Performance Enhancements in Next Generation Multi-Service Broadband Ring Networks

Charlie Kawwas

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

April 2007

Canada

ABSTRACT

Performance Enhancements in Next Generation Multi-Service Broadband Ring Networks

Charlie Kawwas, PhD
Concordia University, 2007

Multi-service broadband networks are forming the basis of a converged service and transport infrastructure. These next generation networks use the Internet Protocol (IP) and Multi-Protocol Label Switching (MPLS) to deliver service and network convergence. The dominant topology, especially in the Edge and Core of such networks, is the ring. Resilient Packet Ring (RPR) has emerged as a key technology and IEEE standard (IEEE 802.17) in delivering Edge and Core transport convergence enabling high utilization, resiliency and fairness. However, the current standard has several drawbacks such as slow convergence, severe oscillation, inefficient utilization of the dual-ring and inconsistent class of service handling. We examine multiple classes of IP services in the Edge and Core networks connected over ring topologies and review the pros and cons of RPR versus alternative solutions. We devise a new technique to increase the performance of RPR and optimize the use of the dual-ring bandwidth while distinguishing between various classes of service. We refer to this technique as Flow-based Priority Distribution (FPD); we take into account the total bandwidth available on both counter-rotating rings, and allocate flow bandwidth based on availability, cost and traffic prioritization. We also introduce a new transit buffer design, referred to as 3-TB design, which provides consistent service differentiation across the network. We analyze the performance of our techniques, and compare the results with current and recent techniques in the RPR research. We recommend the combination of FPD and 3-TB by Service Providers for evolving their networks to a packet-based infrastructure and delivering new IP-based services.

ACKNOWLEDGMENTS

DEDICATION

This thesis is dedicated in loving memory of my father Bishara Kawwas.

TABLE OF CONTENTS

Performance Enhancements in Next Generation Multi-Service Broadband Ring Networks

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\oplus$	Kronecker sum
$\otimes$	Kronecker product
$1/\alpha$	Mean time spent in the On state
α_{Pi}	Average number of transit buffers a packet traverses for a CoS P_i
$1/\beta$	Mean time spent in the Off state
Δ	Timeslot length in ms
φ	Variance coefficient of the fractional Brownian motion process
$\Gamma(x)$	The gamma function
$\lambda_{1\text{-}TB}$	Mean arrival rates for 1-TB (PTQ)
$\lambda_{2\text{-}TB,PTQ}$	Mean arrival rates for 2-TB PTQ
$\lambda_{2\text{-}TB,STQ}$	Mean arrival rates for 2-TB STQ
$\lambda_{3\text{-}TB,PTQ}$	Mean arrival rates for 3-TB PTQ
$\lambda_{3\text{-}TB,STQ}$	Mean arrival rates for 3-TB STQ
$\lambda_{3\text{-}TB,TTQ}$	Mean arrival rates for 3-TB TTQ
$\lambda_{client,Pi}$	Mean arrival rate for client traffic with CoS P_i
λ_i	Mean arrival rate during state s_i
Λ	Arrival rate matrix
π	Equilibrium probability vector
ρ	Traffic throughput or offered load
ρ_k	Throughput for traffic with priority k
ρ_{r0}	Ringlet $r0$ utilization
ρ_{r1}	Ringlet $r1$ utilization
$\rho_{x,P_i}^{ringlet}$	Throughput on $ringlet$ for node x with traffic CoS P_i
ρ_{Th}	Throughput threshold
σ^2	The variance of a random sample

τ_b	Feedback delay from the link to the source
τ_f^i	Forward delay from source node i
ω_i	Weight of a flow from node i
ζ_k	Auto-correlation function at lag k
a	The shape parameter of a Pareto probability density function
b	The scaling parameter of a Pareto probability density function
b_n	Bit-rate during frame n
B_w	Link bandwidth
$\{B_t\}$	A Brownian motion stochastic process for $t \geq 0$
$\{fB_t\}$	A fractional Brownian motion process for $t \geq 0$
C	Ringlet capacity in Mbps
$C_{x,P_i}^{ringlet}$	*ringlet* condition that traffic with priority P_i is destined to node x
$d_{P_i,Th}$	Delay threshold for traffic CoS P_i
$d_{x,P_i}^{ringlet}$	Delay on *ringlet* to node x with traffic CoS P_i
d_f	Total link delay for flow f from source to destination
d_i	Mean sojourn time in state s_i
$\underline{D}$	Denotes equality in distribution
e	The unit column $= (1, 1)^T$
$E[X]$	Expected value (or mean) of random variable X
$f(cos, i, j)$	Bandwidth required by traffic flow f with priority cos from node i to node j
$F_i(n)$	Estimated fair rate for link i during the nth iteration
G	Generalize service time for a packet in a node
H	The Hurst parameter
I	Identity matrix
$K_i(n)$	Estimated number of unbounded flows in link i during the nth iteration
L_{max}	Maximum size of a packet
m	Mean arrival rate $E[N_t]$

M	Number of states in space S
N	Number of RPR nodes on the ring
N_t	Number of arrivals in the time interval $(0,t)$
P_i	Class of Service (or traffic priority) i
P_T	Total number of client and control traffic queues
$P(s_i)$	Probability of being in state s_i
$q_{x,Pi}$	Probability of node 0 transmitting to node x traffic with priority P_i
Q_{max}	Maximum quota allowed for any node to transmit on the ring
$Q(t)$	Buffer occupancy at time t
Q_0	Target value of $Q(t)$
r_i	Transition rate from state s_i
$r_{i,j}$	Transition rate from state s_i to s_j
R	Mean residual service time
$\boldsymbol{R}$	Infinitesimal generator matrix
$R(t)$	Adjusted sending rate of a flow at its source
R_{max}	Maximum allowed sending rate of a flow
s_i	State i
S	State space $\{ s_1, s_2, \ldots, s_M \}$
$\overline{S_{P_i}}, \overline{S_{P_i}^2}$	1st and 2nd moment packet sending time, respectively
$T_{1\text{-}TB,Pi}$	Average packet queuing delay for 1-TB with traffic priority P_i
$T_{2\text{-}TB,Pi}$	Average packet queuing delay for 2-TB with traffic priority P_i
$T_{3\text{-}TB,Pi}$	Average packet queuing delay for 3-TB with traffic priority P_i
T_{link}	Latency between two nodes (i.e. latency per link)
T_{RD}	Round trip delay
W_i	Waiting time of a packet with priority i
$\bar{x}$	The mean of a random sample
X_n	Random variable X that defines S at time n
Y_n	White Gaussian Noise in frame n
Z_t	A normalized fractional Brownian motion process

LIST OF ABBREVIATIONS

ABA	Adaptation Bandwidth Allocation
AC	Attachment Circuit
ADPCM	Adaptive Differential Pulse Coded Modulation
ADSL	Asymmetric Digital Subscriber Line
AF	Assured Forwarding
AM	Aggressive Mode
AR	Auto-Regressive model
ARMA	Auto-Regressive Moving Average
ARIMA	Auto-Regressive Integrated Moving Average
AS	Autonomous Area
AT&T (at&t)	American Telegraph and Telephone
ATM	Asynchronous Transfer Mode
BGP	Broader Gateway Protocol
BIST	Buffer Insertion/Self-Token
BIR	Buffer Insertion Ring
BPC	Block Packet Count
bps	Bits per second
BRAS	Broadband Remote Access Server
BSD	Berkeley Software Distribution
BW	Bandwidth
CBR	Constant Bit Rate
CE	Customer Edge
CIR	Committed Information Rate
CM	Conservative Mode
CO	Central Office
CoS	Class of Service
CPP	Compound Poisson Process

CR-LDP	Constraint-based Routing Label Distribution Protocol
CRC	Cyclic Redundancy Check
CS	Class Selector
DAR	Discrete Auto-Regressive model
DBRR	Distributed Bandwidth Reallocated in Rings
DF	Default Forwarding
DLCI	Data Link Connection Identifier
DNS	Domain Name Server
DiffServ	Differentiated Services
DS	Differentiated Services
DSCP	Differentiated Services Code Point
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DTB	Dual Transit Buffer
DTV	Digital Television
DVB	Digital Video Broadcast
DVD	Digital Video Disk
DVMRP	Distance-Vector Multicast Routing Protocol
DVSR	Distributed Virtual-time Scheduling in Rings
DWDM	Dense Wave Division Multiplexing
E-LER	Egress-Label Edge Router
ECN	Explicit Congestion Notification
EF	Expedited Forwarding
EGP	Exterior Gateway Protocol
EIR	Excess Information Rate
Eth	Ethernet
EVDO	Evolution-Data Optimized
F-ARIMA	Fractional Auto-Regressive Integrated Moving Average
FBM	Fractional Brownian Motion

FCS	Frame Check Sum
FDDI	Fiber Distributed Data Interface
FEC	Forward Equivalency Class
FPD	Flow-based Priority Distribution
FR	Frame Relay
FRR	Fast Re-Route
FTN	FEC-to-NHLFE
FTP	File Transfer Protocol
Gbps	Gigabits per second
GE	Generalized Exponential
GbE	Gigabit Ethernet
GFP	Generic Framing Protocol
HDTV	High-Definition Television
HEC	Header Error Check
HOL	Head-Of-Line
HS IA	High-Speed Internet Access
HSDPA	High-Speed Download Packet Access
I/F	Interface
I-LER	Ingress-Label Edge Router
IA	Ingress Aggregate
ICMP	Internet Control Message Protocol
IDC	Index of Dispersion for Counts
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IGMP	Internet Group Management Protocol
IGP	Interior Gateway Protocol
IHL	IPv4 Header Fields
ILM	Incoming Label Map
iid	Independent and Identically Distributed

IntServ	Integrated Services
IP	Internet Protocol
IP PSN	Internet Protocol Packet Switched Network
IPP	Interrupted Poisson Process
IPTV	Internet Protocol Television
ISO	International Organization for Standardization
ITU	International Telecommunication Union
L2/3 Switch	Layer 2 and Layer 3 Switch
LAN	Local Area Network
LCFA	Low Complexity Fairness Algorithm
LCP	Lowest Cost Path
LDP	Label Distribution Protocol
LER	Label Edge Router
LRD	Long-Range Dependent model
LSP	Label Switched Path
LSR	Label Switching Router
MAC	Media Access Control
MAN	Metropolitan Area Network
Mbps	Megabits per second
MEF	Metro Ethernet Forum
MC	Markov Chain
MM	Markov Modulated
MMDP	Markov Modulated Deterministic Process
MMPP	Markov Modulated Poisson Process
MMS	Multimedia Messaging Service
MOSPF	Multicast Open Shortest Path First
MPEG	Motion Picture Experts Group
MPLS	Multi-Protocol Label Switching
MPLS-TE	Multi-Protocol Label Switching – Traffic Engineering

