IMPLEMENTATION OF AUTOMATIC PROTECTION SWITCHING IN AN OPTICAL CROSS CONNECT

by

Jason Uy
Bachelor of Applied Science, University of British Columbia, 2000

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING

In the School
of
Engineering Science

SIMON FRASER UNIVERSITY

Spring 2005

APPROVAL

Name:	**Jason Uy**
Degree:	**Master of Engineering**
Title of Project:	**Implementation of Automatic Protection Switching in an Optical Cross Connect**

Examining Committee:

Chair:

____

Dr. John Bird
Professor of Engineering Science

____

Dr. Stephen Hardy
Senior Supervisor
Professor of Engineering Science

____

Dr. Paul K. M. Ho
Supervisor
Professor of Engineering Science

Date Defended/Approved: April 11, 2005

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

ABSTRACT

Having a reliable network is a hard requirement for Telecommunication companies when deploying new networks. Service providers and enterprise customers lose a lot of money any time an interruption of internet service occurs. The SONET/SDH specification specifies several different types of topology that support redundancy. An Automatic Protection Switching (APS) mechanism is specified for each topology to dictate how a network behaves in a failure event. For this project, a software implementation of APS is coded and tested on a PMC SONET/SDH Demonstration System Platform.

DEDICATION

To my caring wife Alison, and to my parents.

ACKNOWLEDGEMENTS

I would like to thank the following professors for providing me with the knowledge, education, and support I needed to complete this degree. Dr. Marek Syrzycki for sharing his expertise in the field of analog and digital design. Dr. Kris Iniewski for explaining how optical networks are designed and built. Dr. Ljiljana Trajkovic for teaching how each network protocols work. Dr. Gary Robinson for providing project management techniques. Dr. Glenn Chapman for introducing the field of IC fabrication. Dr. John Dill for showing how data can be better presented and visualized. Dr. Stephen Hardy and Dr. Paul Ho for reviewing my Masters project.

TABLE OF CONTENTS

Approval ii
Abstract iii
Dedication iv
Acknowledgements v
Table of Contents vi
List of Figures vii
List of Tables viii
Glossary ix
1 Introduction to SONET/SDH Networks 1
1.1 Optical Network Topology 2
1.2 Optical Cross Connect (OXC) 4
2 Automatic Protection Switching (APS) 6
2.1 K1 and K2 APS Message Bytes 6
2.2 1:N Protection 8
2.3 1+1 Protection 10
2.4 SONET/SDH Alarms 11
3 PMC Sierra's OXC Demonstration System 12
3.1 Board Description 13
3.2 Creating a Connection 15
4 Software Implementation 18
4.1 PMC Sierra Device Driver Interface 18
4.2 APS Module 19
4.3 Open Path Algorithm 28
4.4 VxWorks Real Time Operating System 34
4.5 Control Module 35
5 Test Result 37
Bibliography 40

LIST OF FIGURES

Figure 1: SONET STS-1/SDH STM-0 Frame Format 2

Figure 2: Linear Network 3

Figure 3: Sprint Ring Network 4

Figure 4: OXC Block Diagram 5

Figure 5: Linear 1:N Protection Scheme 9

Figure 6: Linear 1+1 Bridging 10

Figure 7: Linear 1+1 Selecting 10

Figure 8: Chassis Slot Allocation 13

Figure 9 Time-Space-Time Architecture 16

Figure 10 Clos Architecture Logical Diagram 17

Figure 11 Hardware Control Software Architecture 18

Figure 12 APS Module Functional Blocks 20

Figure 13 1+1 Association Task State Machines 21

Figure 14 1+1 Example System Before Switchover 23

Figure 15 1+1 Transition Diagram 25

Figure 16 1+1 Example System After Switchover 25

Figure 17 1:N Transition Diagram 27

Figure 18 System View of a Simple Cross Connect 28

Figure 19 OPA Internal Model 29

Figure 20 First Connection Created 29

Figure 21 System View of the First Connection 30

Figure 22 Internal Model After 6 Connections 31

Figure 23 System Model After 6 Connections 31

Figure 24 Merged Timeslots 32

Figure 25 Rearrangement Steps 33

Figure 26 Separated Space Switch Mapping 34

Figure 27 Rearranged Connections - System View 34

Figure 28 Interrupt Flow Diagram 36

Figure 29 Test Setup for 1+1 Protection 37

Figure 30 Test Setup for 1:N Protection 38

LIST OF TABLES

Table 1: SONET/SDH Bandwidth Nomenclature..1

Table 2: K1 Bits 1 through 4 – Type of Request..7

Table 3: K1 Bits 5 through 8 - Channel Number..7

Table 4 Four Different Scenarios..22

Table 5 External Events/Conditions..22

Table 6 New Connections..31

Table 7 Switchover Latency Result..38

GLOSSARY

APS

Automatic Protection Switching: Mechanism for automatically selecting a redundant datapath for the purpose of resuming flow of traffic during an alarm condition on the active datapath.

Mbps

Megabits per second: The unit used in measuring the bandwidth of networks.

OC-N

Optical Carrier Level N: The terminology used to define the rate of an optical network. The different level N specifies a speed multiple of the slowest supported optical rate (51.84 Mbps).

OPA

Open Path Algorithm: A PMC Sierra developed algorithm that determines an optimized setting for creating connections in a Clos-based switch architecture.

OXC

Optical Cross Connect: A network element that grooms/switches traffic coming from multiple fiber optic connections.

SDH

Synchronous Digital Hierarchy: The optical standard used in Europe and Asia. The SDH standard is defined by the ITU G.707.

SONET

Synchronous Optical Network: The optical standard used in North America. The SONET standard is defined by Belcore/Telcordia GR253.

STM

Synchronous Transport Module: Terminology used to refer to rates of links/devices in a SDH network.

STS

Synchronous Transport Signal: Terminology used to refer to rates of links/devices in a SONET network.

WTR

Wait to Restore: Timer function used in APS to determine how long after a reestablishment of a datapath should the system wait before utilizing it.

1 INTRODUCTION TO SONET/SDH NETWORKS

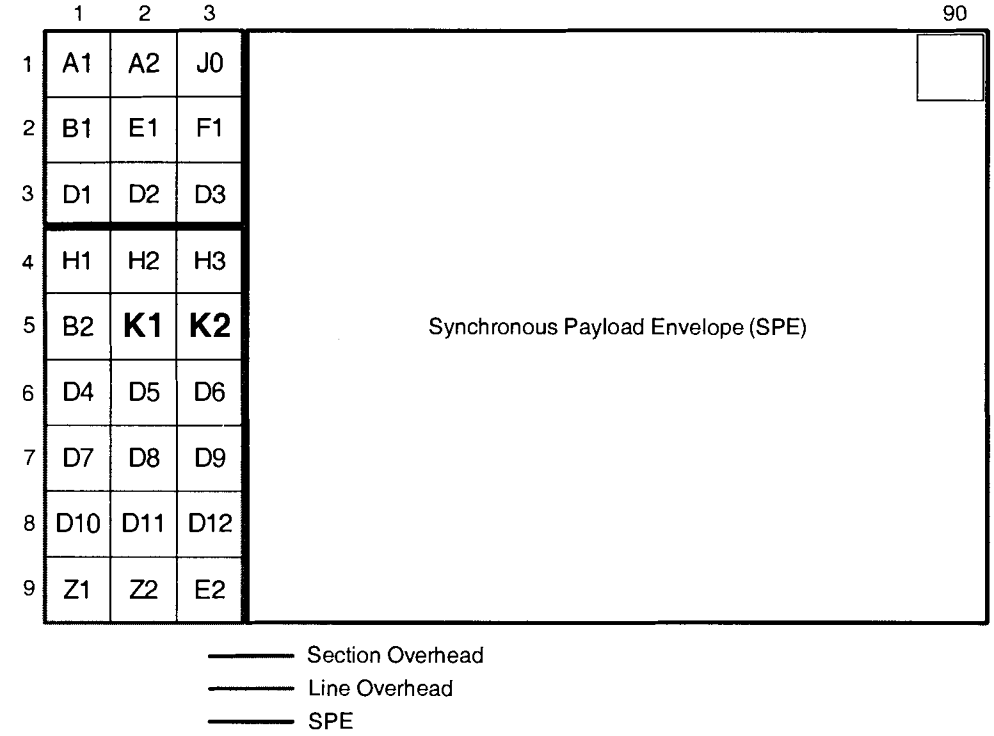
SONET/SDH is a time division multiplexing layer 1 protocol that is used on optical networks. The maximum bandwidth of an optical fiber is in the terabits per second range, but current optical networks are running at a much slower rate than this. The type of laser and the speed of the SONET/SDH Framer dictate the maximum bandwidth for an optical network. Current SONET/SDH Framers can handle speed from 51.84 Mbps up to 40 Gbps. A rate of 51.84 Mbps is referred to as OC-1 when talking in terms of the fiber optic bandwidth, and STS-1/STM-0 when talking in terms of device bandwidth. Table 1 below shows a list of common bandwidth used in the industry as well as the optical and device bandwidth name associated with it.

Table 1: SONET/SDH Bandwidth Nomenclature

Bandwidth (Mbps)	Fiber Optic Terminology	SONET Terminology	SDH Terminology
51.84	OC-1	STS-1	STM-0
155.52	OC-3	STS-3	STM-1
622.08	OC-12	STS-12	STM-3
2488	OC-48	STS-48	STM-12
9953	OC-192	STS-192	STM-48
39812	OC-768	STS-768	STM-192

The SONET/SDH protocol uses a frame structure to align the data it carries. A STS-1/STM-0 frame consists of 90 columns and 9 rows for a total of 810 bytes (as shown in Figure 1). A STS-3/STM-1 frame consists of 270 columns and 9 rows for a total of 2430 bytes. A SONET/SDH frame is always 125 μs long regardless of the bandwidth. This means that as the bandwidth increases, the transmission duration for each byte decreases. The K1 and K2 bytes are highlighted in Figure 1 as they play an important role in the automatic protection switching process.

Figure 1: SONET STS-1/SDH STM-0 Frame Format



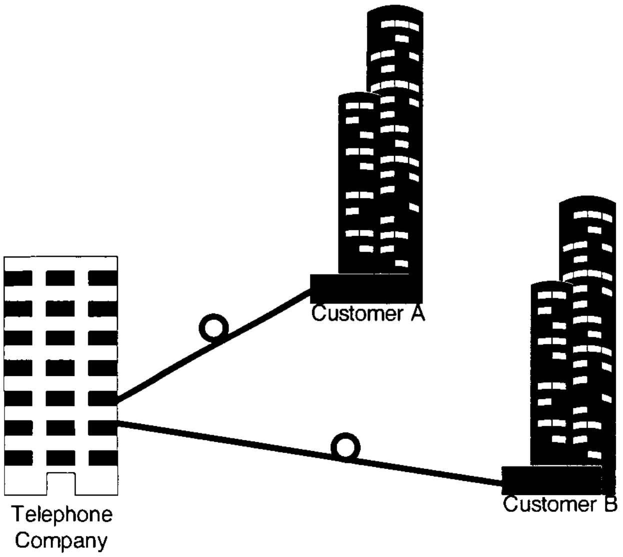
	1	2	3	90
1	A1	A2	J0	Synchronous Payload Envelope (SPE)
2	B1	E1	F1	
3	D1	D2	D3	
4	H1	H2	H3	
5	B2	**K1**	**K2**	
6	D4	D5	D6	
7	D7	D8	D9	
8	D10	D11	D12	
9	Z1	Z2	E2	

1.1 Optical Network Topology

There are currently two different types of optical network topology: Linear, and Ring Networks. The most common type of network used by Internet Service Providers is a ring network.

