to the phase where the error was detected. The source node in a lightpath segment is considered to be the master and is the one that decides what action should be taken, given the various error conditions. For example, if the Source is notified that an error occurred in the Setup Phase, then it should immediately send a Teardown OBGp Message to release the resources. Otherwise, the wavelength that was partially set up will appear to be in use and will not be available for other connections.

Sources of Error can be as follows:

- No available wavelength is found for setting up a lightpath in the Conservative-Discovery Phase.

- Various competition cases where two lightpath setup requests both discover that a certain wavelength is available, but one of them reserves it first. The “winner” will reserve the wavelength, while the “loser” assumes that the wavelength is still available and makes an attempt to reserve it, failing, since the status has changed. These errors are detected in the Reservation Phase and the Reservation/Setup Phase.
- Faults found by the SCS can also cause errors. The source of these errors is hardware faults. These types of errors are detected in the Setup and Reservation/Setup Phases, when the SCS tries to configure the OXC.

4.4 The OBG Router

As each message arrives to an OBG router, the border gateway router, the SCS and the OXC must work together to fulfill a lightpath request. Figure 4-1 shows the block diagram of a typical OBG Router and is similar to the SRSO Architecture [Greenberg+ 00]. The Border Gateway Router contains the OBG Process, the Wavelength Table and an incoming message queue. As a message arrives into the Border Gateway Router, it is enqueued, until the OBG Process is free to process the message. Depending on the message type, the OBG Process will act as described in Section 4.3. If the message is a Setup or Reservation/Setup Message, then a SCS switch request is generated and sent to the SCS block, which then enqueues the message into the SCS Service Queue. The SCS Service Queue is serviced in a FIFO manner and can be infinitely large. The SCS Block manages all the activity of the OXC, since it is a “dumb” device. By managing the current connections, the SCS waits until completion of an OXC switch before processing the next request in the SCS Service Queue.

It should be noted that if the OXC Switch Time is zero, then the OBG Process can bypass the SCS Queue.

5 Simulation Results and Performance Evaluation

5.1 Introduction

In this chapter, the performance of the OBGp architecture is compared and analyzed under different traffic loads. The effect of the OXC time is analyzed under varying OXC Switch Times and compared under a standard scenario. This chapter is organized as follows: A brief description of the simulation model is found in the following section. Simulation results then follow, breaking them down by the different scenarios.

5.2 Simulation Model and Scenario

The OBGp protocol was implemented using a network simulation tool, called OPNET Modeler. OPNET Modeler was chosen because it is a discrete event simulation tool, that realistically models networking protocols such as the TCP/IP Protocol stack, Ethernet, and BGP Routing. By adding OBGp extensions to the OPNET BGP model, the simulation network models a small interdomain network. Appendix B has a detailed description of the simulation model and the configuration of each node in the network.

A standard scenario for testing the simulation model was created and is called the *competition scenario* [Pezoulas 03b]. Figure 5-1 illustrates the simulation network configuration. Each node on the network represents an Autonomous System, with the internal networks removed. The nodes have been identified by their AS number. Table 5-1 shows the labels that will be given to the S-D Pairs, to easily distinguish the three

lightpaths to be established within the network. The list of Intermediate nodes the path traverses is found in the third column of Table 5-1.

Figure 5-1 - Simulation Network in OPNET

Label	S-D Pair	Intermediate Nodes
<i>Path A</i>	AS 1000 to AS 1002	AS 2000, AS 3000, AS 4000
<i>Path B</i>	AS 1005 to AS 1006	AS 2000, AS 3000
<i>Path C</i>	AS 1001 to AS 1007	AS 3000, AS 4000

Table 5-1 - Source-Destination Pair Labels

This scenario was chosen to illustrate competition within the same network. *Path A*'s lightpath requests span links that are shared by *Path B* and *Path C*. The two interfering paths have shorter hop lightpaths, and thus should interfere with the lightpath requests of *Path A*. The overall effect shows that *Path A*'s requests are blocked more than those of *Path B* and *Path C*.

The following are the parameters for the simulation:

- (i) Sources generate requests using a Poisson distribution with varying rates of requests per second.

- (ii) If a lightpath is established, then it is held with an exponential distribution, with a mean time of 1 second.
- (iii) The propagation delay on each link of the control network is fixed to 1ms.
- (iv) The number of wavelengths maintained by each link is set to 30 wavelengths.
- (v) The Optical Cross Connect Switch Time is varied from 0 to 100 ms and is only applied to the Intermediate AS's, which are 2000, 3000 and 4000.
- (vi) The wavelength selection algorithm was changed from First-Fit to Random at the destination node.
- (vii) The SCS serves all requests in a FIFO order and maintains an infinite queue.
- (viii) Each AS is represented by a Border Gateway Router, with one to four IP layer interfaces.
- (ix) OBGp is enabled on each router and route determination is made by the shortest path policy.
- (x) OBGp operates under 2PC, 4PC and 4PA modes initiated at the source node.

5.3 Simulation Results

The different metrics used to measure performance of the system are the lightpath setup round-trip time and the percentage of blocked requests. The blocking probability is a standard metric for signalling systems, as it measures the system utilisation. The lightpath setup time is measured since the Optical Cross Connect Switch Time adversely affects this value. Analysis was conducted in scenarios by varying the request rate and OXC Switch Time. The two following subsections describe the results of the network under the First-Fit and Random Wavelength Selection Algorithms combined with the Conservative signalling algorithm. The third subsection shows that the Aggressive

scheme underperforms in comparison to the Conservative schemes when combined with both First-Fit and Random Wavelength Selection Algorithms. Lastly, a brief description of the system performance is described when the network has fewer wavelengths.

5.3.1 Conservative-First-Fit Wavelength Selection Algorithm

The first scenario tested uses the First-Fit Wavelength Selection Algorithm at the destination nodes, coupled with the Conservative scheme. Under this scenario, the call rate was varied with the OXC Switch Time under the 2PC and 4PC signalling schemes.

Figure 5-2 (a) and (b) show the average percentage of blocked calls for all paths as the request rate is increased by factors of 10, starting at 0.01 calls per second for the 2PC scheme. The graph is shown with a 95% confidence level, but the intervals, in this case, are too small to be noticeable. As the rate increases from 0.1 to 1 call per second, the blocking probability increases dramatically. With the 95% confidence interval, the blocking percentage does not increase as the OXC time is increased. *This shows that the OXC Switch Time does not make a significant contribution to the blocking of the lightpath requests at low request rates.*

It is also interesting to note that between the 4PC and 2PC schemes, there is very little difference in the performance under the overall blocking probability.