MR	Minimum Rate
ms	Milli-second
MTB	Mono Transit Buffer
NHLFE	Next Hop Label Forwarding Entries
NP	Network Processor
NTT	Nippon Telegraph and Telephone
OAM	Operations, Administration and Management
OL	Over-Load
OSI	Open System Interconnection
OSPF	Open Shortest Path First protocol
PCM	Pulse Coded Modulation
PCP	Priority-based Congestion Prevention
PD	Proportional Derivative
PDA	Packet Delay Analysis tool
PDH	Plesiochronous Digital Hierarchy
PDU	Protocol Data Unit
PE	Provider Edge
PHB	Per Hop Behavior
PHY	Physical media
PIM	Protocol Independent Multicast
PIM-DM	Protocol Independent Multicast – Dense Mode
PIM-SM	Protocol Independent Multicast – Sparse Mode
PLR	Point of Local Repair
PLP	Packet Loss Probability
PMPP	Pareto Modulated Poisson Process
PON	Passive Optical Network
PPP	Point-to-Point Protocol
pps	Packets per second
PSN	Packet Switched Network

PSTN	Public Switched Telephone Network
PT	Protocol Type
PTQ	Primary Transit Queue
PW	Pseudo-Wire
PWE3	Pseudo-Wire Emulation Edge to Edge
QoS	Quality of Service
RG	Residential Gateway
RFC	Request For Comments
RIAS	Ring Ingress Aggregated with Spatial Reuse
RIP	Routing Information Protocol
RPR	Resilient Packet Ring
RSVP	Resource ReSerVation Protocol
RSVP-TE	Resource ReSerVation Protocol – Traffic Engineering
RTP	Real-Time Protocol
RV	Random Variable
SBC	Southwestern Bell Corporation
SDU	Service Data Unit
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SMS	Short Message Service
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SONET/SDH	Synchronous Optical NETwork / Synchronous Digital Hierarchy
SP	Service Provider
SPP	Switched Poisson Process
SRD	Short-Range Dependent model
STB	Set-Top Box
STQ	Secondary Transit Queue
TB	Transit Buffer

TCP	Transmission Control Protocol
TFTP	Trivial File Transfer Protocol
TE	Traffic Engineering
TELNET	TELetype NETwork
TES	Transform Expand Sample model
ToS	Type of Service
TTL	Time To Live
TTQ	Tertiary Transit Buffer
UDP	User Datagram Protocol
UL	Under-Load
VBR	Variable Bit Rate
VDQ	Virtual Destination Queue
VDSL	Very high bit-rate Digital Subscriber Line
Ver (or v)	Version
VLAN	Virtual Local Area Network
VoD	Video on Demand
VoIP	Voice over IP
VLAN	Virtual Local Area Network
VSQ	Virtual Source Queuing
WAN	Wide Area Network
WAP	Wireless Application Protocol
Wi	Wireless
WiMAX	Worldwide Interoperability for Microwave Access
XDR	eXternal Data Representation
μs	Micro-seconds

CHAPTER 1 - Introduction

1.1 Multi-Service Broadband Networks

The Internet is rapidly evolving into a vital part of our lives both for consumers and businesses. In fact, the Internet is becoming more crucial to the business community than the telephone network [1]. As new services and applications evolve leveraging the Internet infrastructure and protocols, Service Providers (SP's) are struggling to deliver such services and requirements. This is becoming more challenging as they are attempting to maintain their existing infrastructure and services. Figure 1-1 illustrates a high-level architecture of an SP's network [2].
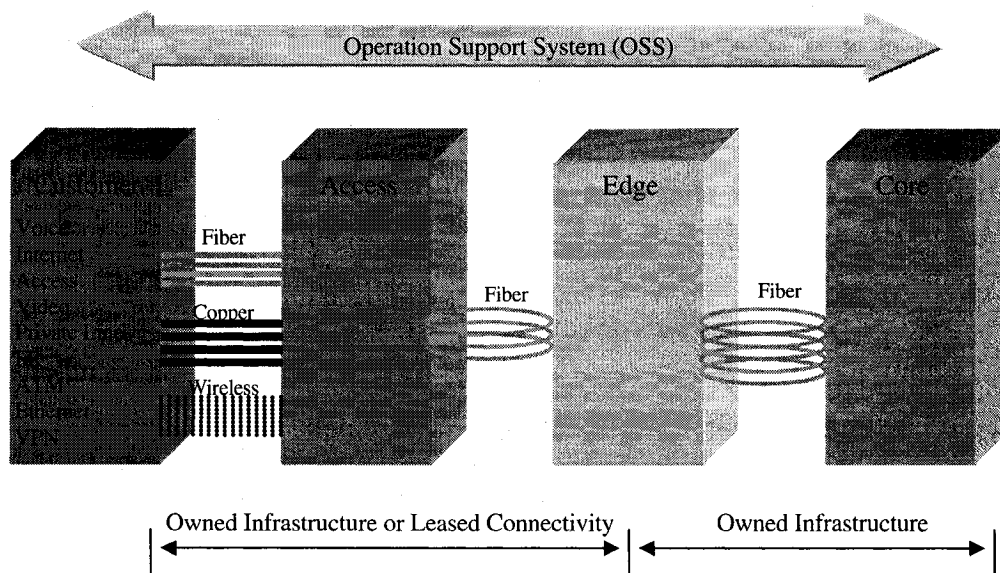
Figure 1-1: Service Provider's Network Architecture

1.1.1 Challenges in Broadband Networks

The introduction of new services based on IP, such as Voice over IP, IP TV and Video Streaming, etc., delivered over multi-service broadband networks requires significant changes to the architecture and implementation of current broadband networks. These modifications include key capabilities such as multicast, QoS, security, authentication, service identification, availability and scalability.

1

A key challenge is to provide the required capabilities for these services and applications to meet their service level agreements. However, the challenge becomes significantly tougher as the SP's need to maintain the circuit-based revenue-generating legacy services while evolving the network to the new packet-based services. Hence, new technologies and techniques that can deliver today's circuit-based services and provide a migration path to the next generation packet-based services are regarded as the foundation for broadband convergence.

1.1.2 Broadband Convergence

There are multiple levels of convergence in the Service Provider (SP) network, namely: Service, Network Connectivity and Protocol, as shown in Figure 1-2. It is clear that IP has won the networked desktop; hence, convergence at the service level is achieved via IP to carry Voice, Video and Data services. At the Network Connectivity and Protocol levels, we need to consider separately the Core, Edge and Access portions of the network. For the Core, MPLS over an Optical infrastructure has emerged as the key technology to deliver convergence for IP, ATM and FR broadband services [2, 3].

This is not the case in the rest of the network, as many technologies continue to provide valid options for Network Connectivity today, such as SONET/SDH, DWDM and Ethernet in the Edge. Two options are emerging as valid for Network Connectivity convergence in the Edge, a pure multi-service packet play based on an Ethernet Layer 2 across the network, and Resilient Packet Ring (RPR) in the Edge and Core of the network. The first option is based on Ethernet only approach and other services are emulated. The second option is a multi-service circuit and packet Network Connectivity, where RPR and Generic Framing Protocol (GFP) are used to transport packet traffic over existing circuit-based infrastructure [4]. Protocol encapsulation and convergence are emerging across both options with Pseudo-wire Emulation Edge to Edge (PWE3) as the Service Provider protocol of choice. PWE3 enables Service Providers to deliver multiple services across a common Packet Switched Network (PSN), each as emulation in a tunnel over an IP PSN [5].

2
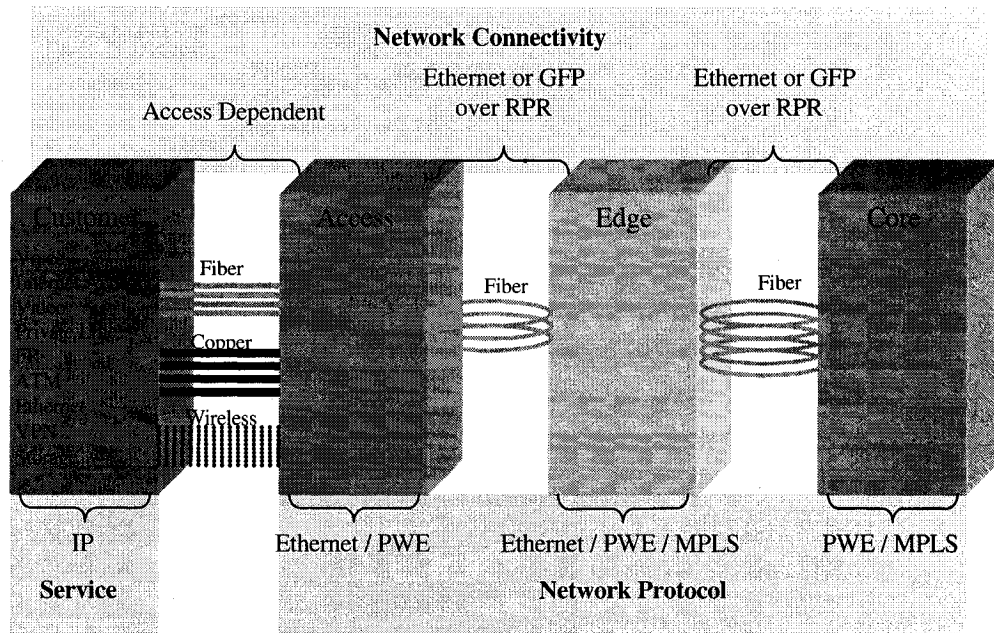
Figure 1-2: Service, Network Connectivity & Protocol Convergence

In the Access part of the network, Network and Protocol convergence is more challenging. PDH, DSL, PON, 802.11, HSDPA, etc., are excellent candidates for Network Connectivity. We expect to have these technologies play and evolve for capital cost, operational costs and regulatory constraints. As for the Network Protocol, ATM is migrating to Ethernet with potential evolution to MPLS & PWE3. We expect the Network Protocol to converge across the network to PWE3 with a converged PSN, while Network Connectivity to remain specific in the Access due to various network, cost and topology constraints.

1.2   Recent Research and Motivation

The dominant topology in the Edge and Core segments of the network is a ring, as shown in Figure 1-2. Typically, in metro networks, fiber is physically laid down in ring configurations. A key reason is resiliency and fault tolerance, i.e. network connectivity is maintained to all nodes even with any single link failure. Moreover, rings have lower deployment costs as opposed to mesh or star topologies [6].

RPR is an IEEE 802.17 standard, which has been defined recently to deliver high bandwidth utilization, differentiated Classes of Service (CoS) and fairness. It plays a key role in enabling broadband convergence especially in the Edge and the Core of the network. However, RPR has several drawbacks [7]:

(a) The slow convergence of the fairness algorithm

(b) Permanent oscillations of the fairness algorithm

(c) Long ring access delay

(d) Total (i.e. dual-ring) bandwidth inefficient utilization, and

(e) Inconsistent class of service traffic handling.