The Linear network topology is the simplest form of optical network and is designed to be a point to point connection. Such an architecture can be used to connect a customer to the Internet Service Provider (ISP), or to connect a gateway of one continent to another. Figure 2 illustrates a typical linear network that can be found in any major city.

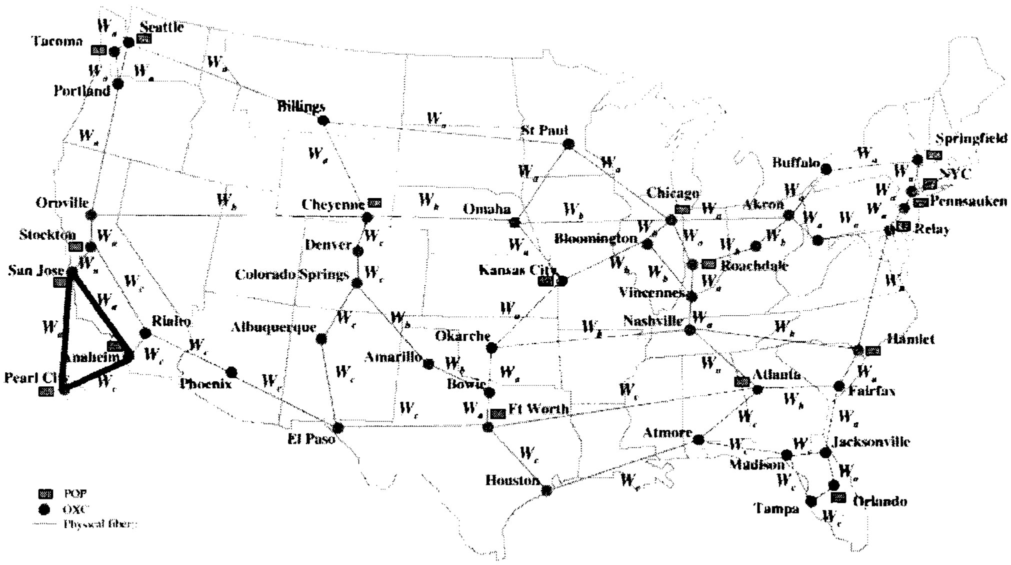
Figure 2: Linear Network



The ring network architecture is a common topology used for connecting multiple nodes to form a single network. Each node is usually located in a heavy traffic area (such as a major city). Figure 3 shows the Sprint network in the United States. The highlighted section in the diagram shows the SONET ring in Southern California that connects the cities of San Jose, Pearl City, and Anaheim. Multiple ring networks can be connected together to provide connectivity and coverage to all users in a large area.

To create any type of optical network, a switching element must be used to allow the network administrator the flexibility to route traffic to different locations. An Optical Cross Connect (OXC) is the most common type of switching element used by ISPs (indicated as a red dot in Figure 3).

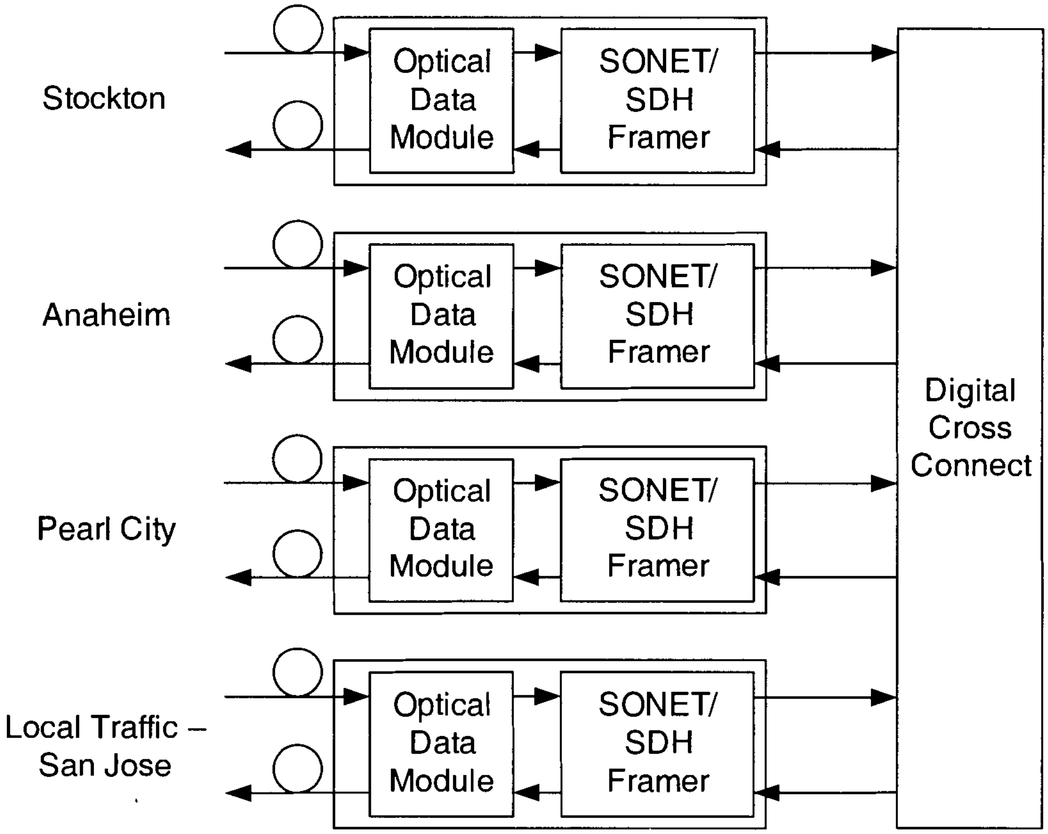
Figure 3: Sprint Ring Network



1.2 Optical Cross Connect (OXC)

An OXC is a network element that takes in optical input (from a fiber optic) and transmit an optical output (to another fiber optic). The OXC transforms optical signals into electrical signals using Optical Data Module (ODM). The ODM connects to SONET/SDH framers where data integrity is analyzed and clock timing is extracted. Figure 4 shows a block diagram of an OXC that might be used in an area like San Jose. Figure 3 shows that there are three optical links connected to San Jose: Stockton, Pearl City, and Anaheim. A fourth optical link (not shown in the diagram) connects the local San Jose internet traffic to the rest of the network. A network administrator determines the bandwidth allocation for each node in order to maximize the traffic going through the network. For example, if maximum utilization is detected for the direct link connecting San Jose and Anaheim, the network administrator can choose to allocate additional bandwidth coming from San Jose to go through Pearl City before reaching Anaheim.

Figure 4: OXC Block Diagram



Designing an optical cross connect is relatively simple as there are many standard products that can be connected together to perform the desired functionality. For this project, the SONET/SDH Optical Cross Connect demonstration system of PMC is used as a hardware platform to implement one of the most important features of an optical network: Automatic Protection Switching.

2 AUTOMATIC PROTECTION SWITCHING (APS)

Optical networks are usually designed with redundancy in mind to guarantee a high degree of reliability that Telecommunication companies require. During the design stage, all single points of failure are eliminated by creating backup circuitry/connections. This means that all active cards in an OXC have a redundant card waiting to take over once an error is detected. Fiber connections between two nodes are also duplicated creating two identical links called a Working link and a Protect link. Since fiber optic links can be disrupted due to unavoidable causes (e.g. during an earthquake), the protect links are ideally laid out far away from the links that are being protected. Switching between the working and protect links require communication between at least two nodes in the network. To facilitate the communication, two bytes in the SONET/SDH frame is allocated for this purpose: K1 and K2 bytes.

2.1 K1 and K2 APS Message Bytes

The APS mechanism uses the K1 and K2 bytes of the line overhead (see Figure 1) to transmit information to the relevant nodes in the network. The K1 byte is used to carry a request from a channel for a switch operation. The first four bits of the K1 byte carries the Type of Request, and the last four bits indicate the channel number initiating the request. Table 2 shows the different types of request that can be sent through the first four bits of the K1 byte.

Table 2: K1 Bits 1 through 4 – Type of Request

Bit Value	Request
1111	Lockout of Protection
1110	Forced Switch
1101	SF – High Priority
1100	SF – Low Priority
1011	SD – High Priority
1010	SD – Low Priority
1001	(not used)
1000	Manual Switch
0111	(not used)
0110	Wait-to-Restore
0101	(not used)
0100	Exercise
0011	(not used)
0010	Reverse Request
0001	Do Not Revert
0000	No Request

Table 3 lists the valid values on the last 4 bits of the K1 byte. Note that there are 14 values allocated for the working channels. This means that the SONET/SDH standard can only accommodate up to a maximum of 14 working channels for a single protection group.

Table 3: K1 Bits 5 through 8 - Channel Number

Bit Value	Request
0000	Null channel
0001 – 1110	Working channels
1111	Extra traffic channel

The K2 byte carries the channel number using the protection link, as well as the protection type and mode of operation. The first four bits of the K2 byte contains the channel number using the protection link (see Table 3). The fifth bit notifies the receiving end what type of protection is being used. A bit value of 0 signify a 1+1 architecture while a bit value of 1 signify a 1:N architecture (see sections 2.2 and 2.3). A bit value of 101 on the last three bits signify a bidirectional protection mechanism, while a bit value of 100 signify a unidirectional protection mechanism. All other values are reserved for future use.

SONET/SDH specification states that when an error is detected, the network has up to 50 ms to do an APS in order to prevent traffic disconnection (e.g. telephone conversation). The time restriction is independent on what type of architecture is being used. If a ring architecture with many nodes are used, the propagation delay from one node to another must be taken into account in the APS as communication between the relevant nodes will take longer.

There are many different kinds of protection schemes available for both Linear and Ring network topology. For this project, the two protection schemes for Linear network topology are implemented: 1:N and 1+1.

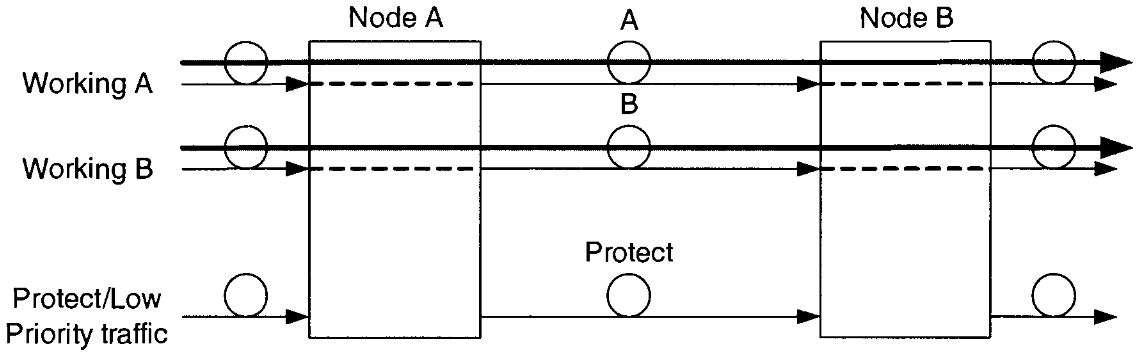
2.2 1:N Protection

A 1:N protection scheme uses a single protection link for a group of active working links (up to a maximum of 14 active links). Each working link is assigned an APS priority number to determine which links can utilize the protection link during a multiple failure scenario. The priority number is only used when the protection link is already carrying redundant traffic. Low priority data is allowed to go through the unused protect link in order to maximize utilization of the network. If the protection link is not carrying any traffic or carrying low priority traffic, the failed working link will be immediately switched.