(a) 2PC Average Blocked Calls with 95% Confidence

(b) 4PC Average Blocked Calls with 95% Confidence

Figure 5-2 - Average Percentage of Blocked Calls with First-Fit Wavelength Selection Algorithm

Figure 5-3 shows the blocking summary for the Discovery and Reservation Phases at each node for Path A, with a 95% Confidence Level. This value is not the average blocking, but the number of blocked messages over the total number of messages passing through that node. 4PC and 2PC schemes behave exactly in the same way. The majority of failures occur at the destination during the Discovery Phase. Reservation Messages are blocked at AS 4000 and AS 3000 and there are no messages blocked at AS 2000.

As the majority of messages are eliminated by the destination (for *Path A*'s case, AS 1002), fewer requests compete for resources during the Reservation Phase. At AS 4000 a large number of messages flow to two destinations, AS 1007 and AS 1002. More than 35% of those messages are eliminated at the destinations and of the remaining messages, another 10% fail at AS 4000 (for sources 1000 and 1004). This means that a smaller number of messages flow through AS 3000 for *Path A* and *C*'s requests.

Figure 5-3 - Nodal Blocking Summary at 1 call per second with 95% Confidence

Figure 5-4 - Average Lightpath Setup Time with 95% Confidence

Figure 5-4 shows the average setup time for the establishment of a lightpath. The lightpath setup time was recorded from the time that the request was generated, until the time that the Confirmation Phase message returns to the source, for both 2PC and 4PC. Failed requests are not included in this measure.

2PC lightpath setup takes less time than the 4PC lightpath setup, the difference being roughly one round-trip time. In 4PC this is significant, because it takes two round-trip times to complete the lightpath reservation (excluding the queueing delay). If a request is successful through the Reservation phase, the round trip time is $13.106 \text{ ms} \leq t \leq 13.223 \text{ ms}$ with an average of 13.175 ms, at a 95% confidence level. It is understood that 4PC takes more time to completely establish a lightpath, but one must remember that after the first round-trip time, the wavelength is already guaranteed to be reserved.

(a) 2PC Average Wavelength Utilization with 95% confidence

(b) 4PC Average Wavelength Utilization with 95% confidence

Figure 5-5 - Average Wavelength Utilization per link at 1 call per second with First-Fit Wavelength Selection Algorithm

Another measure of performance of the network is the wavelength utilization. Figure 5-5 (a) and (b) show the average wavelength utilization at 1 call per second of each Path with a 95% confidence interval. Each link maintains 30 wavelengths. There are more wavelengths used at the destination than at the source for all paths. The reasoning for this is that lightpaths that fail during the Reservation or Reservation/Setup phase have reserved the wavelengths at the destination. Since they are reserved and the failure occurs during the Reservation or Reservation/Setup phase at some Intermediate node, the Source is never reached because the failure is caused by a lack of resources. Thus, the wavelength reservation along the path is torn down and the source never has the wavelengths reserved. This implies that the wavelengths are held at the destination and causes the Discovery messages to fail for contention of the shared resources.

The discoveries about the Conservative-First-Fit scenario are listed as follows:

1. The OXC has no large impact on the blocking probability of interfering paths at low request rates.
2. The 4PC scheme behaves similarly to the 2PC scheme with respect to blocking probability.
3. The majority of the messages are blocked at Discovery Phase only at the destination. It is also important to note that there are no failed Discovery Messages caused by competition of resources along the shared links.
4. Lightpath setup time requires more time as the OXC time is increased. 4PC requires a longer time than 2PC when setting up the wavelength.
5. Wavelength utilization at the destination is higher than the wavelength utilization at the source.

5.3.2 Conservative-Random Wavelength Selection Algorithm

The 2PC and 4PC signalling schemes, coupled with the First-Fit Wavelength Selection Algorithm yielded interesting results. It is known that the Random Wavelength Selection Algorithm outperforms the First-Fit Wavelength Selection Algorithm with respect to overall blocking [Pezoulas+ 03]. In this section, we study the performance of the 2PC and 4PC signalling schemes coupled with the Random Wavelength Selection Algorithm.

Figure 5-6 shows the overall average percentage of messages blocked under the Random Wavelength Selection Algorithm for the 2PC scheme. At lower request rates, below 2 calls per second, the three curves are equal and show no significant difference as the OXC Switch Time is increased. As the rates increase above 2 calls per second, the percentage blocking of 10 ms is slightly higher than that of 0 ms. After 2 calls per second, the percentage blocked at 100 ms increases at a steady rate, much higher than that of the lower OXC Switch Times.

Figure 5-6 - 2PC Average Percentage of Blocked Calls using Random Wavelength Selection Algorithm with 95% Confidence

Figure 5-7 - 4PC Average Percentage of Blocked Calls using Random Wavelength Selection Algorithm with 95% Confidence

Figure 5-7 illustrates the same scenario under the 4PC algorithm. The interesting part is that the 4PC algorithm shows identical behaviour to the 2PC scenario. While the difference between the blocking percentage when the OXC Switch Time is 0 ms and 10 ms is small, the blocking increases sharply when the OXC Switch Time is 100 ms after 2 calls per second. Both Figure 5-6 and Figure 5-7 show the same average number of blocked calls, with the 95% confidence interval.

It should be pointed out that despite the very similar behaviour between both scenarios with respect to blocking probability, both schemes do in fact behave differently. The overall blocking average is an aggregated metric of all requests from all paths. By analyzing a single request path, their differences are exposed. Figure 5-8 (a) and (b) trace the percentage of blocked messages at each node for *Path A* at 5 calls per second. The values shown at each node show the ratio of messages blocked at each node versus the number of messages that arrive at the node. The horizontal axis is placed in such a way that the nodes illustrate a round trip for the lightpath setup. Thus, they are divided in the Discovery Phase and the Reservation Phase. For 2PC (Figure 5-8(a)), the Reservation Phase represents the Reservation/Setup Phase.