The first three items have been researched quite extensively in the past four years with proposed enhancements and solutions. Drawbacks (a) and (b) are related to the RPR fairness algorithm. In [8], a new bandwidth allocation scheme, Adaptation Bandwidth Allocation (ABA), addressed the oscillation and performance issues in RPR. In [9], the RPR performance deficiencies of fair rate estimates and their distribution on the ring were discussed, and a scheme for improvement was proposed. In [6], a new scheme called Distributed Virtual-time Scheduling in Rings (DVSR) proposed a simple lower bound on temporally and spatially aggregated virtual time. Another research focused on providing a fair medium sharing using an access control scheme called Virtual Source Queueing (VSQ) [10]. In [11], this research attempted to address the fairness algorithm by estimating the number of unbounded flows in each link. Also, research [12] addressed the fairness limitations using a scheme called Low Complexity Fairness Algorithm (LCFA).

Drawback (c) is another key area of research in RPR focused on delay performance around the ring. In [13], analytical approximation of the average packet delay is presented for two CoS. A follow on to this research is in [14] which models the incoming traffic with three classes of services as opposed to only high and low priorities. This is obviously closer to the RPR CoS standard; nevertheless, it does not capture the subclass CoS support. Recent contributions to the delay analysis and its steady state bound for RPR single TB design are

4

discussed in [15]. A follow on research focused on simulating access delay for bursty traffic such as streaming video and interactive on-line gaming was published [16]. Lastly, ring access delay bound was proven analytically under steady state [17]. This approach demonstrates that the delay does not depend on the ring size; hence, it is a scalable solution for the MAN/WAN applications.

Limitations (d) and (e) have not been resolved yet. Resolving these unaddressed drawbacks is critical to the success of RPR as a key technology to meet the requirements for SP convergence in next generation multi-service broadband networks. Our research is focused on addressing these two items with our new and proposed techniques in way that is as compatible as possible with the current methods and operations of SP's.

1.3   Research Objectives and Thesis Contributions

1.3.1   Research Objectives

Bandwidth efficiency and service differentiation are fundamental elements to any SP. Addressing limitations (d) and (e), discussed above, is a must for SP's to adopt RPR and make it a key building block in delivering convergence while leveraging the current infrastructure and operations. Hence, we target our research at addressing drawbacks (d) and (e) with new techniques that can be adopted and deployed by SP's. The objectives and key steps of our research are summarized below:

(i) Illustrate using simulation techniques current ring utilization inefficiency

(ii) Introduce a new algorithm that increases dual-ring utilization and a per flow basis

(iii) Simulate the new algorithm and compare it to current RPR techniques

(iv) Analyze packet delay for all RPR classes and subclasses (vs. only 2 or 3 CoS)

(v) Derive packet delay for the new algorithm and compare results to item (iv)

(vi) Demonstrate current Transit Buffer designs cannot differentiate between RPR CoS

(vii) Propose a new Transit Buffer design which addresses item (vi)

5

(viii) Extend packet delay analysis in item (v) to support our new Transit Buffer design

(ix) Compare results of our new design to existing designs.

Finally, we also consider how our proposed techniques can be adopted by SP's and highlight what drawbacks they might bring. We also propose recommendation to ensure that such cases are either avoided or addressed.

1.3.2   Thesis Contributions

The first contribution of our research is a new RPR algorithm that addresses RPR limitation (d). Our algorithm is called Flow-based Priority Distribution (FPD), which enables higher utilization on both RPR ringlets while respecting CoS requirement per flow [18].

A second contribution is our delay performance analysis on all RPR classes and subclasses of service and control traffic (vs. only 2 or 3 RPR CoS). We derived the packet delay for RPR and our proposed algorithm FPD while considering all supported CoS [19].

Third, we proposed a new Transit Buffer (TB) design called 3-TB, which addresses limitation (e). We demonstrated the advantage of our design by extending our performance analysis to cover 3-TB design and comparing it to the existing designs.

Lastly, we also proposed a list of recommendations to network equipment vendors and service providers by combining our new techniques and using them within existing infrastructure and operations.

1.4   Thesis Summary

In this thesis, the present chapter provides an overview of SP's broadband networks architecture and capabilities needed to deliver convergence. We discuss our research motivation and highlight current RPR limitations. We list recent research that addressed some of these limitations. Then, we list our research objectives and discuss our contributions. Finally, we provide a summary of our thesis.

6

Chapter 2 describes how present networks should migrate to next-generation networks to deliver service and network convergence. Then, we provide a detailed review of IP and MPLS protocols, as well as the key service parameters and benefits to SP's. In addition, we discuss SP network topologies and how RPR can enable a multi-service transport solution. Finally, we review the RPR architecture and key protocols required to deliver this convergence.

The service and traffic models are detailed in Chapter 3. We discuss the various source models and their applicability per service. We also provide a detailed review of traffic modeling and ring performance research. We highlight how the unresolved RPR drawbacks impact SP's from seamlessly migrating to a converged infrastructure. This leads us to our new proposed algorithm that addresses such drawbacks.

In Chapter 4, we describe our proposed scheme called Flow-based Priority Distribution (FPD). We discuss the traffic and network model assumptions. Then, we demonstrate the performance improvement obtained using our method by presenting our simulation results. This performance increase is obtained at the expense of packet delay. We derive the packet delay and compare the analysis of RPR vs. FPD for various TB designs.

We discuss the end-to-end inconsistent CoS handling of the current RPR designs in Chapter 5. To address this drawback, we propose a new design, called 3-TB. We analyze packet delay for our new design. We demonstrate through our results the advantage of our proposed scheme in differentiating among the main classes of services proposed by RPR.

Finally, we summarize our research and outline the conclusion of our work in Chapter 6. We also propose several avenues to expand our work and provide additional enhancements.

CHAPTER 2 - Services, Protocols & Architecture of Broadband Networks

2.1 Evolution or Revolution of Broadband Networks?

As the demand for new IP-based services increases and the competitive pressure on Service Providers mounts, it is critical to avoid multiple overlaid service-specific networks. As a result, service providers focus on constructing a converged and efficient multi-service broadband network to deliver on such requirements. In addition, one of the key challenges in building such a network is the maintenance and operations required, which typically represents over 80% of the cost of the network. Consequently, it is imperative to create a network that satisfies existing and new services. More importantly, such a network should have the right building blocks to support a low cost yet scalable operational model. This conclusion has been publicly endorsed by major SP's such as AT&T/SBC, British Telecom, Bell Canada, Verizon, NTT and many others.

In order to deliver on the requirements for multi-service broadband networks, key technologies have been adopted across the industry and in the research community. As discussed in subsection 1.1.2, IP is the corner stone for the new services. Across the network, PWE and MPLS are the key protocols of choice to deliver network transport convergence. At the Edge of the network, both Ethernet and GFP are key and valid options for connectivity. While in the Access, various access technologies will continue to be applied and used. At Layer 2, a strong migration from ATM to Ethernet/MPLS has started, which is forming the basis of a consistent end-to-end Control Plane.

This adoption of new key technologies across the network results in significant changes to the architecture and implementation of the current broadband networks. These modifications include key capabilities such as multicast, QoS, security, authentication, service identification, availability and scalability. Hence, we will see an evolution of today's infrastructure but with revolutionary technologies to deliver the required capabilities and build the new multi-service broadband network.

8

2.2 The Internet Protocol and Next Generation Services

The Internet Protocol (IP) is the most popular protocol in the telecom industry. It is the base protocol of the Internet and the nucleus of new services such as VoIP and VoD. IP is a Network Layer Protocol corresponding to Layer 3 in the OSI protocol stack, as shown in Figure 2-1. It is a connectionless protocol, which defines a basic unit of data transfer called "packet" or "datagram" used through the inter-network, choosing a path over which the data is sent. It is defined in RFC791 to provide addressing and fragmentation [20].

OSI Model   **TCP/IP Stack**

	OSI Model		TCP/IP Stack	
7	Application Layer		App. Protocols (FTP, SMTP, TELNET, etc)	App. Protocols (SNMP, TFTP, XDR, etc)
6	Presentation Layer			
5	Session Layer		DNS	
4	Transport Layer		TCP	UDP
3	Network Layer		Internet Protocol (IP)	
2	Data Link Layer		Network Interface	
1	Physical Layer		Hardware	
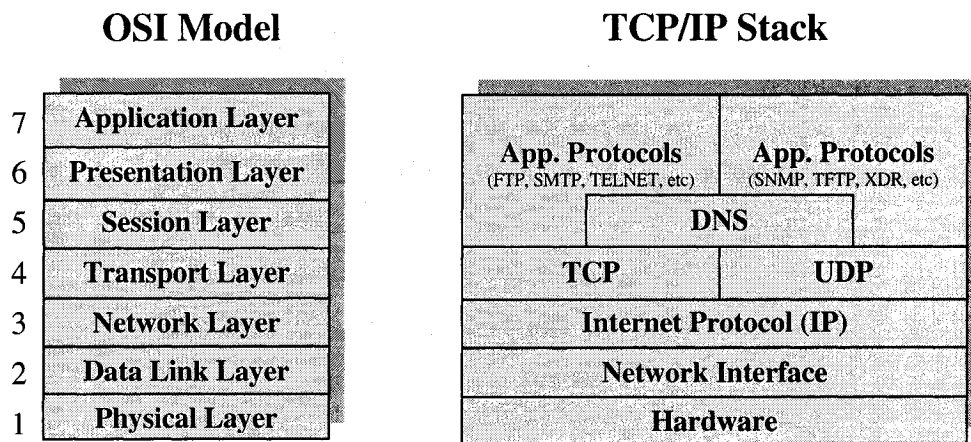
Figure 2-1: OSI Model and TCP/IP Protocol Suite

It has the responsibility of forwarding packets across all the interfaces that make up an inter-network. Therefore, the network layer specification covers both packet forwarding, as well as host addressing and protocols required to deliver diagnostic information about the inter-network. The datagram consists of a header and a data area. The first field in the header is the Version field which describes the version of the Internet Protocol used, hence the format for the rest of the header. There are two key versions of IP: v4 and v6. Figure 2-2 illustrates the format of these two versions of IP [20, 21].

2.2.1 IPv4 and IPv6

IPv4 has been around for over two decades. As new applications and services are introduced, and the number of hosts and the size of the inter-networks significantly grows

9

with the Internet, it became evident that a newer version of IP is required. IPv6 (RFC2460) [22] was designed as the successor of IPv4 (RFC791) [20]. The key enhancements are:

- Expanded addressing: IPv6 increases the address size from 32 to 128 bits in order to support more levels of addressing hierarchy, a much greater number of addressable hosts and simpler auto-configuration of addresses.