When an alarm is detected by an OXC, the low priority data is stopped, and the redundant traffic is sent to the protect link. To complete a switchover from working link to protect link, a two-step process is initiated: Bridging and Selecting.

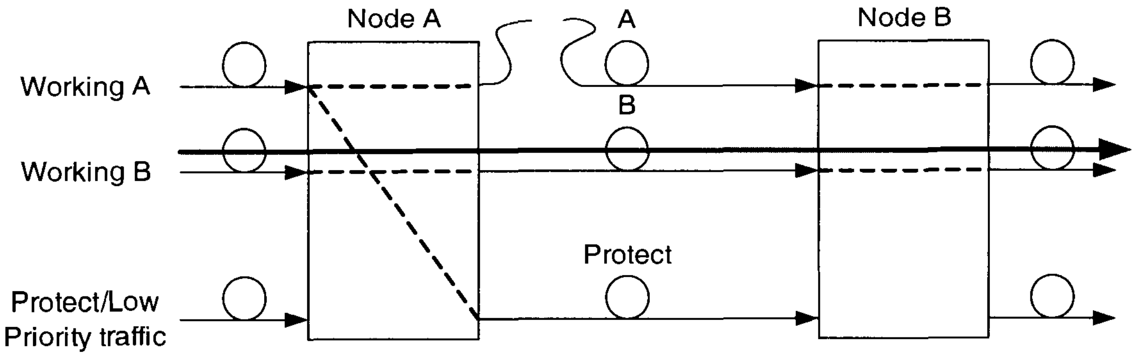
Figure 5 shows an example of a 1:N protection scheme. Node A is connected to Node B through three links: working A, working B and protect links. The diagram illustrates a unidirectional link, but the same concept applies to the other direction.

Figure 5: Linear 1:N Protection Scheme



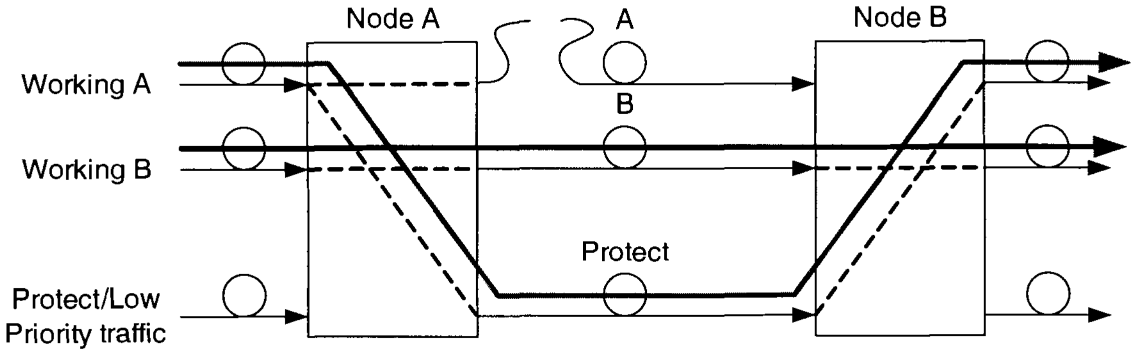
When no errors are detected, Working A and Working B traffic passes through its respective fiber optic link, while optional low priority traffic passes through the protect link. Both nodes A and B are configured for a 1:1 mapping (data from input 1 goes to output 1). When an error is detected (e.g. fiber disconnect or cut), Node B will send the failure information to Node A and this will cause Node A to initiate an automatic protection switchover. When Node A receives the error message in the K1 and K2 bytes, it will initiate a bridging action as shown in Figure 6. Node A will create a duplicate copy of the Working A traffic and sends it to the protect link. This action essentially stops any low priority traffic from being transmitted. After Node A completes the bridging request, it will send a selection command to Node B using the protect link (link A is assumed to be unavailable).

Figure 6: Linear 1+1 Bridging



Once Node B receives the selection command, it will configure its internal cross connect setting to source data from the protect link (see Figure 7). After the selection command is done, the APS procedure is complete.

Figure 7: Linear 1+1 Selecting



2.3 1+1 Protection

A 1+1 protection scheme allocates a protect link for every working link. This means that traffic can always be protected. Low priority traffic can also be optionally sent through the protect link but this is not usually done. The procedure for protection switching is the same as the 1:N protection.

2.4 SONET/SDH Alarms

There are many alarms that can trigger a protection switchover. Each of the alarms are mapped to one of four request type as listed in Table 2: SD – High Priority, SD – Low Priority, SF – High Priority, and SF – Low Priority.

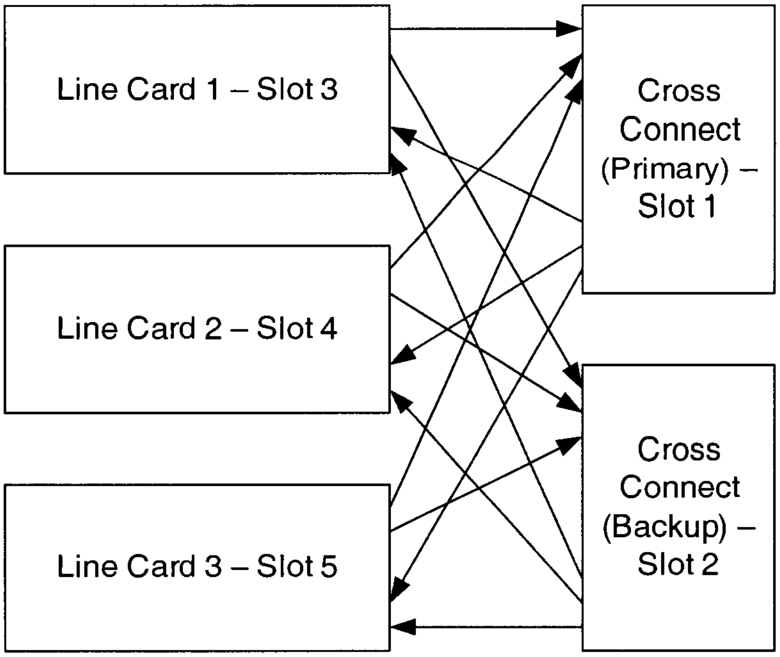
The only alarm that can be easily produced for demonstration purposes is the Loss of Signal alarm (LOS). A LOS alarm is triggered when the SONET/SDH framer detects a long stream of 0s on its input. A long stream of 0s is produced when a fiber optic cable is disconnected or cut. The LOS alarm is mapped into the SF – High Priority request type. Other types of alarms can be used to trigger a switchover. For example, a Loss of Pointer (LOP) condition can trigger a switchover. However, a SONET/SDH tester is required to corrupt the pointer values in the data stream. The requirement of having an optical tester prevented such alarm to be used for demonstration purposes.

3 PMC SIERRA'S OXC DEMONSTRATION SYSTEM

The demonstration system used for this project was designed to be portable so that sales engineers can easily carry the whole system and demonstrate to customers the capability of PMC's SONET/SDH devices. The chassis contains five slots to support a variety of cards that perform different functionality and support different bandwidth. The cards connect to a backplane in order to facilitate transmission of information from one card to another. A single board computer based on an Intel processor is also present and runs the VxWorks real time operating system and the demonstration software.

The first two slots on the chassis are for the cross connect card. As mentioned previously, redundancy is important so it is crucial that this is showcased in the demonstration system. The three other slots are for connecting different types of cards depending on what type of functionality is required. Figure 8 shows the logical connection between all five slots. Each line card is connected to both active and backup cross connect to guarantee that data flow can continue when the primary cross connect goes down. When the system is operating normally, the line cards select data from the active cross connect. When an error is detected on the primary cross connect, each card will select the data coming from the backup cross connect. The selection between the primary and backup cross connect is a manual process and must be managed by the demonstration software.

Figure 8: Chassis Slot Allocation


Line Card 1 – Slot 3
Line Card 2 – Slot 4
Line Card 3 – Slot 5
Cross Connect (Primary) – Slot 1
Cross Connect (Backup) – Slot 2

3.1 Board Description

There are five different types of boards that are supported by the demonstration system: Cross connect card, 1xOC48 Line Card, 4xOC12 Line Card, 4xOC3 Line Card, and an OC48 ATM Card. Different varieties of cards are required to demonstrate the wide range of functional capabilities of the demonstration system.

Cross Connect Card:

The cross connect card is made up of PMC Sierra's PM5372 TSE device that is capable of cross connecting 64 STS-12 links. The overall capacity of the device is 40 Gbit/s, but only a small portion of this bandwidth is utilized in the demonstration system. The cross connect card also acts as the timing card and provides both clock and frame reference pulse to all the cards for synchronization purposes.

1xOC48 Line Card:

The 1xOC48 Line Card uses two of PMC Sierra's SONET/SDH devices: PM5316 SPECTRA 2488 and PM5310 TBS. The SPECTRA 2488 is a SONET/SDH Framer that is capable of receiving OC48 data link from a single fiber optic, or 4xOC12 data link from four fiber optics. For this line card, the SPECTRA 2488 is configured to receive OC48 from a single fiber optic. Data that comes out of the framer is a 32 bit wide parallel bus. It is generally a poor practice to route a parallel bus through long traces (e.g. backplane). Due to variation in trace length, the bits contained in each signal of a parallel bus can arrive at different times. To get around this problem, the TBS device is used to serialize the incoming parallel data into a serial bit stream format. The TBS also duplicates the incoming parallel traffic in order to send the same serial data to two different cross connect cards (for redundancy purposes).

4xOC12 Line Card:

Just like the 1xOC48 Line Card, the 4xOC12 Line Card uses two of PMC Sierra's SONET/SDH devices: PM6316 SPECTRA 2488 and PM5310 TBS. The SPECTRA 2488 is configured to receive OC12 data from four fiber optics. The TBS performs the serializing function to communicate with the two cross connect cards.

4xOC3 Line Card:

The 4xOC3 Line Card uses the PM5315 SPECTRA 4x155 and the PM5310 TBS. Just like the SPECTRA 2488, the SPECTRA 4x155 is a SONET/SDH framer that is able to receive four channels of OC3 data from four fiber optics.