(a) 2PC at 5 calls per second with 95% Confidence

(b) 4PC at 5 calls per second with 95% Confidence

Figure 5-8- Nodal Blocking Summary at 5 calls per second using Random Wavelength Selection Algorithm

(a) 2PC at 5 calls per second with 95% Confidence

(b) 4PC at 5 calls per second with 95% Confidence

Figure 5-9 - Percentage of blocked Discovery Phase at 5 calls per second

At the destination, AS 1002, 4PC blocks a larger number of Discovery messages than 2PC does. This causes the Reservation Phase blocking in 4PC (Figure 5-8 (b)) to be less than the Reservation/Setup phase blocking in 2PC (Figure 5-8 (a)), specifically at AS 4000 and AS 3000. AS 2000 does not suffer any Reservation or Reservation/Setup Phase blocking, because there is very little competition in that stage between Path A and Path B's lightpath requests.

Figure 5-9 (a) and (b) illustrate the Discovery Phase blocking that cannot be seen in Figure 5-8, for 2PC and 4PC scenarios respectively. The Y-Axis scales are different by a factor of 10. Notice that as the OXC Switch Time is increased, more Discovery messages are blocked at AS 2000 and 3000. On average, the 4PC scheme blocks more Discovery messages at each AS than 2PC does. When the OXC Switch Time is increased to 100 ms, the difference is more significant between 2PC and 4PC. 4PC blocks up to 40% more Discovery messages than the 2PC scheme under the same scenario.

What this suggests is that the 4PC signalling algorithm blocks more requests during the Discovery phase than the 2PC signalling algorithm. *By detecting that the lightpath cannot be completed earlier, the source is notified sooner and fewer resources are reserved.* The other benefit is that by detecting a failure earlier, 4PC's resources are left free and fewer Reservation Phase requests fail. 4PC reacts faster, since in a single round trip, wavelengths are reserved and guaranteed to be used for a specific connection. By blocking messages earlier, 4PC has a surprising effect on the lightpath setup time.

Figure 5-10 - Average Lightpath Setup Time with 95% Confidence

Figure 5-10 shows the average time it takes to complete the creation of a lightpath. It is interesting to note that 4PC acts similarly to 2PC for all switch times below 2 calls per second. At 2 calls per second and above, there is a divergence between 4PC and 2PC when the OXC Switch Time is 10ms and 100 ms. The 100ms scenario is interesting, because the improvement of 4PC is as large as 3 seconds. 4PC under the 10 ms scenario diverges from the 2PC switch time up to a maximum of 200 ms. The 10 ms scenario is interesting because the lightpath setup time requires more time at 2 calls per second than it does at 5 to 10 calls per second.

The discovery of 4PC taking a smaller amount of time than 2PC to establish a lightpath goes against the common belief that it should take longer than the lightpath setup of 2PC.

To understand why 4PC signalling requires less time than the 2PC signalling, a closer examination of the SCS Request Queue must be taken. As messages are placed

into the SCS request queue, the Setup requests wait at that node until the request is serviced. Figure 5-11 shows the average SCS queue size at the intermediate nodes for 2PC and 4PC under OXC Switch Times 10 ms and 100 ms. The case with no switch time (0 ms) is excluded since the SCS request queue is always empty. The call request rate was at 5 calls per second.

Figure 5-11 - Average SCS Request Queue Size with 95% Confidence

Most messages are queued at AS 3000 for both 2PC and 4PC. This is obvious, since all three paths flow through AS 3000 and require that the requests be queued there. Attention should be drawn to the fact that under the 2PC scheme, the queue at AS 2000 is relatively small, but is large at AS 4000. The opposite can be found of the 4PC scheme, where the queue at AS 4000 is small and is quite large at AS 2000. The reasoning behind having opposite queue sizes is that in the 2PC scheme, the messages are queued during the Reservation/Setup Phase. In this phase, the messages travel from the destination back

to the source. The reverse can be said about the 4PC scheme. Under that signalling scheme, messages are queued during the Setup Phase. Messages travel from the source to the destination and the wavelength is already reserved. The reason why the queue is large is that all messages are aggregated and require being set up. When nearing the source (or destination in the opposite case), fewer messages are rejected.

In Figure 5-8, it was observed that AS 2000 did not incur any blocking in both 2PC and 4PC schemes. For 2PC, a large number of Reservation/Setup Phase messages are eliminated at AS's 3000 and 4000. This results in fewer requests traveling to AS 2000 and thus, the SCS request queue at AS 2000 is smaller. Under the 4PC algorithm, there are no Setup phase messages that can be blocked since the wavelength is already reserved in the Reservation Phase and thus, even if the SCS queue size is large at AS 2000, there are no blocked messages.

Early detection of failed lightpath requests is the primary source of the smaller queue sizes in 4PC signalling. Since there are fewer wavelengths reserved, there will be fewer Setup Phase messages traveling and being queued at each node. Since in 2PC signalling, the messages are blocked during the Reservation Phase meaning that the lightpath reservation gets further along the path before failing. This increases the signalling traffic along the path and increases the SCS Queue size, which then causes a longer wait at each node resulting in a longer setup time.

(a) 2PC Wavelength Utilization with 95% confidence

(b) 4PC Average Wavelength Utilization with 95% confidence

Figure 5-12 - Average Wavelength Utilization per link at 5 calls per second with Random Wavelength Selection Algorithm

Figure 5-12 (a) and (b) show the average wavelength utilization of each S-D pair for all three paths, when there are 5 calls per second. For each node, the OXC time is varied. Figure 5-12 also shows that as you increase the OXC Switch Time from 0 to 10 ms, there is very little change in the behaviour for this scenario. The source wavelength utilization is lower at the source than the destination's utilization. This verifies that more wavelengths are utilized at the destination. If more wavelengths are held at the destinations, then Discovery Messages have a higher possibility of being blocked because of a lack of resources. The 4PC algorithm holds the wavelengths longer than those in 2PC. *Path B* and *Path C* share a higher utilization in 4PC than 2PC, but *Path A* has a lower utilization at the destination, because requests are blocked during the Discovery Phase.

Note that *Path A* has the lowest wavelength utilization, at both the source and the destination. The reasoning behind *Path A*'s low utilization is the fact that it has the highest blocking percentage, caused by the interfering path. In both scenarios, *Path C* has the same average utilization at the destination but their source utilization differs by a full wavelength, on average. Note that in 4PC, *Path B* has a higher destination usage than *Path C*, but the source usage is equal.