- Header format efficiency: some IPv4 header fields have been dropped or made optional, to reduce the packet processing cost and to enable the IPv6 header to be used in a bandwidth efficient way.

- Enhanced options support: changes in the way IP header options are encoded allows for more efficient forwarding, less stringent limits on the length, and greater flexibility for future options.

- Flow labeling: a new capability is added to enable the labeling of packets belonging to particular traffic "flows" for which the sender requests special handling, such as "real-time" service.

- Authentication and privacy: IPv6 introduced extensions to support authentication, data integrity, and (optional) data confidentiality.

IPv4 Header

4	8	16	32
Ver	IHL	ToS	Total Length
Identification		Flags	Fragment Offset
TTL	Protocol	Header Checksum	
Source Address			
Destination Address			
Options + Padding			
Data			

IPv6 Header

4	8	16	32 bits
Ver	Pri	Flow Label	
Payload Length		Next Header	Hop Limit
Source Address (128 bits)			
Destination Address (128 bits)			
Options + Other Headers			
Data			
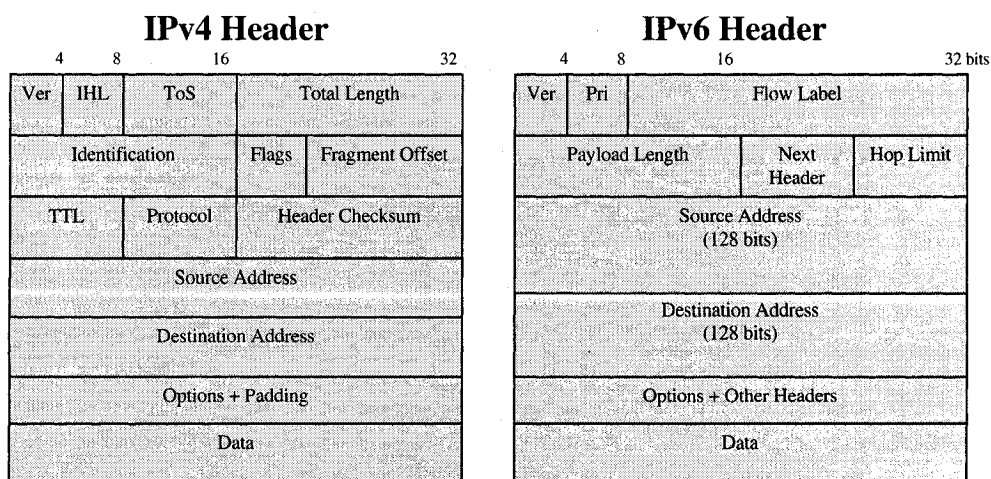
Figure 2-2: IPv4 and IPv6 Packet Headers

10

2.2.2 Routing Protocols

In a packet switched network (PSN), *routing* refers to the process of choosing a path to deliver packets, and *router* refers to any node making such a decision. It is important to understand that IP routing does not alter the original PDU; instead it consists of deciding where to send a datagram based on its IP address and header. A large IP inter-network, such as the Internet, is organized into *Autonomous Systems* (AS). Each AS consists of a collection of routers under the control of a single administrative entity. Typically, the collection of AS's are organized in a hierarchical fashion. In order to provide the IP routing functions, routers need to exchange routing information within and between AS's. Several key protocols developed for the IP layer will be discussed below. We focused here on IPv4, but equivalent submissions also exist for IPv6:

- Internet Control Message Protocol (ICMP) RFC792: is an integral and required part of IP that allows hosts to find routers attached to their segments and provides certain diagnostic capabilities to the hosts when the routers are not able to deliver packets to their destinations. Two of the most frequently used messages are ICMP *echo request* and *echo reply* [23].

- Interior Gateway Protocols (IGP): are used to exchange routing information within an AS. An IGP has to efficiently calculate routes and be able to quickly recalculate when changes happen in the network. There are two types of IGPs. The first type is based on the distance vector algorithm, such as *Routing Information Protocol* (RIP) RFC1058 and RFC2453, also known by *routed* from the 4BSD UNIX system, which is simple. Yet, it can be slow to converge under certain network failures. The second type is based on link-state algorithms, which are generally considered to have good convergence properties and employ a replicated distributed database approach, such as *Open Shortest Path First* Protocol RFC2328 [24, 25, 26].

- Exterior Gateway Protocols (EGP): are used to exchange routing information between AS's, express particular routing polices and aggregate routing information.

11

The first EGP was simply called Exterior Gateway Protocol and did a poor job of loop detection and forced a strict hierarchy of AS's. *Broader Gateway Protocol 4* (BGP-4) RFC1771 replaced it, as it allowed AS's to be organized without a strict hierarchy. The primary function of a BGP system is to exchange network reachability information with other BGP systems. It is used between SPs [27].

- Multicast Routing Protocols: services and connections which require transmitting data to many (not all) receivers require multicast capability. Video distribution is one prime driver for IP multicast support. An integral part of IP is Internet Group Management Protocol (IGMP) RFC2236 which handles multicast group membership requests and forwards multicast traffic on its attached subnets [28].

- Distance-Vector Multicast Routing Protocol (DVMRP): is defined in RFC1075 which constructs source-based multicast delivery trees using Reverse-Path Multicasting (RPM) algorithm. DVMRP was derived from RIP with a major difference of calculating the previous hop toward a source as opposed to next hop towards a destination [29].

- Multicast Extensions to Open Shortest Path First (MOSPF): a popular multicast protocol specified in RFC 1584, which is a link-state routing protocol with better and fast convergence properties than distance-vector algorithms. MOSPF interoperates with OSPF and can be used to forward both unicast and multicast. It also leverages the flexible-path metrics of OSPF to support QoS and cost-based routing. Some of its drawbacks are its lack of tunneling support, not being well suited for sparse topologies and being a computationally intensive routing protocol [30].

- Protocol-Independent Multicast (PIM): consists of two protocols – Dense Mode (PIM-DM) and Sparse-Mode (PIM-SM). PIM-DM is designed for an environment where group members are densely packed and bandwidth is available. PIM-DM is defined in RFC 3973 in the IETF [31]. While PIM-SM is optimized for

environments where group members are distributed across many regions of the inter-network and bandwidth might not be abundant. PIM-SM is defined in RFC4601 [32].

2.2.3 IP-based Services

New IP-based services are becoming increasingly more popular with consumers, businesses and Service Providers. Such services and their respective network topologies receive increased attention from the research community. They not only offer significantly better experience, new and enhanced features as well as flexibility; but equally important, they provide a convergence mechanism to deliver on the promise of operational and capital cost savings [33].

The top next generation services are VoIP, High-Speed Internet Access with security and Digital TV (IPTV with both Standard and High Definition TV). Figure 2-3 illustrates a next generation multi-service broadband network. The diagram demonstrates the need for various connectivity protocols at Layer 1 and 2 across the network. At the consumer or business premises, IP/Ethernet is a given. In this example, we use VDSL2 as access connectivity to the first Central Office (CO). Then, we move into the Edge of the network with GbE over Generic Framing Protocol (GFP), where such IP based services are handled, delivered or transported to the Core.

In order to deliver these services appropriately, the IP protocols that we cover in this section are fundamental. Consider the Digital Video Broadcast and Video on Demand services (IPTV), IP and IGMP are absolutely critical to deliver digital TV broadcast and Video on Demand (VoD) across such a network. However, these services must be delivered reliably across the entire network. In addition, as the number of subscribers and coverage increases, it becomes increasingly more important to have a scalable network that can guarantee a minimum availability and be easily managed with a common Control Plane. These requirements are the premises of what Service Providers are counting on for IP/MPLS and RPR to enable across the network.

13
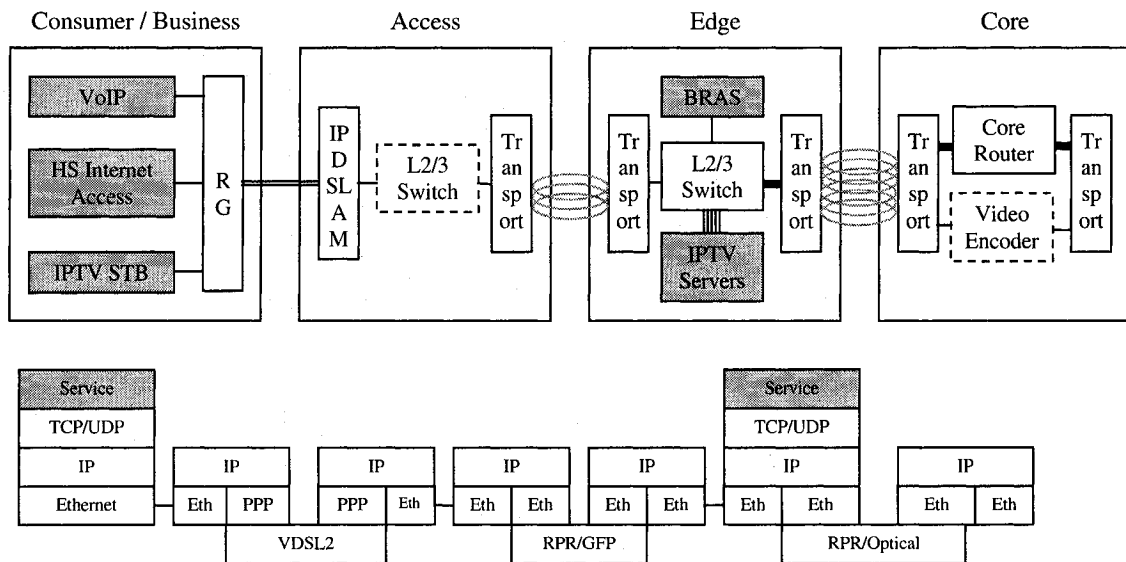
Figure 2-3: Example of a NG Multi-Service Broadband Network and Services

2.3 MPLS Protocol and Service Requirements

Key challenges in traditional IP networks are the bottleneck in the IP header processing and next-hop determination. A key mechanism to address this issue is cut-through switching which maps Layer 3 paths into Layer 2 circuits, allowing next-hop lookup in hardware. One of the proposals to deliver on this promise is Multi-Protocol Label Switching (MPLS) which is being touted as the technology that will enhance speed, scalability and service-provisioning in any network layer protocol [1].

MPLS is specified by RFC 3031 to address such challenges by assigning a particular packet to a certain Forwarding Equivalency Class (FEC) only once, as the packet enters the network. This is accomplished by encoding the FEC into a fixed-length value called the "label" which is used to eliminate any further header analysis by subsequent routers. All forwarding is driven by the labels, resulting in resolving the challenge faced in conventional network layer forwarding [34].

Another key benefit to MPLS is its promise to provide a common infrastructure for Service Providers to build cost-effective networks enabling a multi-service environment for current and next generation services: the basis for convergence!