OC48 ATM Card:

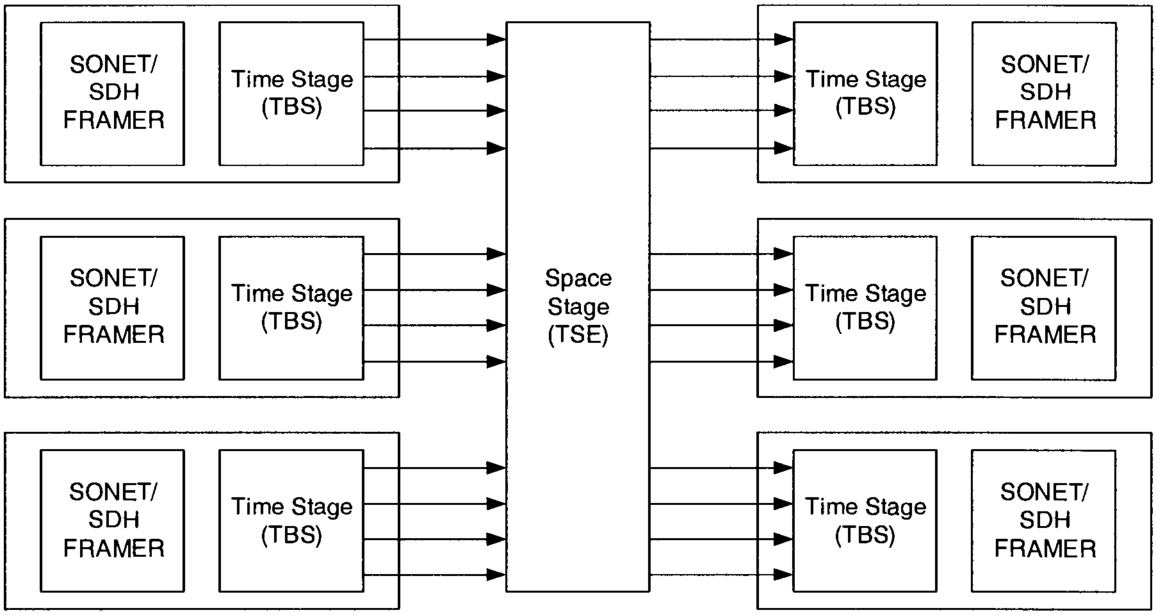
The OC48 ATM Card extracts and processes the ATM cells embedded into a SONET/SDH data stream. PMC Sierra's PM7390 MACH48 device implements the extraction and processing functionality of the OC48 ATM Card. The MACH48 handles cell delineation and supports virtual concatenation for bandwidth monitoring/limiting functionality.

3.2 Creating a Connection

An optical cross connect allows the network administrator to dynamically create a connection from any input to any output data stream. Each data stream carries individual timeslots that contains data from users of the network. For an optical cross connect, the switching granularity is STS-1/STM-0 (51.44 Mbit/s). As mentioned in the previous section, the cross connect card supports 64 STS-12 links. A 1xOC48 Line Card will be connected to the cross connect card through four STS-12 links. A 4xOC3 Line card will be connected to the cross connect card through one STS-12 link.

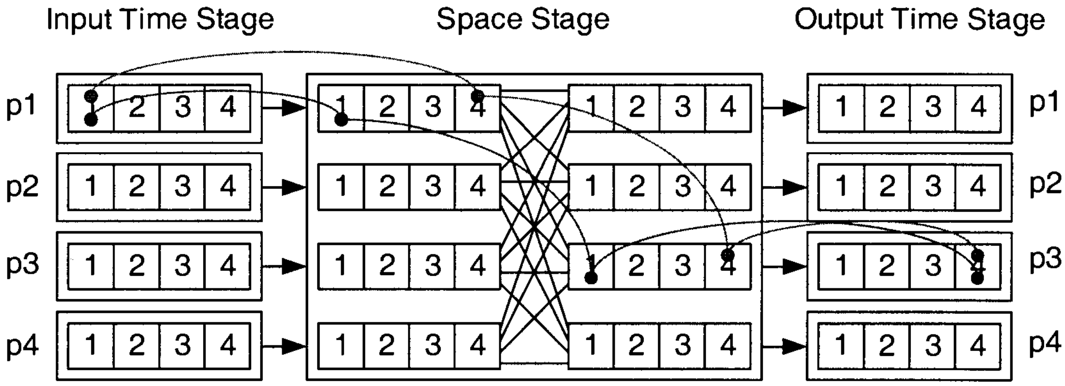
The switching architecture of the optical cross connect is referred to as Clos Architecture. A Clos architecture uses three stages to create every connection: Input Time Stage, Space Stage, and Output Time Stage. A time stage allows switching of any input timeslot to any output timeslot within the same port/link. A space stage allows an input timeslot from any input port to be switched to the same timeslot of any output port. Figure 9 shows an illustration of the cross connect mechanism, and indicates the PMC Sierra device that implements each stage. Every connections created in the cross connect requires configuration changes on all three stages. Clos architectures are rearrangeably non-blocking in nature. Blocking is a term referred to when there are unused timeslots in the system, but a connection cannot be created. For a Clos architecture, blocking is possible as the load of the fabric increases. When a blocking scenario occurs, rearranging existing connections can allow new connections to be created.

Figure 9 Time-Space-Time Architecture



Creating a connection is trivial in a Clos architecture. However, without a proper switching algorithm to derive the connection settings for all three stages, it is hard to determine what rearrangements can be done to allow more connections to be made. There are usually more than one way to create a connection from an input timeslot to an output timeslot. Figure 10 shows a logical diagram of a simple Clos architecture (4 port and 4 timeslots), and highlights two different configurations that creates a connection from input port 1 timeslot 1 to output port 3 timeslot 4.

Figure 10 Clos Architecture Logical Diagram

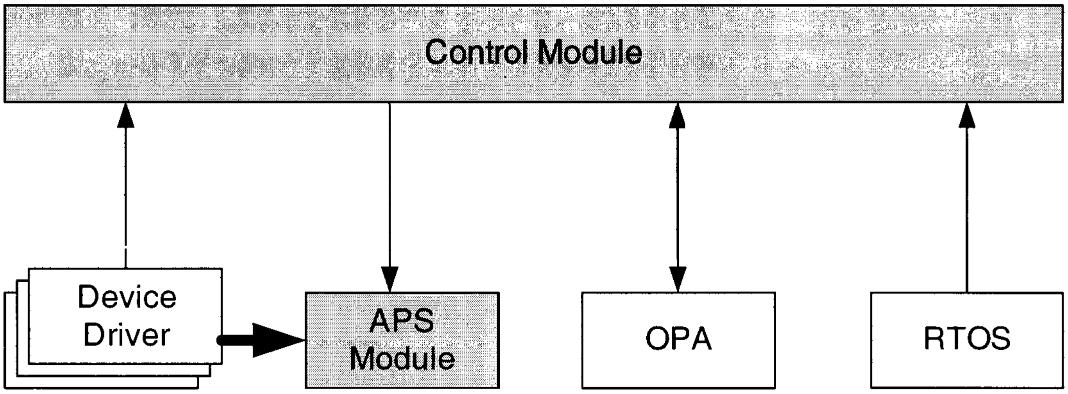


In the demonstration system, an Open Path Switching Algorithm is used to determine the optimal connection and rearrangement settings (see section 4.3 for more details).

4 SOFTWARE IMPLEMENTATION

The software for the OXC demonstration system can be divided into two categories: Application software and Hardware Control software. The Application software runs on a Windows platform, and provides the user interface to control all the functionality of the demonstration system. The Hardware Control software runs on a single board computer and controls all the different cards in order to implement features such as automatic protection switching. The Hardware Control software has an architecture shown in Figure 11.

Figure 11 Hardware Control Software Architecture



Shaded blocks represent software written for the demonstration system. Non-shaded blocks were written prior to the project.

4.1 PMC Sierra Device Driver Interface

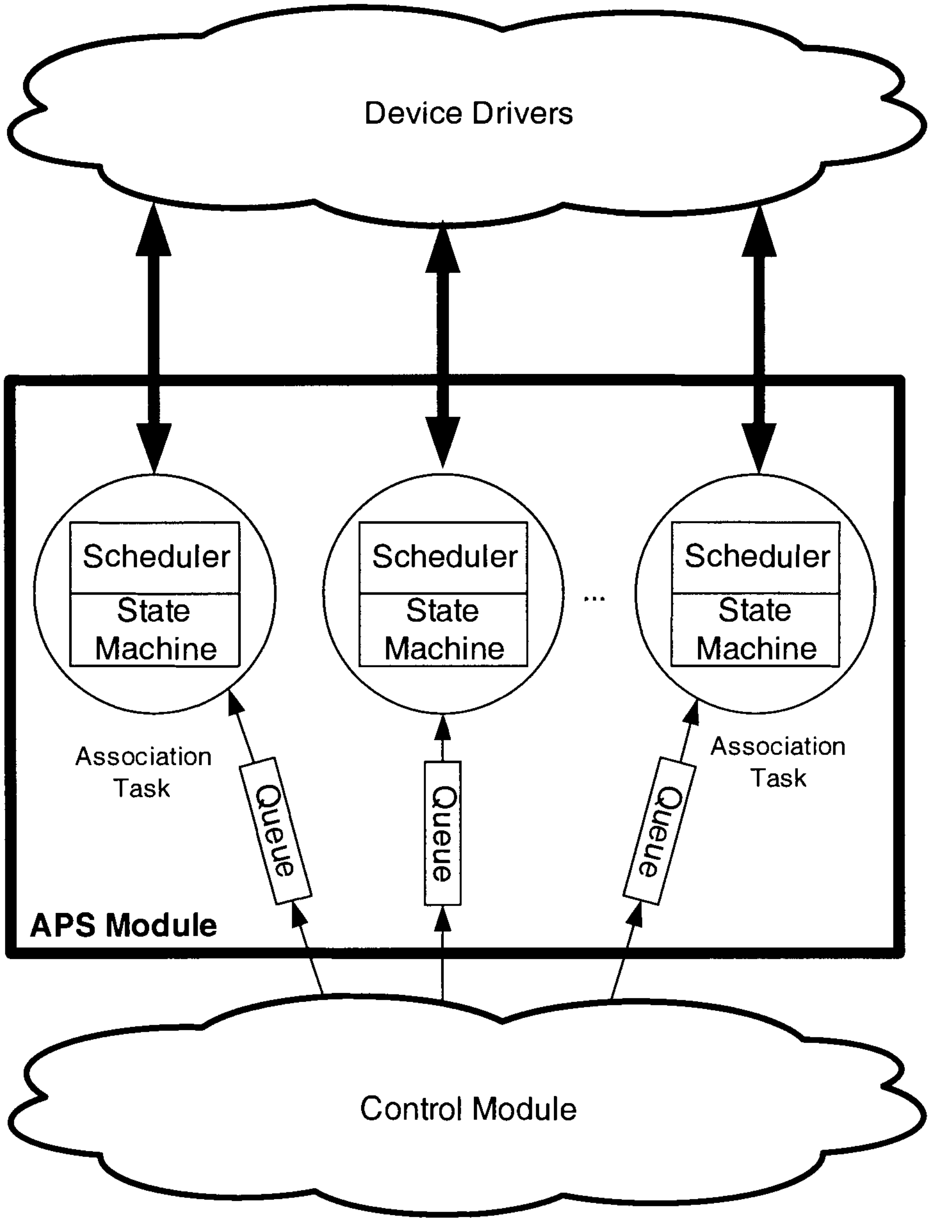
The device drivers for the SONET/SDH framers are used to raise interrupts when either one of the following events occur: error condition, and change of K1/K2 bytes.

When an interrupt is triggered by the device driver, an Interrupt Service Routine (ISR) will handle the raised interrupt by reading the device registers to determine the nature of the interrupt. The ISR will send a message to the Control Module indicating which device triggered the interrupt and the nature of the interrupt.

4.2 APS Module

The APS Module keeps track of the state of each protection group provisioned in the demonstration system. A protection group is created by assigning physical ports in the demonstration system as working and protect. To support protection switching, the APS Module is subdivided into smaller functional blocks as shown in Figure 12.

Figure 12 APS Module Functional Blocks



An Association Task is created for every protection group provisioned in the system. Within each Association Task, two functional blocks are instantiated: Scheduler and State Machine. The Scheduler uses a round robin algorithm to poll relevant interrupts from all the devices associated to the protection group. The Scheduler also polls a control queue to determine if there are manual commands received from the Control Module (e.g. Manual

Switchover). Each Association Task has state machines that determines how switching should occur given a received Kl/K2 bytes or given any incoming interrupts from the device driver. Figure 13 shows the state machines used to handle four different system scenarios. The Working State is the state where valid traffic is sourced from the working channel. The Protect State is the state where valid traffic is sourced from the protect channel. The Association Task changes the active state machine used when a change in scenario is detected.