What was found in the Conservative-Random scenario is as follows:

1. Using the Random Wavelength Selection Algorithm, a lower blocking probability can be achieved and high request rates can be achieved.
2. The OXC has no large impact on the blocking probability of interfering paths at lower request rates. When the OXC Switch Time is 100 ms and the request rate is at 2 calls per second and above, there is a sharp increase of blocked calls.

3. The 4PC algorithm holds wavelengths longer at the destination, thus blocking more requests earlier.
4. At higher rates, 4PC has the ability of setting up lightpaths faster than 2PC, because the SCS Request Queue size is smaller.

5.3.3 Aggressive Scenario Results

The 4PA results were compared with the 2PC and 4PC scenarios under the same parameters. It was found that when the call rate was 1 call per second, the blocking probability of 4PA was much higher than that of 4PC or 2PC. Figure 5-13 shows this result.

In Figure 5-13, we can see that the results of 4PA are much higher than those of 4PC and 2PC. This result is not surprising, since it follows the findings in previous literature [Yuan+ 96].

Figure 5-13 - Average Percentage of Blocked Messages under different algorithms with 95% Confidence

5.3.4 Varying the Number of Wavelengths per Link

The same simulations were run for scenarios where the links had 10 wavelengths. It should be noted that having fewer wavelengths per link increases the overall blocking probability, but does not change the final conclusions of the system by any means, as described in Sections 5.3.1, 5.3.2, 5.3.3.

It was also discovered that the same nodal behaviour is found with the exception of one case. Under the Random wavelength selection algorithm, the lightpath request required less time than its counterpart where 30 wavelengths are available per link. Under the First-Fit wavelength selection algorithm, 2PC and 4PC have the same number of requests blocked and the same lightpath setup time when there are fewer wavelengths.

6 Conclusions and Recommendations for Further Study

6.1 Conclusion

6.1.1 Interdomain Control Protocol in All-Optical Networks

Carrier networks have been evolving to meet the current demands of the users. As Internet usage increases, the accepted core technology will migrate from a static fibre optic network, to a dynamic all-optical WDM network that is inherently circuit-switched, rather than packet-switched. Certain considerations must be made to integrate all-optical technologies into the existing Internet framework.

As individual domains maintain their all-optical networks, the methodology for connecting with other domains and establishing end-to-end lightpaths becomes a challenge. Intradomain networks can freely share optical information, but cannot in the interdomain. This thesis has proposed a network protocol called the Optical Border Gateway Protocol (OBGP) that marries the physical all-optical network to the Internet, allowing standard routing capabilities and coupling them with methods to establish end-to-end lightpaths in the interdomain. OBGP is an extension to BGP, which is the de-facto standard for interdomain routing in the Internet.

We briefly summarize the contributions made from the creation of OBGP, addressing the requirements stated in Section 4.2 :

- Successfully combined flexible signalling algorithms (4PC, 2PC and 4PA) to a routing protocol (BGP) in OBGP.
- Leveraged existing properties of BGP to determine fixed optical routes for lightpaths.

- Created the Four-Phase Conservative and Four-Phase Aggressive Signalling Schemes. It was found that Four-Phase Aggressive underperformed compared to the Conservative counterparts.
- Introduced the OBGW Wavelength Table (λ -table). This table maintains current wavelength and connection information to the neighbouring OBGW nodes.
- Introduced a new BGP Message Type, called the OBGW Message. This message is used for lightpath signalling.
- Added the SCS Message Queue, which served OXC Switch Requests in a FIFO Order. The SCS Message Queue is essential to the performance of an All-Optical Network, as the OXC Switch Time is increased.

The purpose of simulating OBGW was to prove that BGP can be extended to consider the optical network and be used as an interdomain control protocol. It was shown through simulations that OBGW can be considered as a viable control domain for optical networks in the interdomain. OBGW's extensions are simple and can be integrated into BGP through the BGP Route Capabilities feature [RFC 2842]. Since it is based on BGP, which is widely deployed within the entire Internet, OBGW can be introduced as the network migrates from a static network topology to a dynamic all-optical network. The benefit of having OBGW as the primary interdomain control protocol is to enable edge customers control and manage their wavelengths.

6.1.2 OBGW Signalling and OXC Switch Times

OBGW introduces three schemes to establish end-to-end lightpaths: Two-Phase Conservative, Four-Phase Conservative and Four-Phase Aggressive, all of which use a

backward reservation method. In this iteration of OBG, the signalling algorithms and the routing algorithm are tightly coupled, similar to the VPN approach [RFC 2547].

These schemes were then used to investigate the effect of the OXC Switch Time on blocking probability. Since the OXC requires a fixed time for the device to reconfigure the switch fabric and stabilize the signal, the transmission of the signal cannot be guaranteed until all OXCs along the path have completed their switch. It was discovered that the OXC Switch Time has a unique effect on the performance of establishing lightpaths in all-optical networks. As competing request paths were introduced the lightpath setup time increased for both 2PC and 4PC algorithms.

The following results and conclusions were found through simulations conducted in Chapter 5:

- The OXC Switch Time had an effect on the average blocking probability and the lightpath setup time at higher request rates.
- Blocked requests occurred primarily at the destination under the Discovery Phase, for both 2PC and 4PC at all request rates.
- Under the First-Fit Wavelength Selection Algorithm, 2PC and 4PC behave similarly, except 4PC required 1 round trip time longer to establish a lightpath, due to the two extra phases.
- Under the Random Wavelength Selection Algorithm, the 4PC algorithm required less time to establish a lightpath at higher rates than the 2PC scheme. This is a direct result of 4PC blocking requests earlier, removing possibly failed lightpath requests in the SCS Queue.

- The 2PC algorithm blocked more requests than the 4PC algorithm at Intermediate nodes, during the Reservation/Setup Phase. This resulted in a larger SCS Queue Size and a dramatic increase in the average lightpath setup time at higher request rates.
- More wavelengths are utilized at the destination under the 4PC algorithm than the 2PC algorithm.
- The 4PC algorithm blocked requests earlier, notifying users of a failed lightpath request.

In previous literature, the OXC Switch Time was never considered in the calculations of blocking probability. This thesis provides results that the blocking probability is affected by the increasing OXC Switch Time, especially at higher request rates. Even though it is unnatural to run a network at a high request rate with lower wavelength utilization, this behaviour is important as it demonstrates that the OXC Switch Time impacts the blocking probability as it is increased. This research can be used as an indicator that if the switching time inside a domain exceeds 10 ms, then there is a higher probability that the success of a request is lowered, due to out of date state information.