2.3.1 Forward Equivalency Class (FEC) and MPLS Label

One of the key MPLS capabilities is categorizing a packet to belong to a particular flow only once, known as Forward Equivalency Class (FEC), and subsequent packets are switched based on this relation. This leads to a significant reduction in lookup time, hence reducing the latency of the system. The FEC assigned to a packet is encoded in a 20–bit unsigned integer known as the label which is a part of a 32-bit fixed length identifier inserted between the L2 and L3 header. Packets are labeled before they are forwarded. At subsequent nodes, the label is used as an index to specify the next hop and a new label. The old label is replaced with the new one, and the packet is forwarded. MPLS also enables label stacking allowing for hierarchical operation in the MPLS domain, as shown in Figure 2-4 [34].

| L2 Header | Label 1 | Label 2 | _ _ _ | Label N | L3 Header | L3 Data |

Bottom of Stack: 1 bit

| Label | CoS | S | TTL |

MPLS Label: 20 bits   Class of Service: 3 bits   Time-To-Live: 8 bits
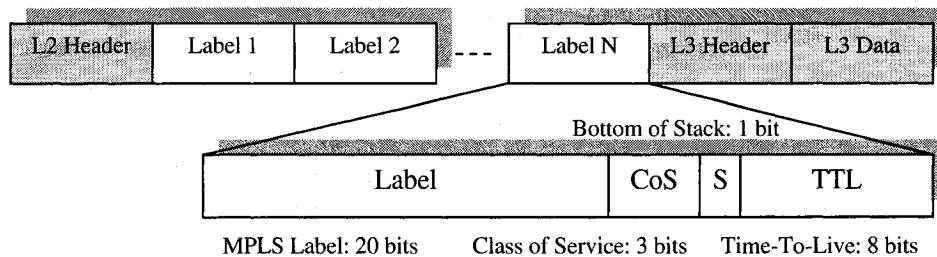
Figure 2-4: MPLS Label Formats and Stacking

2.3.2 MPLS Architecture

MPLS nodes are capable of Layer 2 switching and Layer 3 routing, in addition to switching and forwarding labeled packets. A node has two architectural planes: forwarding and control planes. The **forwarding plane** uses an indexing operation into an *Incoming Label Map* (ILM) to forward labeled packets. ILM is maintained by the node and maps each incoming label to a set of *Next Hop Label Forwarding Entries* (NHLFEs) to forward a labeled packet. The NHLFE consists of a sequence of entries each containing the packet's next hop, the operation to perform on the label stack (replace, pop, and replace/push), data

15

link encapsulation and label stack encoding when transmitting a packet. Moreover, each FEC is mapped to the set of NHLFEs using the *FEC-to-NHLFE* (FTN) map which is used when forwarding packets that arrive unlabeled and to be labeled before forwarding [34].

The **control plane** is responsible for populating and maintaining the maps. All MPLS nodes run IP routing protocols such as OSPF to exchange IP routing information with all other MPLS nodes. Label-binding information can be distributed using Label Distribution Protocol (LDP) or extensions to BGP, since it is only distributed amongst adjacent routers. This makes Link-State protocols not suitable for label distribution as they flood routing information among a set of routers that are not necessarily adjacent. The labels exchanged amongst the MPLS nodes are used to build the node maps. MPLS uses a forwarding paradigm based on label swapping, which can be combined with a range of control modules that are responsible for assigning and distributing a set of labels. Control modules include: unicast, multicast, traffic engineering, and QoS modules which will be discussed in the next section. Figure 2-5 shows MPLS node architecture and its planes [34].

2.3.3   MPLS Components

MPLS identifies two basic types of nodes: a *Label Edge Router* (LER) and a *Label Switching Router* (LSR), as shown in Figure 2-6. Both types implement an IP protocol such as OSPF, and participate in a label distribution protocol such as MPLS-LDP, MPLS-BGP etc. An LER is located at the edge of the MPLS domain; it acts as an MPLS ingress or egress node. In the ingress direction, the I-LER examines the packet header and determines the destination and FEC based on FTN. Then it inserts the appropriate label and forwards the encapsulated packet to the outbound interface. In the egress direction, the E-LER simply extracts the label prior to forwarding to a non-MPLS interface. LERs are also referred to as Edge LSRs or MPLS Edge nodes.
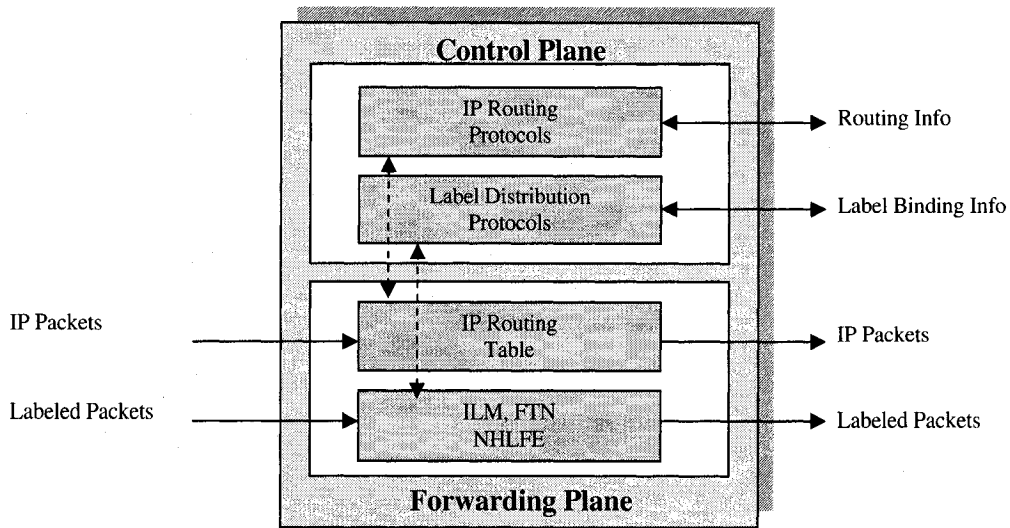
Figure 2-5: MPLS Node Architecture

The LSR, located in the MPLS core, performs label switching on all its interfaces, and it is capable of forwarding native L3 packets. All forwarding is based on the labels of incoming packets. An LSR uses the ILM to map each incoming label to a set of NHLFEs. It decides to swap the incoming label to the outgoing one or add another label on top of the existing one, then forward the labeled packet to its next hop outgoing interface. Both LER and LSR are logical components that are likely to be implemented on the same device.

MPLS nodes use a *label distribution protocol* to establish label paths across the network in conjunction with network layer routing protocol. A label distribution protocol is a set of procedures by which an MPLS node (LSR) informs another of the label/FEC binding it has made. Such two LSRs are known as "label distribution peers" with respect to the exchanged binding information, and have a "label distribution adjacency" between them. They distribute their labels over TCP as it provides a reliable transport with robust flow-control and congestion handling. In addition, label distribution protocols encompass any negotiations in which two label distribution peers need to engage to learn of each other's MPLS capabilities. Such protocols can be extensions of the existing IP protocols like MPLS-BGP, or MPLS related such as MPLS-LDP. As an example, LDP defines four

categories of messages: Discovery, Session, Advertisement, and Notification messages for distributing the labels and binding information.

FEC	In Label	In I/F	Out Label	Out I/F
10.1	200	1	-	-
10.2	-	-	200	1

LER 1 FTN

FEC	In Label	In I/F	Out Label	Out I/F
10.1	-	-	400	1
10.2	400	1	-	-

LER 2 FTN

FEC	In Label	In I/F	Out Label	Out I/F
10.1	300	2	200	1
10.2	200	1	300	2

LSR 1 ILM

FEC	In Label	In I/F	Out Label	Out I/F
10.1	400	2	300	1
10.2	300	1	400	2

LSR 2 ILM
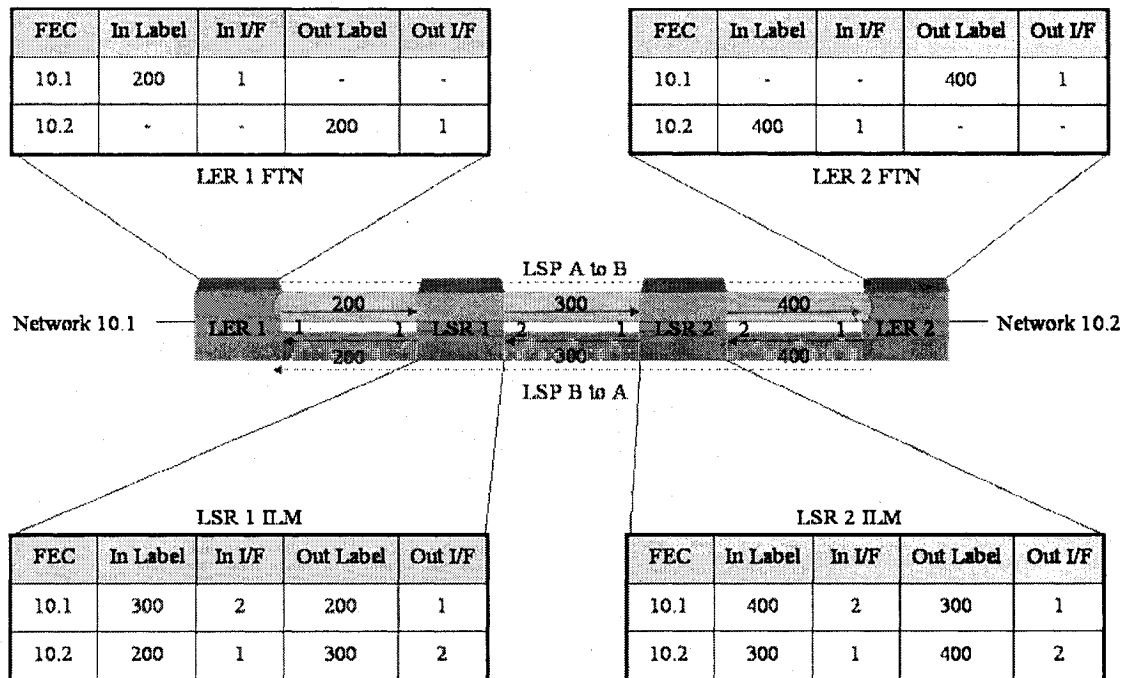
Figure 2-6: An Example Demonstrating MPLS Components

LSRs are connected using a *Label Switched Path* (LSP) which is used to forward packets between them. An LSP is a specific unidirectional traffic path through the network over a set of LSRs that packets belonging to a certain FEC travel to get to their destination. LSPs are provisioned using LDP, RSVP-TE, CR-LDP or multi-protocol BGP. Typically, RSVP-TE which runs over UDP is used to setup LSPs across the network. CR-LDP which runs over TCP can be used as well. The advantage of RSVP-TE is that it is based on IP. LSPs can be established using two ways: independent or ordered control. The independent method provides faster convergence and establishment. The ordered control method distributes the bindings before LSPs are established; its advantage is a better loop dedication capability. Figure 2-6 provides an example of the MPLS components [1] and [34]. The figure illustrates the functionality of LERs and LSRs in IP and MPLS networks as well as their associated FTN and ILM tables. When a packet is generated on the 10.1 network and destined to the 10.2 network, LER1 examines the packet header and

determines its respective FTN table entry based on its FEC. An MPLS label 200 is pushed on the packet header, and the new packet is forwarded on "Out I/F 1". Then, LSR1 inspects the incoming packet header on "In I/F 1". Based on the ILM table entry, LSR1 pops MPLS label 200 and pushes label 300, then forwards it to "Out I/F 2". A similar function is carried out in LSR2, where MPLS label 300 is popped and label 400 is pushed. The packet is forwarded on "Out I/F 2" to LER2. Finally, LER2 receives the packet and inspects its header and FEC. It pops label 400 and forwards the original packet to its network destination 10.2. A similar operation takes places when a packet is generated on 10.2 and is destined to 10.1 network.