Figure 13 1+1 Association Task State Machines

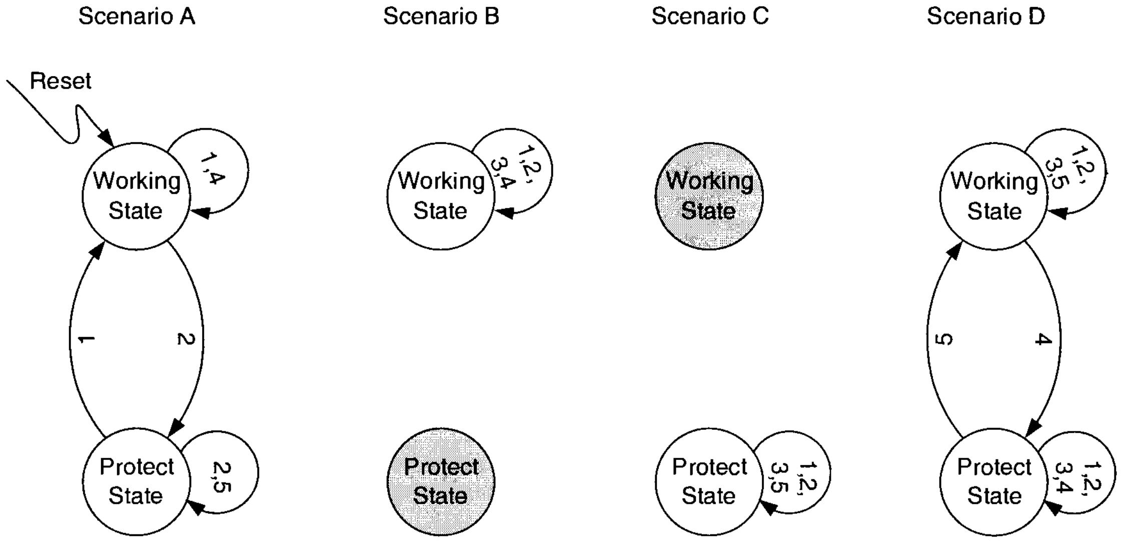


Table 4 shows the different scenarios that the system can be in . The states that are coloured grey indicate an inactive state; therefore, no transitions are possible. Table 5 shows the different external events that can be applied to the system.

Table 4 Four Different Scenarios

Scenario	Working Cable	Protect Cable
A	Connected	Connected
B	Connected	Disconnected
C	Disconnected	Connected
D	Disconnected	Disconnected

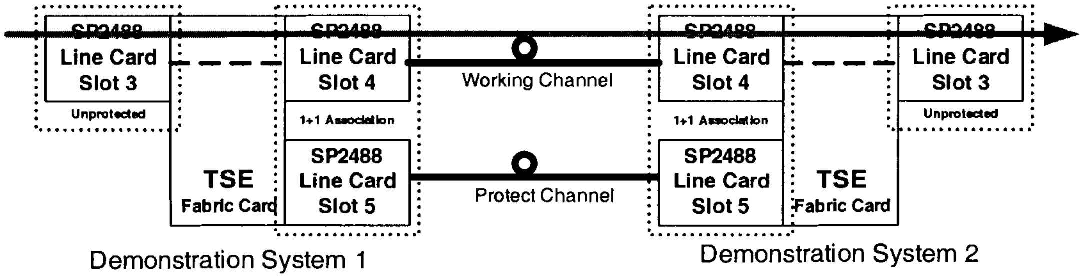
Table 5 External Events/Conditions

Event Number	Event Description
1	Protect Cable Disconnected
2	Working Cable Disconnected
3	Protect Channel Not Available
4	Protect Cable Connected
5	Working Cable Connected

The following example is given to illustrate how external events affect the algorithm of the Association Task. When both working and protect cables are connected, the Association Task will use Scenario A state machine. For this example, it is assumed that traffic is being sourced from the working port. If the working cable is disconnected, an event number 2 is detected by the Association Task, and the appropriate message is sent to the state machine algorithm. From Figure 13, an event number 2 will trigger a state change from the Working State to the Protect State. The control module will do the necessary configuration on the devices to enable the protect port to source the traffic. Since the working cable is now disconnected, the Association Task will change the active state machine to use the one shown in Scenario C of Figure 13.

To demonstrate how a 1+1 protection switchover scenario occurs, the example system in Figure 14 is used. Note that the diagram illustrates a unidirectional data path even though the protection switching applies to both directions. The dark blue arrow indicates the data path of the traffic flow.

Figure 14 1+1 Example System Before Switchover



Two demonstration systems are connected to each other using two pairs of fiber optic cables. One connection is designated as the working channel (using slot 4), and the other connection is designated as the protect channel (using slot 5). Traffic coming in from slot 3 of Demonstration System 1 is connected to slot 4. During the initial setup, slot 4 is used to source traffic going to slot 3 of Demonstration System 2.

Figure 15 shows how the two demonstration systems communicate with each other to execute an automatic protection switchover. The fiber optic cable connected to slot 4 is disconnected to cause a Loss of Signal (LOS) condition. The receiving SONET/SDH framer in slot 4 of Demonstration System 2 will detect the condition and transmit a Signal Fail error message to Demonstration System 1 using the K1 and K2 bytes. The K1 byte will contain the Signal Fail message and the K2 byte will indicate no active channel is currently using the protect channel. After transmitting the error message, Demonstration System 2 will wait for a Reverse Request message before moving further.

Upon receiving the K1/K2 bytes containing the error event and the protection channel to be used, the Association Task in Demonstration System 1 will execute a bridge command by creating a connection from slot 3 to slot 5. After the bridge command, Demonstration System 1 will send a message to Demonstration System 2 indicating a Reverse Request command (through the K1 byte) and indicate that the traffic from the working channel is rerouted to the protect channel. Demonstration System 1 waits for working channel to be bridged by Demonstration System 2 before moving further.

Demonstration Systems 1 and 2 will both select the protect channel to source traffic and the configuration of the system will changed as illustrated in Figure 16. The dark blue arrow indicates the data path of the new traffic flow.

Figure 15 1+1 Transition Diagram

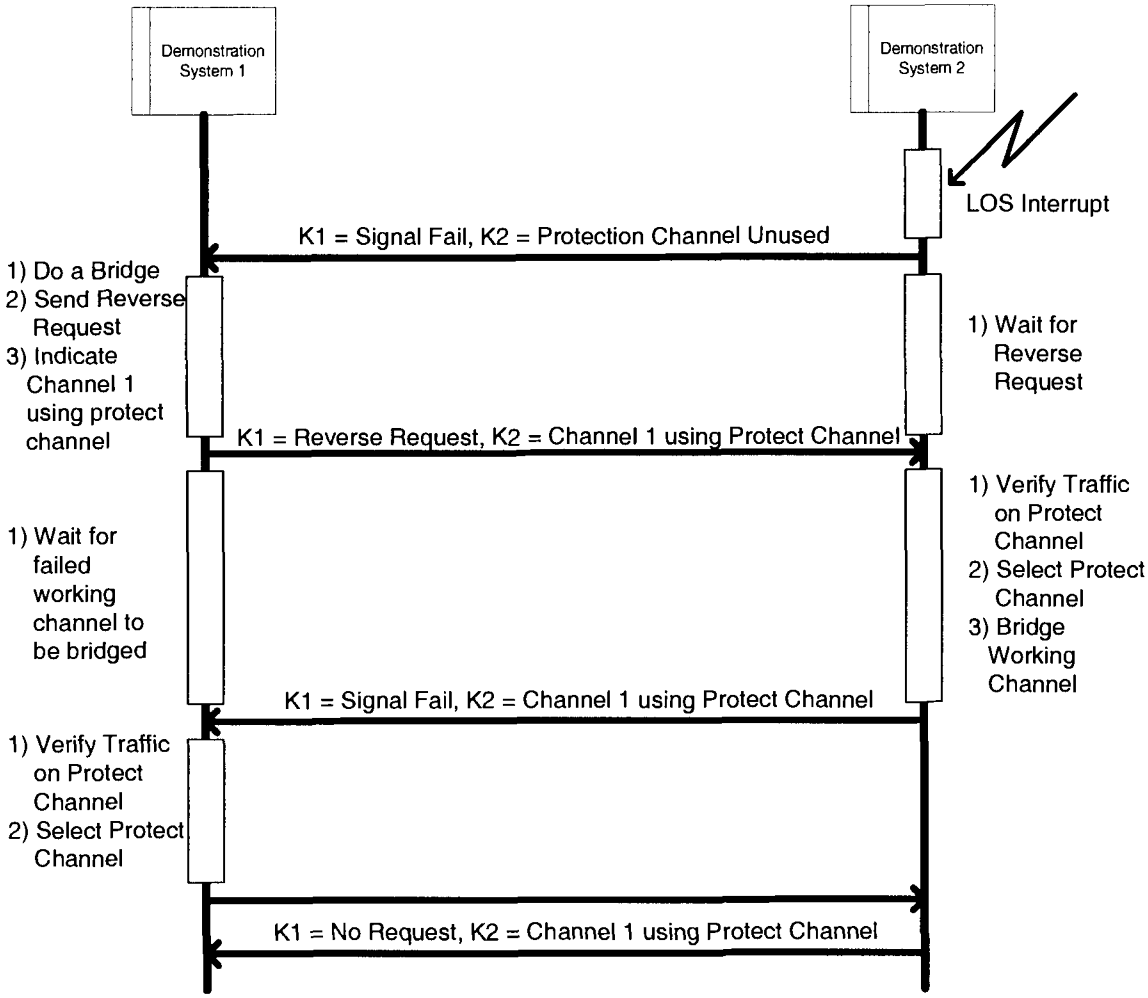
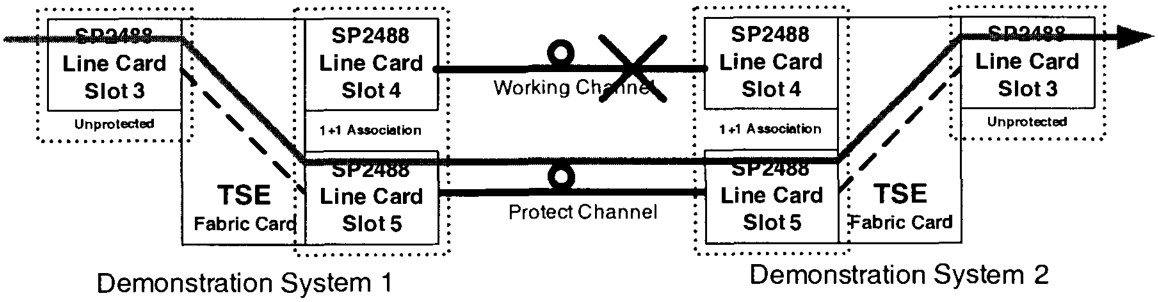

Demonstration System 1
Demonstration System 2
LOS Interrupt
K1 = Signal Fail, K2 = Protection Channel Unused
1) Do a Bridge
2) Send Reverse Request
3) Indicate Channel 1 using protect channel
1) Wait for Reverse Request
K1 = Reverse Request, K2 = Channel 1 using Protect Channel
1) Verify Traffic on Protect Channel
2) Select Protect Channel
3) Bridge Working Channel
1) Wait for failed working channel to be bridged
K1 = Signal Fail, K2 = Channel 1 using Protect Channel
1) Verify Traffic on Protect Channel
2) Select Protect Channel
K1 = No Request, K2 = Channel 1 using Protect Channel