6.2 Recommendations for Further Study

OBGP has been shown as a valid method for establishing lightpaths across multiple domain networks, but further research must be conducted to ensure the validity of the protocol under all scenarios. Several considerations must be made into the actual protocol such as issues of link or path protection, intradomain and interdomain protocol

communication, pricing and billing features, maintenance of the network and security issues related to the requests.

Routing of the wavelength information is a primary role of a control protocol that hasn't been addressed thus far. In this revision of OBGp, a node maintains a static wavelength availability table between itself and its neighbouring nodes. Each node knows if a wavelength *exists*, but the shared information does not show whether or not the wavelength is *available*. The difference is subtle, but must be duly noted. If a wavelength exists, then it can either be available or unavailable. This allows a source node to decide whether or not it can establish a lightpath to a destination. If a wavelength is available, then the source node knows it can establish a clear path to the destination.

The difficulty that arises in routing is the ability to share this information to neighbouring networks and beyond. This information can cause an overabundance of routing information constantly travelling between nodes in the network. This involves a large overhead, but it would be beneficial to determine the correct amount of information to pass between nodes about other areas of the network. The following describes some of the considerations in shared routing information:

1. BGP and OBGp routing tables must maintain a database of the status of all wavelengths in the network between each link. Information must be aggregated in the system since the entire network's information cannot be processed.
2. UPDATE messages need to be modified to contain wavelength information associated with the AS_PATH.

3. Wavelength information must be aggregated to minimize control message size.
4. UPDATE message timers must be implemented to regulate the information sharing. This will be sensitive to the requests being made at the edge since the status changes affect signalling nodes.

It would be compelling to study the relationship between the frequency of the routing UPDATE messages and the request rates, as well as the effect on overall blocking probability in the interdomain environment. Studies such as [Ho+ 02] show that these methods work in a controlled environment, but cannot be scaled to the interdomain network since the shared information will flood the network.

The research presented shows results where the OXC Switch Time affects the blocking probability and the lightpath setup time in one specific competition scenario. While this scenario illustrates a common competition case within a network, other scenarios with different levels of complexity were not tested. The effect of the OXC Switch Time should be the same under those conditions, but cannot be guaranteed when several competing paths are introduced.

Test networks such as NSFNet, hypercube or mesh-torus networks with multiple competing connections are ideal for exploring the effect of OXC Switch Time. The differences between 4PC and 2PC can be further explored and considered under more stressful situations, wherein the competition is high with shorter and longer lightpaths.

For all scenarios, the propagation time, d_{prop} , was fixed to 1 ms suggesting an equal distance to each node. It would be interesting to see how varying the propagation time affects the performance under a backward reservation protocol. I would suggest that

both 4PC and 2PC will behave similarly at lower request rates, but as the request rates increase, the performance of 4PC will be much better than that of 2PC for both Random and First-Fit wavelength selection algorithms.

Finally, it would be interesting to compare OBGP's signalling approaches with the existing approaches presented by GMPLS. Since the OXC Switch Time study is applicable to any signalling protocol, it would be useful to compare the 2PC and 4PC method with the signalling method presented in GMPLS under similar conditions.

7 References

- [Banerjee+ 00] D. Banerjee, B. Mukherjee, "Wavelength-routed optical networks: Linear Formulation, Resource Budgeting Tradeoffs, and a Reconfiguration Study", IEEE/ACM Transactions on Networking, Vol. 8, No. 5, Oct. 2000, pp. 598-607.
- [Banerjee+ 01] A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, Y. Rekhter, "Generalized Multiprotocol Label Switching: An Overview of Signalling Enhancements and Recovery Techniques", IEEE Communications Magazine, Vol. 39, No. 7, July 2001, pp. 144-151.
- [Barry+ 95] R.A. Barry, P. Humblet, "Models of Blocking Probability in All-Optical Networks with and without Wavelength Changers", Proceedings of Infocom '95, Boston, MA, Vol. 2, April 1995, pp. 402-401.
- [Bernstein+ 02] G. Bernstein et al., "Optical Inter Domain Routing Considerations", Internet Draft, Network Working Group, draft-ietf-ipo-optical-interdomain-02.txt.
- [Birman 95] A. Birman, "Computing Blocking Probabilities for a Class of All-Optical Networks", Proceedings of Infocom '95, Boston, MA, Vol. 2, April 1995, pp. 651-658.
- [Chlamtac+ 92] I. Chlamtac, A. Ganz, G. Karmi, "Lightpath Communications: An Approach to High Bandwidth Optical WAN's", IEEE Transactions on Communications, Vol. 40, No. 7, July 1992, pp. 1171-1182.

- [Chu+ 02] P.B. Chu, S. Lee, S. Park, "MEMS: The Path to Large Optical Crossconnects", IEEE Communications Magazine, Vol. 40, No.3, March 2002, pp. 80-87.
- [Foster+ 01] I. Foster, C. Kesselman, S. Teucke, "The Anatomy of the Grid", International Journal of Supercomputer Applications, Vol. 15, No. 13, 2001.
- [Francisco+ 01] M.J. Francisco, L. Pezoulas, C. Huang, I. Lambadaris, W. St.Arnaud, "End-to-End Signalling and Routing for Optical Networks", The Proceedings of the SPIE, OptiComm 2001, Denver Colorado, Vol. 4599, Aug. 2001, pp. 120-129.
- [Francisco+ 02] M.J. Francisco, L. Pezoulas, C. Huang, I. Lambadaris, "End-to-End Signalling and Routing for Optical IP Networks", Proceedings of ICC 2002, New York, NY, Vol. 2, April 2002, pp. 1468-1473.
- [Greenberg+ 00] A. Greenberg, G. Hjalmtysson, J. Yates, "Smart Routers – Simple Optics: A Network Architecture for IP over WDM", Optical Fibre Conference, 2000.
- [Harai+ 97] H. Harai, M. Murata, H. Miyahara, "Performance of Alternate Routing Methods in All-Optical Switching Networks", Proceedings of IEEE Infocom '97, Kobe, Japan, April 1997, pp. 516-524.
- [Ho+ 02] P. Ho, H.T. Mouftah, "A novel distributed control protocol in dynamic wavelength-routed optical networks", IEEE