2.3.4 IP/MPLS Requirements for Broadband Services

Many service providers are looking to consolidate the delivery of multiple services and traffic types onto a single IP-optimized network. MPLS-based infrastructure enables this convergence with innovative and revenue-generating service enhancements. Such infrastructure requires key capabilities, namely: in the control plane, QoS, scalability and resiliency.

2.3.4.1 Pseudo-Wire Emulation & Control Plane

One of the enablers to deliver convergence is *Pseudo Wire Emulation Edge-to-Edge* (PWE3) RFC3916 which is a mechanism that emulates key attributes of a service such as FR or Ethernet over a *Packet Switched Network* (PSN). A *pseudo-wire* (PW) encapsulates service-specific PDUs at the ingress port and carries them across a PSN tunnel while maintaining their timing and ordering. From a customer perspective, the PW seems as an unshared link for a given service. As shown in Figure 2-7, a PW is a connection between two *provider edge* (PE) nodes, which connects two *attachment circuits* (AC) from *customer edge* (CE). For instance, an AC is an Ethernet port, a VLAN, a FR DLCI, an MPLS LSR, etc.
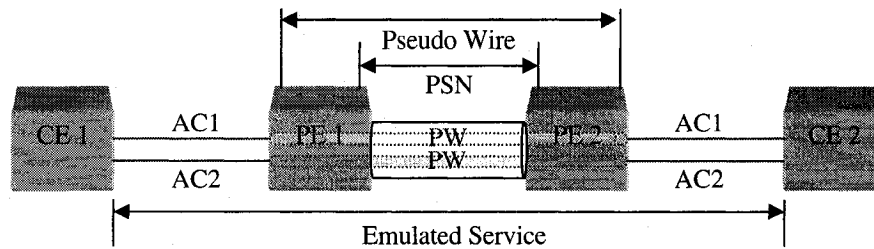
Figure 2-7: PWE3 Reference Model

A PW must be setup prior to establishing an emulated circuit. Similarly, when an emulated circuit is not required, a PW must be torn down. In case of a bi-directional circuit, it can only be established after its associated PW and PSN tunnels are functional in both directions. Setup and teardown of a PW can be established in three ways: a command from the management plane of a PE, Setup/Teardown of an AC such as an ATM SVC, or an auto-discovery mechanism.

Several mechanisms have been proposed in the research and industry. One of the approaches is to use OSPF-TE as an IGP with tunneling extensions for auto-discovery, RSVP-TE for LSP tunnels setup and signaling between PEs, and LDP for PW distribution. Figure 2-8 illustrates an emulated service tagged with a PW MPLS inner label at the De-multiplexer layer, then tunneled into a PSN MPLS network with an LSP tunnel as the outer label. This approach results in significant operational savings due to its auto-discovery capability, and its ability to support stringent QoS for a scalable and resilient network and service, as discussed in the following sections [5, 35].
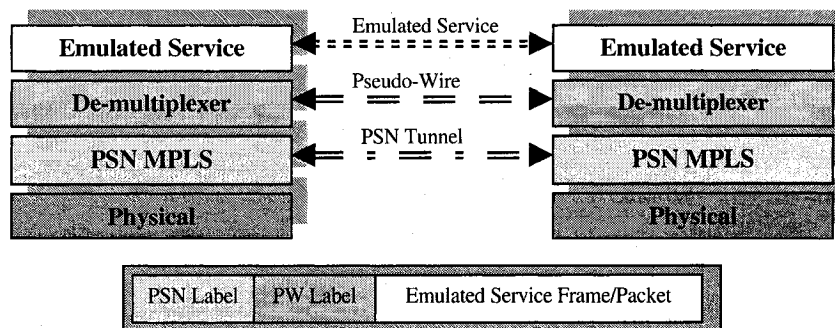
Figure 2-8: PWE3 Protocol Stack and Encapsulation Reference Model

2.3.4.2 Traffic Engineering (TE) & Quality of Service (QoS)

MPLS provides a connection-oriented environment that enables *Traffic Engineering* (TE) of a PSN. TE can guarantee bandwidth for various flows which is fundamental for enabling QoS. A key QoS requirement is to enable latency and jitter-control for time-sensitive applications such as VoD or VoIP. This can be established by combining MPLS-TE with technologies that provide traffic flows with their class-specific treatment such as RSVP-TE signaling and DiffServ-based forwarding.

RSVP-TE RFC3209 provides a set of tunneling extensions to the original RSVP protocol which was developed as a signaling protocol for reservations and explicit admission control for IntServ. RSVP-TE was formed for a variety of network applications one of which is Traffic Engineering. A key advantage is that it operates on RSVP capable routers to enable creation of explicitly routed LSPs as well as provide smooth rerouting, preemption and loop detection. RSVP-TE provides signaling between pairs of LERs to create a traffic trunk (TT) which aggregates same class flows on a single LSP tunnel. This results in reducing the number of RSVP states in the network. RSVP-TE uses extended IGPs and constraint-based routing, e.g. OSPF-TE, to provide additional information for each TT handled at the I-LER, such as: max link bandwidth, reservable bandwidth, current bandwidth reservation at each priority level, etc.

Consequently, MPLS-TE provides the mechanism to guarantee bandwidth and differentiate services, while RSVP-TE provides the means for controlling delay and its variations for time-sensitive flows. However, this solution lacks a packet discard mechanism, which can be addressed by combining MPLS with DiffServ. DiffServ redefines the 8-bit ToS field in the IP header to a 6-bit DiffServ Code Point (DSCP) value and a 2-bit Explicit Congestion Notification (ECN). It provides a QoS to traffic aggregate and does not require per-flow signaling. Meanwhile, MPLS-TE can guarantee bandwidth for FECs. This approach is as described in RFC3270. Combining MPLS with RSVP-TE/DiffServ allows Service Providers to deliver bandwidth assurance, load-balancing, diverse routing and path redundancy which are critical to deliver QoS across the network [5, 36, 37, 38].

2.3.4.3 Resiliency & Scalability

Since MPLS is path-oriented, it can provide faster and more predictable protection and restoration capabilities in the face of topology changes than conventional hop by hop routed IP systems. Resiliency in MPLS can be achieved through end-to-end or local recovery. **End-to-end recovery** can be established through *load-balancing* on multiple LSPs, or *path-protection* with two modes. In, 1+1 protection, a working and a dedicated backup are used in parallel; the receiving node chooses which path to take traffic from. In 1:1 protection, a working and a backup path are pre-provisioned, but only the working path is established, and in case of failure, traffic will be switched to the backup path, which takes longer to switch due to the notification and trigger for recovery.

Another approach is through **fast local recovery** which can redirect traffic rapidly to a backup path, this is also known as Fast Re-Route (FRR). It provides recovery for node and link failures, where the backup (protection) nodes or links are pre-signaled and quickly (in tens of ms) switched to in failure scenarios. There are two types of backup LSPs: one-to-one and one-to-many. In the case of *one-to-one backup*, a detour LSP is provided at each point of local repair (PLR). Also, detour LSPs can merge to improve scalability. For the *one-to-many* or *facility backup* case, a single LSP is created for multiple links and nodes failures instead of creating many detours, and label stacking is used to preserve labels from protected paths [39].

Another important mechanism to provide resiliency in MPLS is *graceful restart*. It provides a self-healing capability to the MPLS control plane (LDP, RSVP-TE, and BGP) without disturbing the MPLS forwarding plane (minimum disruption). Graceful restart is used in case of planned restart such as software upgrade, and unplanned restart such as software crash/restarts [36, 37, 38, 40].

2.3.5 Generalized MPLS (GMPLS)

As discussed above, MPLS-TE FRR is a flexible and reliable protection solution especially for data services. However, the establishment of TE tunnels requires MPLS operations and management with an IGP domain. Moreover, FRR capability is supported by IP/MPLS

routing platforms. As a result, protection LSPs are not capable of detecting intermediate non-routing node status and failures. This resulted in extending the FRR concept into Generalized MPLS (GMPLS) from an IP environment to the Optical and transport space.

GMPLS enables the control and management from an IP routing domain to the TDM/WDM transport domain with a common environment. GMPLS utilizes extended TE signaling protocols such as Generalized CR-LDP or Generalized RSVP to deliver non-blocking protection tunnels across the packet, TDM and wavelength domains. GMPLS provides end-to-end unified protection and recovery schemes for data services and optical transport infrastructure [39].

2.4 Network Topologies and Transport Solutions

2.4.1 Service Provider Network Topologies

Various topologies are supported in Service Provider networks; Table 2-1 provides a summary of these topologies. The simplest form and lowest cost is a linear topology where interconnecting multiple points is achieved through a serial connectivity. The drawback is the distance between non-neighboring points and node isolation (lack of resiliency) in case of a single link failure.

To address the distance issue, the nodes can be interconnected through a central point; this is known as star topology. Even though the distance between points is now minimized through the central point; however, it is not optimal. To optimize connectivity to every node, mesh topology is used. It also provides alternate paths in case of a link failure; hence, providing higher resiliency. The main drawback of mesh topology is the number of links required to connect all nodes; thereby, it is costly and complex.

Another approach to provide resiliency is ring topologies. The advantage of a ring topology is that it provides full connectivity to every node with minimum number of links in a loop. Therefore, it is regarded as a cost effective and resilient method. Its drawback is that the distance between any two nodes (which are not neighbors) is not optimal. The
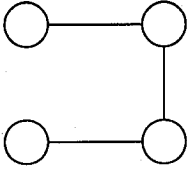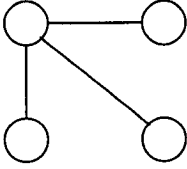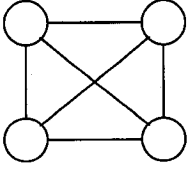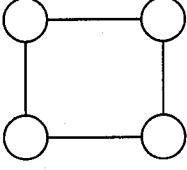
dominant topology in Service Provider networks is a ring due to its low cost deployment, operational simplicity and resiliency.