Figure 16 1+1 Example System After Switchover

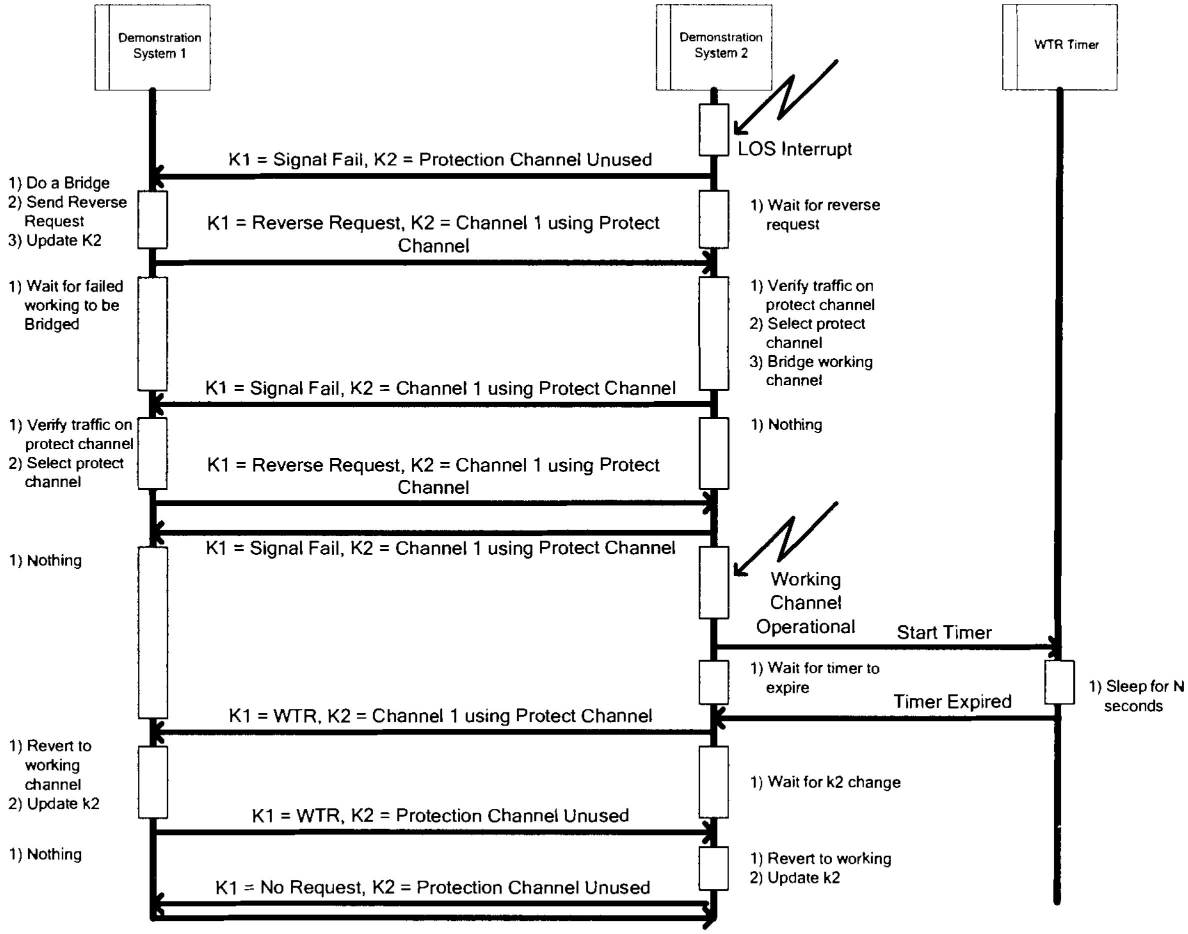


A 1:N protection switchover is similar to a 1+1 protection switchover. Since multiple working ports are protected by a single protect port, a priority scheme is required.

The working port with the lowest channel number is assigned the highest priority and the working port with the highest channel number is assigned the lowest priority. A Wait to Restore (WTR) timer is implemented in the 1:N protection algorithm to minimize the utilization of the protection port (so that other working channels can use the protect port in a failure event). When a working channel is restored, the WTR timer is initiated to monitor the health of the newly connected working channel. When the timer expires and there are no errors detected on the working channel, a switchover is initiated so that traffic from the protect port is reverted back to the working channel. After the switchover, the protect port is free to service any erred working channels.

To demonstrate how a 1:N protection switchover scenario occurs, the transition diagram shown in Figure 17 is used.

Figure 17 1:N Transition Diagram

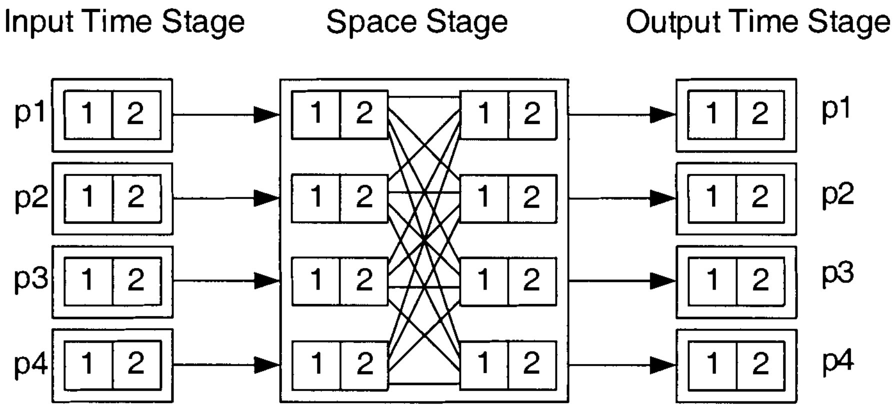

Demonstration System 1
Demonstration System 2
WTR Timer
LOS Interrupt
K1 = Signal Fail, K2 = Protection Channel Unused
1) Do a Bridge
2) Send Reverse Request
3) Update K2
K1 = Reverse Request, K2 = Channel 1 using Protect Channel
1) Wait for reverse request
1) Wait for failed working to be Bridged
1) Verify traffic on protect channel
2) Select protect channel
3) Bridge working channel
K1 = Signal Fail, K2 = Channel 1 using Protect Channel
1) Verify traffic on protect channel
2) Select protect channel
K1 = Reverse Request, K2 = Channel 1 using Protect Channel
1) Nothing
1) Nothing
K1 = Signal Fail, K2 = Channel 1 using Protect Channel
Working Channel Operational
Start Timer
1) Wait for timer to expire
Timer Expired
1) Sleep for N seconds
K1 = WTR, K2 = Channel 1 using Protect Channel
1) Revert to working channel
2) Update k2
1) Wait for k2 change
K1 = WTR, K2 = Protection Channel Unused
1) Nothing
1) Revert to working
2) Update k2
K1 = No Request, K2 = Protection Channel Unused

Executing the first part of a 1:N automatic protection switching is identical to the 1+1 automatic protection switching. When the working channel is operational again, demonstration system 2 will detect activity on the channel and determine if traffic can be sourced from it. To prevent a false detection of a bad channel, a WTR timer is started. The timer will run for a certain amount of time, after which the status of the working channel will be assessed again. When the timer expires and it is determined that the working channel is indeed operational, demonstration system 2 will send a message indicating WTR on the K1 byte and the current working channel using the protect channel. Demonstration system 1 will revert back to using the working channel and send a message to demonstration system 2 indicating a WTR on the K1 byte and indicating an unused protect channel. The demonstration system 2 will revert back to using the working channel and the two systems are now back to the original condition.

4.3 Open Path Algorithm

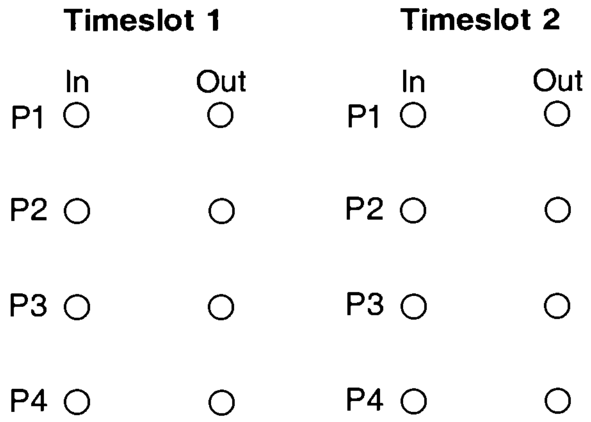
The Open Path Algorithm (OPA) is PMC Sierra's solution for creating a rearrangeably non-blocking connection through a Clos-based cross connect architecture. The OPA uses a bipartite graph approach to determine an optimal way of creating connections. To better understand the algorithm, a model of a simplified cross connect is used. Figure 18 shows a system with four input/output ports and each port contains two timeslots each.

Figure 18 System View of a Simple Cross Connect



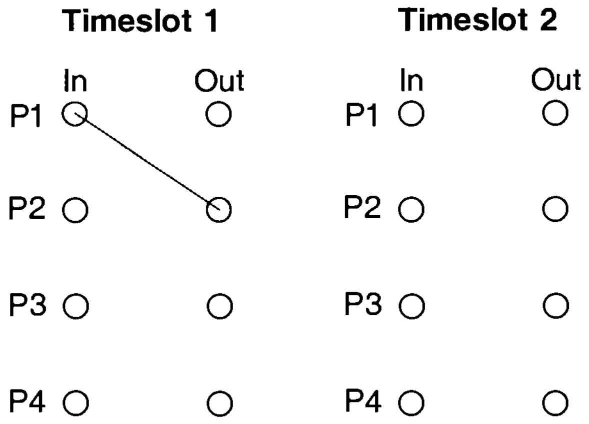
The OPA uses an internal model as shown in Figure 19 to represent the space stage of the system in Figure 18. The input and output time stages are used to do the necessary timeslot switching that the space stage cannot do.

Figure 19 OPA Internal Model



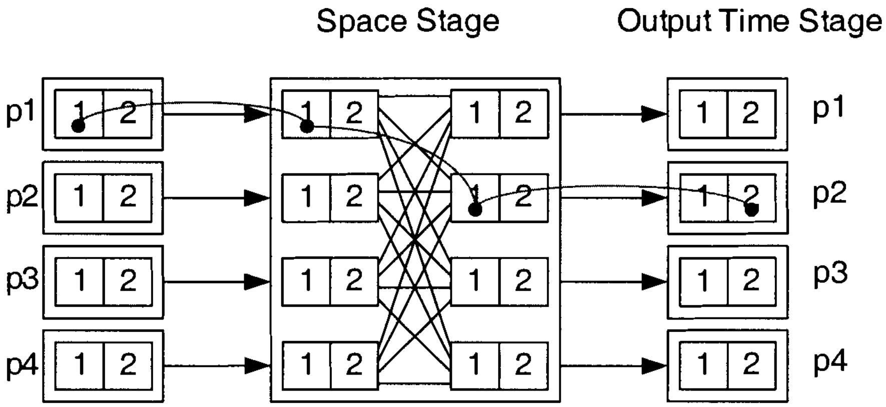
When a connection request is made from input port 1 timeslot 1 to output port 2 timeslot 2, the OPA will go through its internal list to see which space switch timeslot it can use. The algorithm starts from the first timeslot and moves on to the next until it finds a timeslot that has the desired input and output port free. The input and output timeslots are ignored at this point and will be handled by the input and output time stages. Figure 20 shows how the internal memory look after the first connection is created.