- Communications Magazine, Vol. 40, No. 11, Nov. 2002, pp. 38-45.
- [Labourdette+ 03] J.F. Labourdette, Z. Zhang, "Role of Optical Network in Resilient IP Backbone Architecture", Optical Networks Magazine, Vol. 4, No. 5, Sept./Oct. 2003, pp. 124-128.
- [Li+ 02] G. Li, J. Yates, D. Wang, C. Kalmanek, "Control Plane Design for Reliable Optical Networks", IEEE/LEOS Summer Topical Meeting, Invited Paper, July 2002.
- [LiFrancisco+ 03] Y. Li, M.J. Francisco, I. Lambadaris, C. Huang, "Traffic Classification and Service in Wavelength-Routed All-Optical Networks", The Proceedings of ICC 2003, Anchorage, AK, Vol. 2, May 2003, pp. 1375-1380.
- [Ma+ 03] X. Ma, G.S. Kuo, "Optical Switching Technology Comparison: Optical MEMS vs. Other Technologies », IEEE Optical Communications supplement to IEEE Communications Magazine, Vol.41, No.11, Nov. 2003, pp. 50-57.
- [Mei+ 97] Y. Mei, C. Qiao, "Efficient Distributed Control Protocols for WDM All-Optical Networks", The Proceedings of the Sixth International Computer Communications and Networks, Las Vegas, NV, Sept. 1997, pp 150-153.
- [Mokhtar+ 98] A. Mokhtar, M. Azizoglu, "Adaptive Wavelength Routing in All-Optical Networks", IEEE/ACM Transactions on Networking, Vol. 6, No. 2, April 1998, pp. 197-206

- [Mukherjee 97] B. Mukherjee, "Optical Communication Networks", McGraw-Hill, New York, 1997.
- [Pezoulas 03b] L. Pezoulas, "End-to-End Signalling and Routing for Optical Networks", Master's Dissertation, Ottawa-Carleton Institute for Electrical and Computer Engineering, Carleton University, Ottawa 2003.
- [Pezoulas+ 03] L. Pezoulas, M.J. Francisco, I. Lambadaris, C. Huang, "Performance Analysis of a Backward Reservation Protocol in Networks with Sparse Wavelength Conversion", Proceedings of ICC 2003, Anchorage, AK, Vol. 2, May 2003, pp. 1468-1473.
- [Qiao+ 96] C. Qiao, Y. Mei, "Wavelength Reservation Under Distributed Control", IEEE/LEOS 1996 Summer Topical Meetings, Keystone, CO, Aug. 1996, pp. 45-46.
- [Rajagopalan+ 00] B. Rajagopalan, D. Pendarakis, D. Saha, R. Ramamurthy, K. Bala, "IP over Optical Networks: Architectural Aspects", IEEE Communications Magazine, Vol. 29, No. 9, Sept. 2000, pp. 94-102.
- [Rajagopalan+ 04] B. Rajagopalan, J. Luciani, D. Awduche, "IP over Optical Networks: A Framework", Internet Draft, IP over Optical Group, draft-ietf-ipo-framework-05.txt, March 2004.
- [Ramamurthy+ 98] R. Ramamurthy, B. Mukherjee, "Fixed-Alternate Routing and Wavelength Conversion in Wavelength Routed Optical Networks",

- Proceedings of Globecom '98, Australia, Nov. 1998, pp. 2295-2302.
- [Ramaswami+ 95] R. Ramaswami, K. Sivarajan, "Routing and Wavelength Assignment in All-Optical Networks", IEEE/ACM Transaction on Networking, Vol. 3, No. 5, Oct. 1995, pp. 489-500.
- [Ramaswami+ 97a] R. Ramaswami, A. Segall, "Distributed Control for Optical Networks", IEEE/ACM Transactions on Networking, Vol. 5, No. 6, Dec. 1997, pp. 936-943.
- [Ramaswami+ 97b] R. Ramaswami, G. Sasaki, "Multiwavelength Optical Networks with Limited Wavelength Conversion", IEEE/ACM Transactions on Networking, Vol. 6, No. 6, December 1997, pp 744-754.
- [Reardon 03] M. Reardon, "Kompella Backs BGP", Lightreading, http://www.lightreading.com/document.asp?doc_id=42559&print=true, Oct. 28, 2003.
- [RFC 1771] Y. Rekhter, T.Li "A Border Gateway Protocol 4" Internet Engineering Task Force, Network Working Group, Request for Comments 1771, March 1995.
- [RFC 2547] E. Rosen, Y. Rekhter, "BGP/MPLS VPNs", Internet Engineering Task Force, Network Working Group, Request for Comments 2547, March 1999.
- [RFC 2842] R. Chandra, J.Scudder, "Capabilities Advertisement with BGP-4" Internet Engineering Task Force, Network Working Group, Request for Comments 2842, May 2000.

- [RFC 3031] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture", Internet Engineering Task Force, Network Working Group, Request for Comments 3031, Jan. 2001.
- [RFC 3209] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", Internet Engineering Task Force, Network Working Group, Request for Comments 3209, Dec. 2001.
- [Rouskas 02] G. Rouskas, "A Tutorial on Optical Networks", Networking 2002 Tutorials, Lecture Notes in Computer Science, Vol. 2497, 2002, pp. 155-193.
- [Saha+ 03] D. Saha, B. Rajagopalan, G. Bernstein, "The Optical Network Control Plane: State of the Standards and Deployment," IEEE Optical Communications, Vol. 1, No. 3, August 2003, pp. S24 – S29
- [Sharma+ 97] R. Sharma, R.I. MacDonald, "A signal transparent 10/spl times/10 space-division opto-electronic switch core for virtual-transparent path-based multiwavelength networks", Lightwave Technology, Journal of, Vol. 15, No. 8, August 2003, pp. 1522 – 1529.
- [St.Arnaud+ 00] B. St.Arnaud, et al., "Optical BGP Networks" CA*Net Discussion Paper, March 20, 2000. Revised Draft July 16, 2000.
- [Subramaniam+ 97] S. Subramaniam, R.A. Barry, "Dynamic Wavelength Assignment in Fixed Routing WDM Networks", Proceedings of ICC '97, Montréal, Canada, Vol. 1, June 1997, pp. 406-410.