2.4.2 Network Transport Solutions

Service Provider networks are predominantly based on a synchronous transmission network deployment. In North America, ANSI published Synchronous Optical NETwork (SONET), while in Europe and the rest of the world ITU-T published Recommendations G.707, G.708 and G.709 covering Synchronous Digital Hierarchy (SDH). SONET/SDH support the topologies discussed above; however, due to the advantages of ring topologies, fiber optics lines have been laid out in ring configurations across the network primarily in the Metro Edge and Core portions of the network. As a result, the dominant SONET/SDH configuration is dual-ring in which one of the rings is used as the backup ring, and remains unused during normal operation, utilized only in the case of failure of the primary ring. The static bandwidth allocation and network monitoring requirements increase the total cost of a SONET/SDH network.

Another key technology is Ethernet IEEE 802.3, specifically Gigabit Ethernet, which emerged from the LAN, and is now extending its reach across the campus and into the Access and the Edge of certain portions of the network. The key advantage of Ethernet is its simplicity, low cost and total domination of the Enterprise. However, Ethernet has many drawbacks before being able to be deployed across the Service Provider networks. These drawbacks are reach, fairness, strong QoS, and resiliency. The Metro Ethernet Forum (MEF) has been established to enhance the Ethernet standards in order to address the above drawbacks and make it more suitable for Service Provider networks and the support of legacy circuit-based services.

Table 2-1: Service Provider Network Topologies

Topology Name	Example	Advantage	Disadvantage
Linear		Simplest topology & lowest operational cost	Lack of resiliency & longer distance between points
Star		Low cost operational and shorter distance	Lack of resiliency and node isolation
Mesh		Resilient to link failures and single link distance to any point	Expensive operationally and complex
Ring		Resilient to link failures and low cost	Longer distance between non-neighboring points

A third and a recent transport technology aimed at Service Provider networks is Resilient Packet Ring (RPR) IEEE802.17 which was declared a standard in 2004. Its main benefits are: network transport convergence, high utilization, resiliency and fairness. These goals are not achievable with existing technologies such as SONET/SDH, Ethernet or legacy ring protocols such as Token Ring. As a result, RPR has received strong endorsement from many Service Providers to migrate the current circuit-based infrastructure to packet-based next-generation broadband networks. In the next section, we provide an overview of RPR [41].

2.5 Resilient Packet Rings in Next Generation Broadband Networks

2.5.1 IEEE 802.17 RPR Objectives & Services

RPR is an IEEE standard defined in 802.17. It is a new high-speed Media Access Control (MAC) technology for Campus, Metro Edge and Core networks. RPR's objectives are: network transport convergence, high utilization, fairness and fault-tolerance. The technical challenge with RPR is to dynamically deliver on these goals with changing multi-service flows and network demands especially in utilization and fairness [42, 43].

RPR can distinguish and support various services with its three defined classes and subclasses, as shown in Figure 2-9. Class A provides a guaranteed and allocated bandwidth. This class supports the lowest delay and jitter which makes this class suited for real-time applications. There are two subclasses, class A0 which reserves the bandwidth required on the ringlet, and class A1 which allocates the bandwidth but allows a ringlet to reclaim it. Subclass A0 has the highest priority followed by subclass A1.

Class of service			Quality of service			
Name	Example use	Subclass	Guaranteed bandwidth	Delay/jitter	Bandwidth type	Bandwidth subtype
classA	real time[a]	subclassA0	yes	low	allocated	reserved
		subclassA1				reclaimable
classB	near real time[b]	classB-CIR	yes	bounded	allocated	
		classB-EIR	no	unbounded	opportunistic	
classC	best effort[c]	—				

[a]Real-time services need as close to immediate servicing as possible. Examples include interactive telephony.
[b]Near real-time services need short, but not immediate, time frames for servicing. Examples include noninteractive video.
[c]Best effort services are not time dependent, and they do not need a minimum amount of bandwidth within any particular time frame. Examples include mail.

Figure 2-9: RPR Class of Services and Quality of Service Supported Modes [42]

Class B has two subclasses: B-CIR (Committed Information Rate) and B-EIR (Excess Information Rate). Subclass B-CIR provides guaranteed and allocated bandwidth with bounded delay. However, its allocated bandwidth is reclaimable. Subclass B-EIR does not guarantee the bandwidth; nevertheless, it allocates its required bandwidth and allows the

26

ringlet to reclaim it. Also, subclass B-CIR provides a bound on the delay and jitter. Subclass B-CIR has lower priority than subclass A1. Similarly, subclass B-EIR is a lower priority CoS than B-CIR.

Class C is the lowest priority CoS. It does not guarantee bandwidth, jitter or delay bounds. Bandwidth is allocated opportunistically. This class is suited for best effort traffic.

2.5.2 RPR Architecture & Operation

An RPR connects N nodes using a bi-directional ring configuration: ringlet 0 clockwise and ringlet 1 counter-clockwise, where both rings have equal capacity. Figure 2-10 shows an RPR ring with a generic nodal architecture. Each node connects to both ringlets and also has station ingress and station egress traffic. Station ingress traffic is client traffic destined to be transported on ring, while station egress traffic is traffic arriving from the ring and destined to be dropped locally to the client.

Each node also has ringlet ingress and ringlet egress traffic. Ringlet ingress traffic is checked to determine if it is destined locally to the station egress traffic or pass-through where it is queued in the transit buffer (TB). Ringlet egress traffic is either pass-through traffic or station ingress traffic that is destined to be transmitted on the ring. The frame format of RPR ringlet ingress and ringlet egress traffic is shown in Figure 2-11.
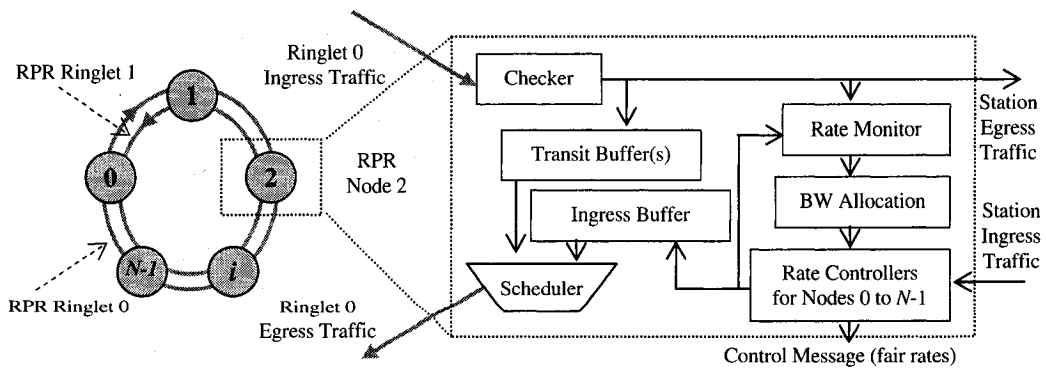
Figure 2-10: RPR Network and Generic Nodal Architecture

The standard defines two TB implementations: single queue (or 1-TB) called Primary Transit Queue (PTQ) serviced with strict priority over ingress buffer traffic. The second is with dual queues (or 2-TB), PTQ for Class A traffic, and Secondary Transit Queue (STQ) for Class B and Class C traffic. Traffic in STQ is serviced with strict priority over ingress buffer traffic when a certain depth (or threshold) is reached in the STQ. Otherwise, STQ traffic is serviced using a round-robin discipline with ingress buffer traffic. The advantage of the 1-TB is hardware simplicity; however, lower CoS traffic blocks higher priority traffic from accessing the ring. The 2-TB design offers lower access delay for higher priority traffic; nonetheless, the STQ buffer must be large enough to store the lower priority traffic while yielding access to the ring [44, 19].

The RPR standard defines a fairness algorithm to avoid starvation. The fairness algorithm uses explicit rate feedback to control the amount of traffic that each node inserts on the ring. There are two modes in RPR: Aggressive Mode (AM) and Conservative Mode (CM). The AM is the default mode and is associated with the 2-TB design, while the CM uses the 1-TB implementation. In both modes, each node measures the output of the scheduler and ingress traffic over a fixed interval. These measurements are used to detect congestion and compute fair rates to be sent in control messages to upstream nodes [42, 44].

In addition, IEEE 802.17 is defining an optional sublayer of the MAC that supports efficient bandwidth utilization for applications on RPR that require bridging clients. This is specified in IEEE 802.17b Spatially Aware Sublayer (SAS) [45].

Another key area being developed in RPR is inter-ring traffic engineering and its impact on packet buffers on such interconnects. A proposal on global fairness for inter-ring traffic is discussed and simulated in [46].

RPR Frame	Number of bytes	Description
Time-To-Live	1	(TTL) Hop count to destination
Base Control	1	Frame type, service class and control
Destination Address	6	RPR destination node address
Source Address	6	RPR source node address
TTL Base	1	For computing hop count to source
Extended Control	1	Extended flooding/consistency checks
HEC	2	Header Error Check (HEC) CRC
Protocol Type (PT)	2	Form and function of SDU
Service Data Unit (SDU)	Variable	Data provided by client
FCS	4	Frame Check Sum CRC for PT/SDU

Figure 2-11: RPR Frame Format

2.5.3 IP and MPLS Traffic Mappings over RPR

As discussed, service packets are based on IP. IP packet transport in the Access and Edge portions of the network can get transported over RPR. However, in the Edge and Core parts of the network, IP packets will be switched using MPLS protocol over RPR transport layer. Hence, mapping of IP and MPLS packets into IEEE 802.17 networks is important. Currently, a proposed IETF Internet-Draft by M. Holenss and G. Parsons was proposed in June 2006 [47].

We consider IP traffic classification using a field called "Differentiated Services" (DS) which is used in IPv4 and IPv6. The DS 8-bit field allocates 6 bits for DS Code Points (DSCP) and 2 bits for Explicit Congestion Notification (ECN) field. The DSCP field is used to select the per hop behavior (PHB) for each IP packet at a network node. The

mappings of the DSCP to RPR are illustrated in Table 2-2. The coding of ECN does not impact the mappings to the RPR service class relevant fields [47].