Figure 20 First Connection Created



Since timeslot 1 of the space stage is used, the time stages will need to align the input and output timeslot for the connection to be made. Since the desired input timeslot is already timeslot 1, the input time stage for timeslot 1 will be set for a straight through connection. The desired output timeslot is timeslot 2. Since the data from the space switch comes out of timeslot 1, the output time stage needs to map timeslot 1 to timeslot 2. Figure 21 shows the system view with the first connection created.

Figure 21 System View of the First Connection



As the number connections increases, there is an increasing chance that going through all the available timeslots will not yield a free input and output port. When a connection cannot be created despite the fact that there are available input and output timeslots left, a blocking scenario occurs. To produce a blocking scenario, the following connections listed in Table 6 is created. Figure 22 shows the internal memory of the OPA after 6 connections are made.

Table 6 New Connections

Connection No	Input Port	Input Timeslot	Output Port	Output Timeslot
1	1	1	2	2
2	2	1	3	2
3	3	1	4	2
4	2	2	1	1
5	3	2	2	1
6	4	2	3	1
7	1	2	1	2

Figure 22 Internal Model After 6 Connections

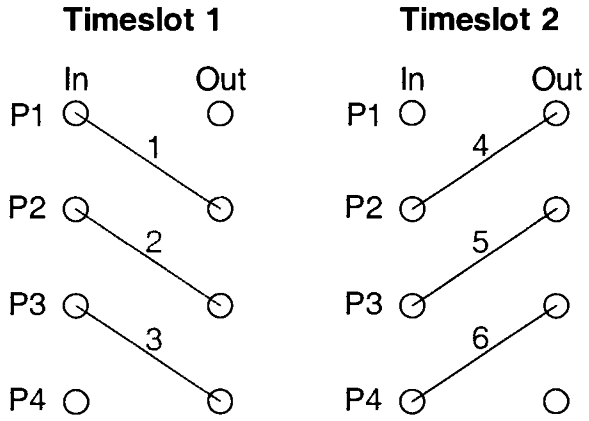
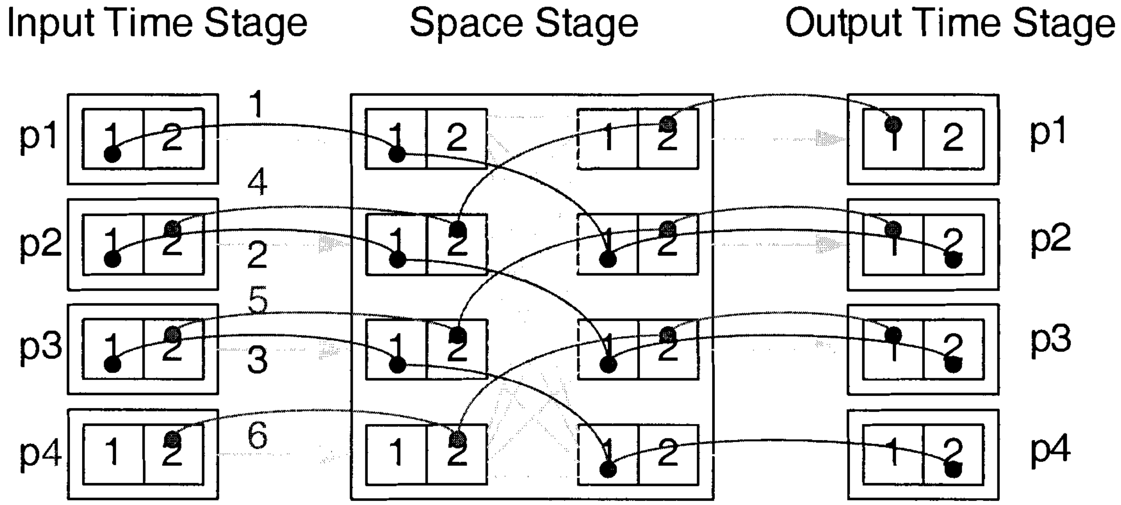


Figure 23 System Model After 6 Connections



From the diagram above, it can be seen that there are no timeslots that can accommodate connection 7 even though timeslot 2 of input port 1 and output port 1 are unused. In order to get around the blocking scenario, a rearrangement has to be made. This means that some existing connections need to be moved around in order to accommodate the new connection (connection 7). To facilitate the rearrangement, the OPA merges its internal model of the space switch to get a diagram shown in Figure 24. Note that a connection made in timeslot 1 of the space switch is marked as red and a connection made in timeslot 2 is marked as blue.

Figure 24 Merged Timeslots

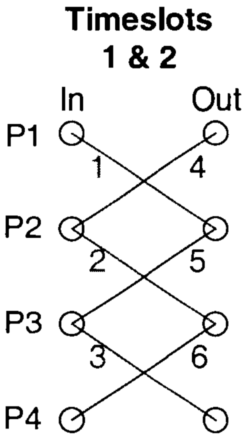
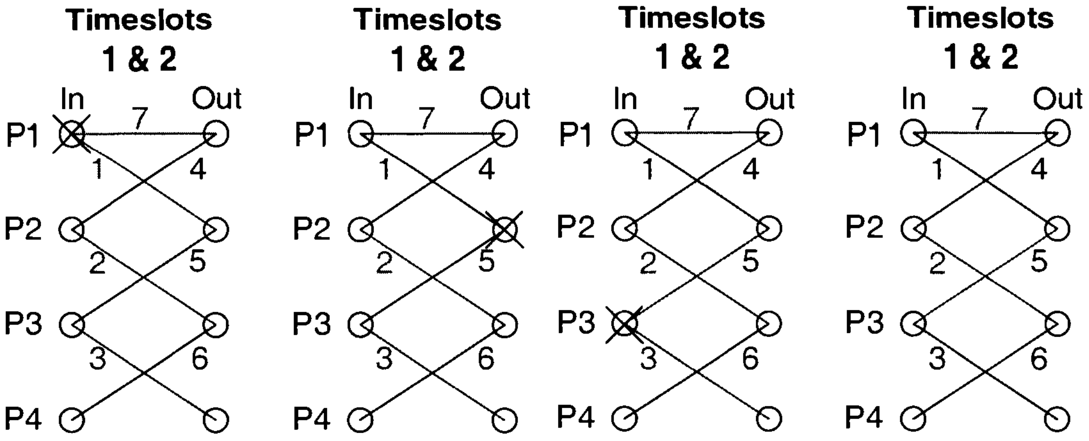


Figure 25 shows the sequence in which rearrangement occurs to resolve a blocking scenario.

Figure 25 Rearrangement Steps



The first timeslot is used to create connection 7. Since input port 1 is already used in timeslot 1 by connection 1, a conflict arises between connection 1 and connection 7. The OPA will rearrange connection 1 to use timeslot 2 instead of timeslot 1. After this rearrangement, there will no longer be any conflict between connections 1 and 7. However, a conflict arises between connection 1 and connection 5. Connection 5 is rearranged to utilize timeslot 1 in order to resolve the conflict with connection 1. The process continues until there is no more conflict.

After the blocking scenario is resolved, the OPA will separate the timeslots in order to determine the new space switch mapping for each timeslot. Figure 26 shows the separated space switch mapping and Figure 27 shows the rearranged mapping on the system view.

Figure 26 Separated Space Switch Mapping

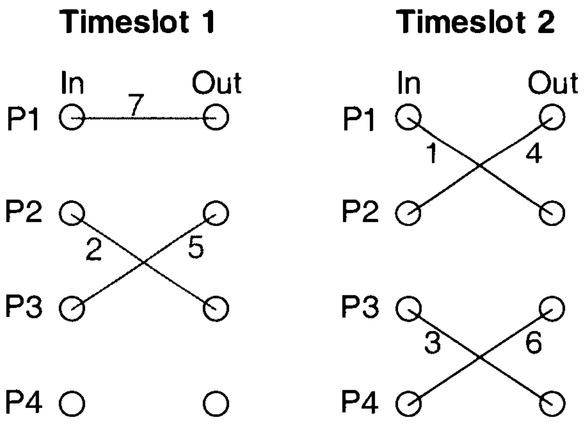
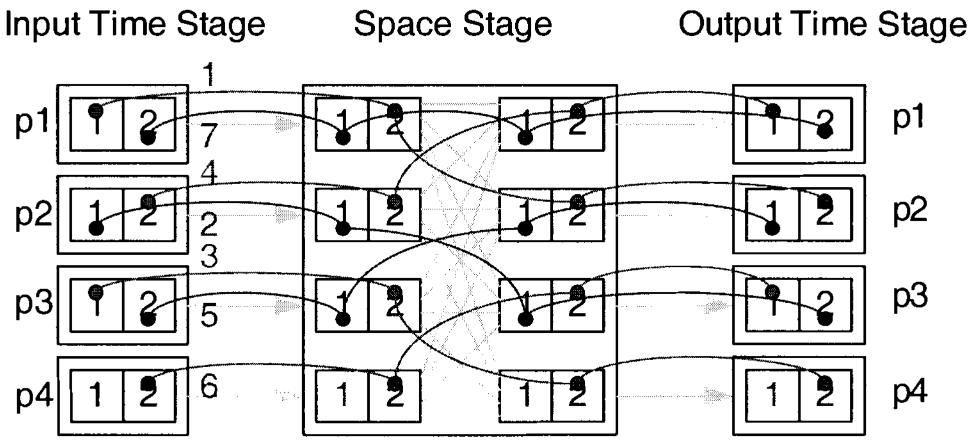


Figure 27 Rearranged Connections - System View



There were a total of 3 rearrangements done to accommodate connection 7.

4.4 VxWorks Real Time Operating System

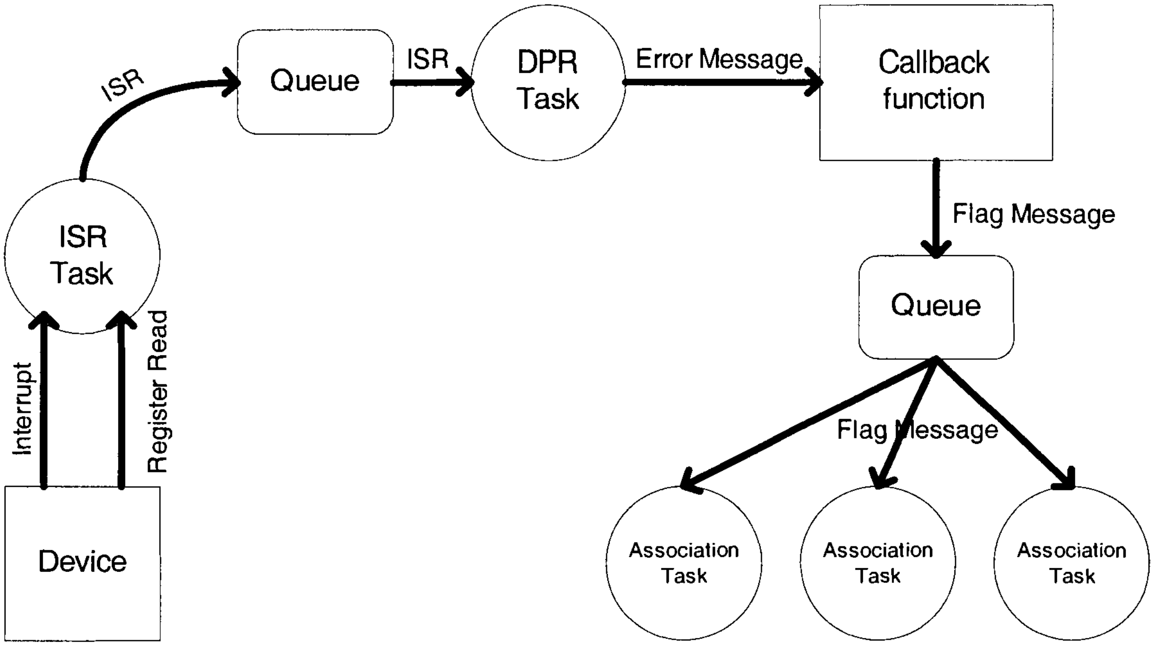
VxWorks is the operating system used for the demonstration system. Since VxWorks is a real time OS, it is possible to perform time critical operations such as protection switching. The Hardware Control software relies on the OS's kernel to provide immediate interrupt notifications on timers and device error events. Priorities of different threads are also controlled by the Hardware Control software to guarantee adequate CPU resource allocation on critical operations.