- [Yuan+ 96] X. Yuan, R. Gupta, R. Melhem, "Distributed Control in Optical WDM Networks", Proceedings of MILCOM '96, IEEE Conference on Military Communications, McLean, VA, Vol. 1, Oct. 1996. pp. 100-104.
- [Yuan+ 99] X.Yuan, Y.Mei, C. Qiao, "Distributed Control Protocols for Wavelength Reservation and Their Performance Evaluation", Photonic Networks and Communications, Vol. 1, No. 3, 1999, pp. 207-218.
- [Zang+ 00] H. Zang, J.P. Jue, B. Mukherjee, "A Review of Routing and Wavelength Assignment Approaches for Wavelength-Routed Optical WDM Networks," Optical Networks Magazine, Vol. 1, No. 1, Jan. 2000, pp. 47-60.
- [Zang+ 01] H. Zang, J.P. Jue, L. Sahasrabudhe, R. Ramamurthy, B. Mukherjee, "Dynamic Lightpath Establishment in Wavelength-Routed WDM Networks", IEEE Communications Magazine, Vol. 39, No. 9, Sept. 2001 pp. 100-108.
- [Zheng+ 01] J. Zheng, H.T. Mouftah, "Distributed Lightpath Control based on Destination Routing for Wavelength-Routed WDM Networks", Proceedings of GLOBECOM '01, IEEE Global Telecommunications Conference 2001, San Antonio, TX, Vol. 3, Nov. 2001, pp. 1526-1530.

A Appendix A

Border Gateway Protocol Version 4

A.1 Background

Basic routing in a network involves two main activities, first determining an optimal routing path and secondly, transporting packets through a network. The Border Gateway Protocol (BGP) [RFC 1771] addresses the task of path determination through connections with other BGP neighbours in an IP network in the interdomain. Similar to other routing protocols, such as OSPF, RIP and IS-IS, BGP maintains routing tables, transmits routing updates and uses metrics to make routing decisions. The main function of BGP is to exchange and maintain network reachability information between network domains. This consists of path information used to connect with other networks.

The goals of an interdomain routing protocol such as BGP are very different from those of an intradomain protocol. The issue of scalability has much more significance for an interdomain gateway router, which must be able to forward any packet destined anywhere in the world, versus a local router on a small subnet. In addition, the autonomous nature of network domains means that it is very difficult to calculate an optimal path that crosses several domains. As a result, interdomain routing will only advertise network reachability, the ability to reach other networks.

The communication of information regarding the internal structure of a domain is limited. Certain Autonomous Systems (ASs) may not want to share their own information to a neighbouring AS. Thus, traffic engineering metrics such as hop count,

available bandwidth, route diversity and many more, cannot be shared for routing purposes in the interdomain. This leaves the primary metric for route determination in the interdomain to be the reachability information. Reachability information, in the Border Gateway Protocol is defined by the *AS_PATH*, which is a path of Autonomous Systems from one AS to another. To share this information each AS will advertise to its neighbours its ability to reach another AS, based on a set of policies, which is described in a Service Level Agreement (SLA).

Each BGP router in a network maintains a set of Routing Information Bases (RIBs), which list all feasible paths to all known networks. The RIBs are updated upon initial data exchange and subsequently updated when new information is received from another BGP neighbour.

A.2 BGP Message Types

There are four types of messages specified for BGP: OPEN, UPDATE, NOTIFICATION, and KEEP-ALIVE. A BGP OPEN message opens a communications session between BGP neighbours and is the first BGP message sent by BGP speakers after a TCP connection is established.

The OPEN message opens a BGP communications session BGP neighbours and is the first message transmitted in a BGP session. Before proceeding with the session, an OPEN message must be confirmed with a KEEP-ALIVE message.

An UPDATE message is used to provide routing updates within a network. These messages allow BGP routers to maintain consistent information regarding network reachability. UPDATE messages can add or withdraw routes from the RIBs depending on the content of the message.

A NOTIFICATION message is sent between neighbours when an error is detected. These messages will close an active session and inform peers of the reasons.

A KEEP-ALIVE message notifies BGP peers that a BGP neighbour is still active. These messages are sent as needed to prevent the session from expiring.

UPDATE messages are comprised of a standard BGP header along with additional fields. Figure B.1 illustrates the standard BGP header, while Figure B.2 illustrates the additional fields that make up a BGP UPDATE message. Upon receiving an UPDATE message, routers will be able to add or delete entries from their respective routing tables.

16 bytes	2 bytes	1 bytes	Variable
Marker	Length	Type	Data

Figure A-1 - A BGP Packet Header

2 bytes	Variable	2 bytes	Variable	Variable
Unfeasible Routes Length	Withdrawn Routes	Total Path Attribute Length	Path Attributes	Network Layer Reachability Information

Figure A-2 - BGP UPDATE Message Fields

BGP UPDATE messages contain the following fields:

Unfeasible Routes Length – This indicates the total length of the withdrawn routes field or that the field is not present

Withdrawn Routes - Contains a list of withdrawn IP address prefixes for routes

Total Path Attribute Length - The total length of the path attributes field

Path Attributes - The characteristics of the path being advertised. The following are possible path attributes:

- *Origin*: A mandatory attribute that defines the origin of the path information
- *AS Path*: A mandatory attribute composed of a series of AS path segments
- *Next Hop*: A mandatory attribute which defines the IP address of the router that should be used as the next hop to all destinations listed in the network layer reachability information (NLRI) field
- *Mult Exit Disc*: An optional attribute which is used to discriminate between exit points to a neighbouring AS
- *Local Pref*: A discretionary attribute which is used to specify the degree of preference for an advertised route
- *Atomic Aggregate*: A discretionary attribute which is used to disclose information about route selections
- *Aggregator*: An optional attribute which contains information about aggregate routes

Network Layer Reachability Information - This list contains IP address prefixes for the advertised routes.

BGP will lend itself in the creation of OBGp and act as the control domain for the all-optical network.

B Appendix B

OBGP Model in OPNET Modeler

B.1 OPNET Simulation Model

The OBGP Simulation Model was integrated into the BGP Process Model. An OBGP Router was created with the SCS integrated into the BGP Process. The model was created to simulate the control domain, which excluded the data domain. The reasoning behind excluding the data domain was that the study was not focused on the transmission of end-user data, but the ability to establish lightpaths between the different users.

The physical layer itself was not modeled, but was considered in the SCS process within the BGP Process Model.

B.1.1 BGP Process Models

The BGP Process Model was pre-defined in the OPNET library. A brief description of the BGP functionality in OPNET will be described, and then the additions will be discussed.

OPNET Modeler implements BGP with two process models: the BGP Process and the BGP Connection Process. The BGP Process acts as the central BGP process on a router/gateway. This is an application layer process running above TCP (port 179). The BGP Process acts as a parent and maintains child processes, which are the BGP Connection Processes, to manage each port leading to a BGP device.

Figure B-1 – The BGP Process Model (The Parent Process)

After initialization, the BGP Process enters an ACTIVE state, illustrated in Figure B.1. The BGP Connection processes an ESTABLISHED state, once a BGP connection is established between itself and the peer, as shown in Figure B.2. When any messages arrive into the BGP Connection Process, the BGP Process examines the message type and enters the proper state to handle the message. For example, if an OPEN message arrives, the BGP Process will be triggered to handle it in the OPEN_INDICATION State.

The initial event is created at the INITIAL state in the BGP Process and is inserted into an Event queue in order of increasing time. This queue was infinitely large, and contained all different types of events for all lightpaths maintained by the node. Events were serviced at each scheduled interrupt, and consisted of an event type (Request or Teardown), the unique index given to the request, the time that the event was scheduled to occur, and the connection ID that is related to the event.

When an event occurred, an interrupt was generated on the BGP Process, forcing the state transition from the ACTIVE state to the OBGP_REQUEST state. In this state, the event type is analysed and is handled appropriately.

As requests are fulfilled, the next interrupt is scheduled with the following event in the event queue. When servicing the events, if more than one occurred at the same time, then they were serviced in a First-Come First-Serve order.

It should be pointed out that each OBGP Router managed a request generator, meaning that each BGP Process had its own event queue and interrupts. There was no universal request generator for the entire scenario, making the request generation of a specific node independent of another's.

B.1.3 OBGP Message Handler

OBGP Messages are another type of BGP message with the type code set to 5. The BGP Process Models had to be modified to handle this new message type. The BGP Process was modified by introducing a new state, called the OBGP_MESSAGE state.

To understand how the BGP Process Models handle an OBGP Message, one must understand the interaction between the child and the parent. As a BGP message arrives into a BGP router, the BGP Process is interrupted, indicating that a new packet has

arrived on one of its ports. The BGP Process then tells all of the BGP Connection Processes to check whether or not the packet arrived on the connection it is managing.

Assuming the packet arrived on one of the BGP Connection Processes, the packet is examined and it is determined what type of packet it is: OPEN, UPDATE, NOTIFICATION, KEEP-ALIVE, or OGBP. Once it is determined, this information is passed back to the BGP Process, who then decides what form of action must be taken for the message.

For the OGBP Message, the BGP Process examines the incoming message and sends it to the OGBP_MESSAGE state. Within this state, the BGP Connection Process that received the message is triggered and asked to process the message. The message is dissected and analysed based on OGBP Message type it is. Once completed, information is passed back to the BGP Process and is then forwarded along to the following BGP Connection Process, based on the message type and the next AS in the AS Path.

B.2 OGBP Configuration

Configuration of the network requires that each node speaks BGP with one another. The domain was represented by a border gateway router, as mentioned above. In this network, there are three S-D pairs interconnected with links that maintain 30 wavelengths each. The following information is provided for the readers to configure a test network. Figure B.3 shows the network diagram.

Figure B-3 – The Network Diagram

B.2.1 Wavelength Configuration Files

Wavelengths are routed statically, through the use of a wavelength configuration file. This file is similar to what a network administrator would use to identify connections to adjacent domains. The format of the configuration file is shown in Figure B.4.

Dest AS	λ ID	OXC Source Port ID	OXC Dest Port ID	Dest Port IP Address	Dest Port IP Mask
---------	--------------	--------------------	------------------	----------------------	-------------------

Figure B-4 – Wavelength Configuration File Format

The Dest AS is the Destination AS Number that the lightpath is to be connected to. The λ ID is the unique wavelength ID that identifies the wavelength to the network. The Source and Destination Port ID are the OXC specific port identifiers, upon which the wavelength is switched to and from. The Destination Port IP Address is the IP Address of the Control Domain port and its mask.

During the INIT State of the BGP Process, each router will be configured with this information. It is important for the user to configure each wavelength configuration file carefully, to ensure that neighbouring nodes maintain a consistent collection of resources between themselves.

B.2.2 Nodal Configuration

Table B.1 show the IP configuration parameters, whereas Table B.2 show the Request generation powers of the three sources.

AS	Peer AS	Port Number	IP Address	Link Rate	MTU	Routing Protocol
1000	2000	10	192.0.10.1	DS3	IP	BGP
1005	2000	11	192.0.15.1	DS3	IP	BGP
1001	3000	13	192.0.11.1	DS3	IP	BGP
2000	1000	10	192.0.10.2	DS3	IP	BGP
	1005	11	192.0.15.2	DS3	IP	BGP
	3000	12	192.0.20.23	DS3	IP	BGP
3000	4000	10	192.0.30.34	DS3	IP	BGP
	1006	11	192.0.16.2	DS3	IP	BGP
	2000	12	192.0.20.32	DS3	IP	BGP
	1001	13	192.0.11.2	DS3	IP	BGP
1006	3000	11	192.0.16.1	DS3	IP	BGP
1007	4000	12	192.0.17.1	DS3	IP	BGP
4000	3000	10	192.30.43	DS3	IP	BGP
	1007	12	192.17.2	DS3	IP	BGP
	1002	13	192.0.12.2	DS3	IP	BGP
1002	4000	13	192.0.12.1	DS3	IP	BGP

Table B-1 – IP Configuration Parameters

Source AS	Destination AS	Dest IP	Dest Mask	Request Interarrival Time	Request Hold Time	Requests Generated
1000	1002	192.0.12.1	255.255.255.0	Poisson(x)	Exponential(1)	Unlimited
1001	1007	192.0.17.1	255.255.255.0	Poisson(x)	Exponential(1)	Unlimited
1005	1006	192.0.16.1	255.255.255.0	Poisson(x)	Exponential(1)	Unlimited

Table B-2 – Request Generation Parameters