4.5 Control Module

The Control Module manages all the communications between the device drivers, APS Module, OPA, and RTOS. Aside from instantiating all the modules in the system, the Control Module handles the interrupt propagation, and device configuration.

Interrupt Service Routines (ISR) and Delayed Processing Routines (DPR) are implemented to properly propagate device interrupts to the APS Module. Figure 28 shows how interrupts are propagated to the respective Association Tasks. When the interrupt pin of the microprocessor gets asserted, an ISR is called by the operating system. ISRs accesses the device registers to determine what type of interrupt was triggered. The nature of the interrupt is then passed to a queue for further processing of the DPR. When an ISR is executing, all other interrupts are masked by the operating system. For this reason, the RTOS gives the ISR a very high priority to guarantee quick execution. The ISR also needs to be short and efficient; hence, all further processing are passed to the DPR using a message queue. The DPR determines the type of interrupt and sends a message to the appropriate Association Task (see section 4.2) for further processing.

Figure 28 Interrupt Flow Diagram



When the APS Module determines that a protection switchover needs to occur, the Control Module passes all relevant information to the OPA. The OPA will calculate the new device settings and returns these settings to the Control Module. The Control Module uses the device drivers to write the new device settings.

5 TEST RESULT

The SONET/SDH specification states that a protection mechanism must successfully complete within 50 ms of a detected failure. To determine whether the software implementation meets the specification, the test setup shown in Figure 29 and Figure 30 are created.

Figure 29 Test Setup for 1+1 Protection

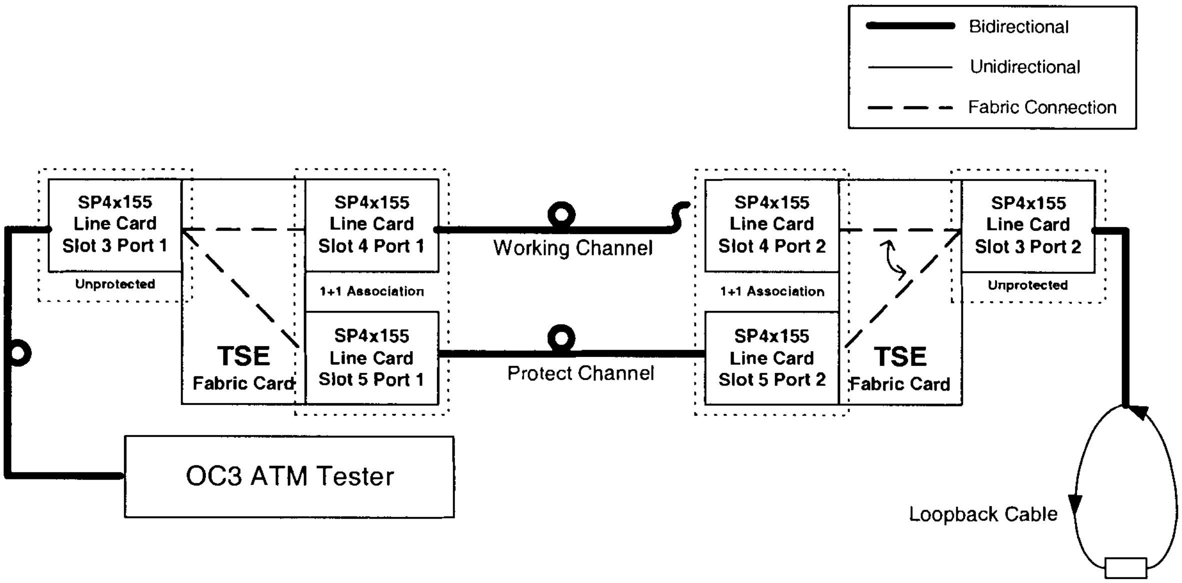
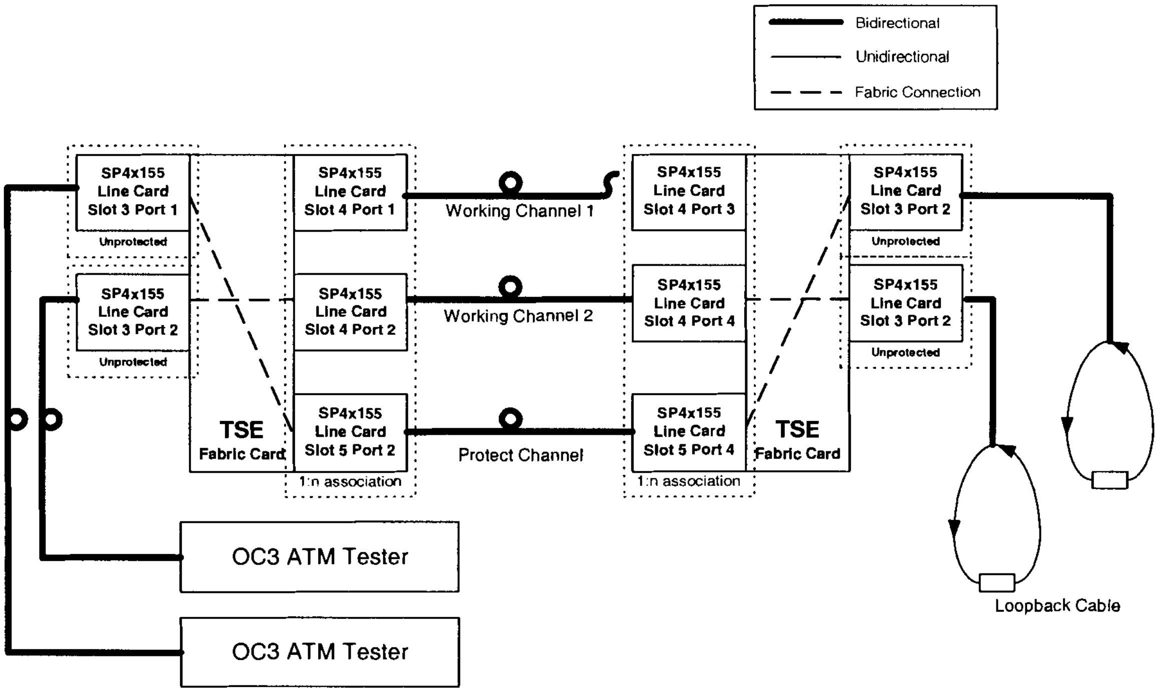

Bidirectional
Unidirectional
Fabric Connection
SP4x155 Line Card Slot 3 Port 1
Unprotected
TSE Fabric Card
SP4x155 Line Card Slot 4 Port 1
1+1 Association
SP4x155 Line Card Slot 5 Port 1
Working Channel
Protect Channel
SP4x155 Line Card Slot 4 Port 2
1+1 Association
SP4x155 Line Card Slot 5 Port 2
TSE Fabric Card
SP4x155 Line Card Slot 3 Port 2
Unprotected
OC3 ATM Tester
Loopback Cable

Figure 30 Test Setup for 1:N Protection


Bidirectional
Unidirectional
Fabric Connection
SP4x155 Line Card Slot 3 Port 1
Unprotected
SP4x155 Line Card Slot 3 Port 2
Unprotected
TSE Fabric Card
SP4x155 Line Card Slot 4 Port 1
SP4x155 Line Card Slot 4 Port 2
SP4x155 Line Card Slot 5 Port 2
1:n association
Working Channel 1
Working Channel 2
Protect Channel
SP4x155 Line Card Slot 4 Port 3
SP4x155 Line Card Slot 4 Port 4
SP4x155 Line Card Slot 5 Port 4
1:n association
TSE Fabric Card
SP4x155 Line Card Slot 3 Port 2
Unprotected
SP4x155 Line Card Slot 3 Port 2
Unprotected
Loopback Cable
OC3 ATM Tester
OC3 ATM Tester

The SONET/SDH tester has a Service Disrupt Monitoring option that determines how long a data pattern loss was detected. A total of 10 tests were conducted for each protection mechanism to determine the performance of the switchover mechanism. Table 7 shows the results of the tests.

Table 7 Switchover Latency Result

Test	Switchover Latency (ms)	
	1+1 Protection	1:N Protection
1	24.3	23.7
2	22.4	24.3
3	23.2	24.3
4	22.6	24.9
5	23.6	24.6
6	23.1	23.5
7	24.7	22.7
8	22.9	24.8
9	22.6	24.5
10	24.7	24.4

From the table above, the performance between the 1+1 and 1:N both meet the maximum switchover latency of 50 ms. The variation in switchover latency between each test can be attributed to how fast a cable is pulled out (to cause a LOS condition) and the polling period of the Association Task's scheduler.

Even though the switchover latency numbers are within the specification, the numbers are still relatively high. For a SONET/SDH network, the 50 ms requirement applies to the whole network. Most networks are comprised of multiple nodes and there are scenarios where each node in the network need to change their configuration settings. For this test case, a network of two nodes is implemented. If a network of 10 nodes is implemented, each node will only have a maximum allowable switchover latency of 5 ms. For such a network, the performance of the software does not meet the performance requirement as indicated by the SONET/SDH specification. Using faster processors will decrease the switchover latency in order to meet more stringent requirements.

BIBLIOGRAPHY

[1] Bell Communications Research – SONET Transport Systems: Common Generic Criteria, GR-253-CORE, Issue 2, Revision 2, January 1999.

[2] ITU, Recommendation G.707 – "Digital Transmission Systems – Terminal equipments – General", March 1996.

[3] PMC Sierra. PMC-2002156 "TSE/TBS Open Path Algorithm Application Note", Issue 1, April 2002.

[4] PMC Sierra. PMC-1990822 "SPECTRA 4x155 ASSP Telecom Stnadard Product Data Sheet, Issue 5, September 2002.

[5] PMC Sierra. PMC-1990821 "SPECTRA-2488 Telecom Standard Product Data Sheet", Issue 4, November 2001.

[6] PMC Sierra. PMC-1991257 "TelecomBus Serializer Data Sheet", Issue 7, November 2001.

[7] PMC Sierra. PMC-1991258 "TSE Transmission Switch Element Datasheet", Issue 7, November 2001.

[8] PMC Sierra. PMC-1990823 "Multi-Service Access Device For Channelized Interfaces Telecom Standard Product Data Sheet", Issue 4, December 2001.

[9] Antonio Nucci, Nina Taft, Chadi Barakat, Patrick Thiran, "Controlled Use of Excess Backbone Bandwidth for Providing New Services in IP-over-WDM Networks", IEEE Journal – Optical Communications and Networking Series, November 2004