Table 2-2: IP DSCP Mapping to RPR CoS's

IP Service Class	DSCP	Per Hop Behavior (PHB)	RPR Service Class
Standard	000000	Default Forwarding (DF)	Class C
Low Priority Data	001000	Class Selector 1 (CS1)	
High Throughput Data	001010	AF11	Class B-CIR
	001100	AF12	Class B-EIR
	001110	AF13	
OAM	010000	Class Selector 2 (CS2)	Class B-CIR
Low Latency Data	010010	AF21	Class B-EIR
	010100	AF22	
	010110	AF23	
Broadcast Video	011000	Class Selector 3 (CS3)	Class B-CIR
Multimedia Streaming	011010	AF31	
	011100	AF32	Class B-EIR
	011110	AF33	
Real-time Interactive	100000	Class Selector 4 (CS4)	Class A1
Multimedia Conferencing	100010	AF41	
	100100	AF42	
	100110	AF43	
Signaling	101000	Class Selector 5 (CS5)	Class A0
Telephony	101110	Expedited Forwarding (EF)	
Network Control	110000	Class Selector 6 (CS6)	
Reserved for Future Use	111000	Class Selector 7 (CS7)	-

The best effort group is in Default Forwarding (DF) and Class Selector 1 (CS1) PHB's which are mapped to Class C. The Assured Forwarding (AF) groups are a means for a DS domain to offer different levels of forwarding assurances for IP packets. In case of congestion, the drop precedence of a packet determines the relative importance of the packet within the AF class. A congested DS node tries to protect packets with lower drop

30

precedence value from being lost by preferably discarding packets with a higher drop precedence value. The DS CS and AF group are mapped according to their priorities to RPR CoS's.

The Expedited Forwarding (EF) PHB group is used to build a low loss, latency, jitter and assured end-to-end bandwidth and service. Hence, it is mapped to RPR CoS A0. Also, note that CS7 is reserved for future use and it is not mapped to any RPR class. Now, let us consider the mapping of MPLS traffic over RPR networks. As shown in Figure 2-4, an MPLS label has 3 bits defined for CoS support, which defined the PHB. The mappings between MPLS CoS bits to RPR service classes are shown in Table 2-3.

Table 2-3: MPLS Mapping to RPR CoS's

MPLS CoS	RPR CoS
000	Class C
001	Class C
010	Class B-CIR
011	Class B-EIR
100	Class A0
101 (reserved)	Class A0
110	Class A1
111 (reserved)	Class A1

All MPLS supported CoS's are mapped into relevant RPR CoS's. Note that the two reserved MPLS CoS's are mapped into its appropriate Class A and its subclasses to provide consistency with its preceding MPLS CoS mapping.

2.6 Summary

In this chapter, we described how multi-service broadband networks have to evolve to achieve convergence. In addition, we discussed the main components of IP protocol

including enhancement of IPv6. We provided a summary of the key IP routing protocols. We also gave an example of next generation IP-based services.

We discussed the MPLS protocol, architecture and IP/MPLS requirements for broadband services convergence, namely: Pseudo-Wire Emulation, Traffic Engineering, Quality of Service, resiliency and scalability. We also provided a view on GMPLS and how it fits with MPLS in a service provider network.

Furthermore, we reviewed SP network topologies and transport solutions; we highlighted the value of ring topologies and namely RPR as a critical transport solution for service providers. We introduced RPR and its role in multi-service broadband convergence and described its architecture. We also discussed the key classes of service supported by RPR and how they map IP and MPLS classes of service.

CHAPTER 3 – Traffic Characteristics, Models & Ring Performance Analysis

3.1   Packet Traffic and IP Services Characteristics

To meet the requirements for broadband services discussed in the previous chapter, it is fundamental to understand their statistical properties and characteristics. These properties are considerably different from the statistical properties of conventional voice traffic or traditional data traffic. We focus our work on four key IP-based classes of services, namely: Voice, Video, Low-Speed Data Access, and High-Speed Data Access.

3.1.1   Voice Services: Voice over IP (VoIP)

VoIP refers to delivering voice information in discrete packets using IP instead of the traditional circuit-based public switched telephone network (PSTN). An advantage of VoIP is that it avoids the tolls charged by ordinary telephone service. VoIP also uses Real-Time Protocol (RTP) RFC 1889 to help ensure that packets get delivered in a timely fashion. Session Initiation Protocol (SIP) RFC3216 is also a critical standard for the establishment of multimedia sessions used for audio, video, instant messaging, or other real-time data communication. Another key protocol for VoIP is H.323 which is an ITU standard for sending voice and video using IP. It addresses call control and management, gateway administration and connectivity with the PSTN [48, 49, 50].

Voice sources generate relatively low bandwidth real-time traffic i.e. very sensitive to delay, but it tolerates moderate packet loss. A voice source generates packets during talk periods at a constant rate and no packets during silent periods. Traditional voice traffic in general is modeled using a well-known Poisson process. In addition, it can be modeled as alternating talk-silent periods. During talk periods, traffic can be generated at 64kbps using Pulse Coded Modulation (G.711 PCM), or at lower compressed rates of 40/24kbps using Adaptive Differential PCM (G.726 ADPCM) [51, 52].

3.1.2   Video Services: Digital Video and Video on Demand (VoD)

Digital Video Broadcast (DVB) is a set of standards that define digital broadcasting using existing satellite, cable, and terrestrial infrastructures - it is a European initiative, with Digital TV (DTV) being its American equivalent. DVB uses MPEG-2 ISO/IEC standard as a lossy-compression of audio and video signals. MPEG-2 is also adopted by SDTV, HDTV and DVD technologies. As new video services demand more flexibility, scalability and interactivity, MPEG-4 standard was developed to satisfy such demands with very efficient near-lossless compression of content over a wide range of bit rates, from 64Kbps to potentially 38Mbps. For authors, it enables greater reusability and flexibility in content creation. For service providers, MPEG-4 provides more transparency, error resiliency and QoS awareness. For end users, it delivers tools for seamlessly integrating broadcast content with equally high-quality interactive MPEG-4 objects.

Video on Demand (VoD) systems allow users to select and watch digital video content over a network as part of an interactive television system. VoD systems are either *streaming*, in which viewing can start as the video streams over the network, or *download*, in which the program is brought in its entirety to a set-top box before viewing starts. VoD service allows the user to have control over the session presentation with capabilities to forward play, reverse play, freeze, and random positioning. VoD's biggest obstacle is the lack of a network infrastructure that can handle the large amounts of data required by video. As a result, major Service Providers have announced to build and deploy a Broadband network to deliver DTV and VoD services, such as at&t, Telecom Italia, Bell Canada, Verizon, Telefonica, etc.

Digital video sources require the largest bandwidth, are sensitive to delay, and generate correlated packets. The statistical characteristics of such traffic is dependent on several aspects such as correlation of a line in an image with the next line, frame correlation, scene correlation, amount of interactivity with the end-user, and the QoS offered [52, 53].

3.1.3 Low-Speed Data Services: WAP and MMS

Wireless Application Protocol (WAP) is an industry-wide specification for developing applications that operate over wireless communication networks. The wireless market is growing very quickly, and reaching new customers and services. To enable SP's to meet the challenges in advanced services, differentiation and fast/flexible service creation, WAP Forum defines a set of protocols in transport, security, transaction, session and application layers.

Multimedia Messaging Service (MMS) is a service environment and a system application by which a WAP client is able to provide a messaging operation with a variety of media types such as text, images, audio and video to mobile handsets. The service is described in terms of actions taken by the WAP MMS Client and its service partner, the MMS Proxy-Relay, a device that operates as a WAP Origin Server for this specialized service. The transition from Short Message Service (SMS) to MMS is as important on mobile phones as the transition from DOS to Windows was for the PC. MMS presents a revolution to the end user in terms of the richness of the messaging services. However, it is delivered in an evolutionary manner for the infrastructure, since it will be reusing components from existing SMS, email, Unified Messaging and other platforms.

MMS is like SMS, a non-real time service, a relay platform that routes multimedia messages to MMS Servers. Hence, it is not sensitive to delay. However, it is sensitive to packet loss, and it is limited in terms of the available bandwidth and the quality of the media sent. Nonetheless, there are recent additions to MMS such as Instant Messaging, which are more sensitive to delay. Generally, such low-speed data services have lower traffic and bandwidth requirements and can be modeled in a similar fashion to email traffic, while taking into account the mobility challenges. Once it starts transmission, it streams at its peak allowable rate to an idle state where no packets are generated [54].

3.1.4 High-Speed Data Services: High-Speed Internet Access

High-speed Internet access, such as xDSL, cable, EVDO, HSDPA or WiMAX, is becoming a necessity for business users and consumers. As we see more services and

applications migrate and get created over this access, it becomes critical to understand the impact on user-behavior, thus source traffic characteristics and models.

It has been proven that Internet traffic exhibits strong correlation and is bursty in nature. It has also been shown that general properties of Internet traffic are almost independent of the access speed, such as the shape of the up/downstream session volume, the shape of the TCP-connection volume and the TCP connection data rate. However, the traffic volume and the obtained data rates were found to increase proportionally to the access speed. With respect to interactive applications, the rise of traffic volume and rate is less than the corresponding increase in modem speed. Also the type of services utilized is dependent on the access speed.

As a result, high-speed access traffic can be modeled to generate traffic at certain burst rates (depending on the access speed and applications: HTTP, FTP, e-mail, peer-to-peer applications, database access etc.), and to an idle state with no packets generated. Current services are very sensitive to packet loss, but they can tolerate packet delay. However, with new services such as VoIP and VoD, delay sensitivity becomes critical and must be taken into account [55, 56].

3.2   Source Levels and Models

Source modeling is essential for a precise definition of the behavior of a traffic source. Traffic models must have good tractability, a manageable number of parameters and a simple method to estimate them. A traffic source can be characterized at three levels: connection, burst and packet, as shown in Figure 3-1 (a). The connection level describes the source traffic from the connection setup until the disconnections. Each connection is composed of alternating silence (no traffic) and burst periods. Each burst period is composed of packets with variable packet lengths. The packet level can be calculated as the average packet length divided by the source transmission rate.
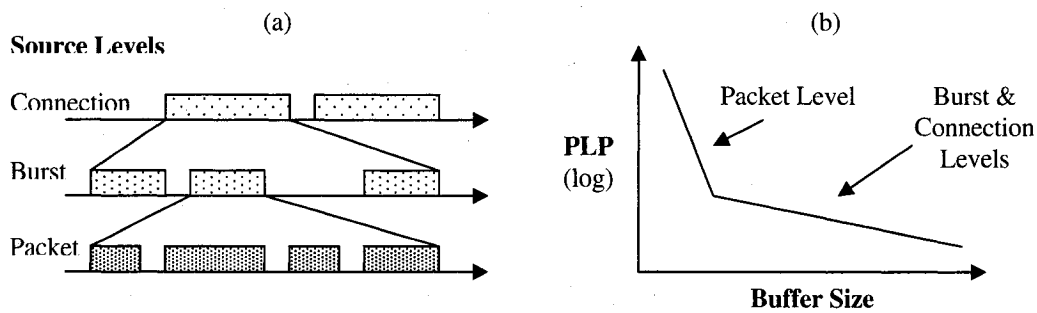
Figure 3-1: (a) Source Levels and (b) Packet Loss Probability (PLP)

Traffic modeling is used in traffic engineering to predict network performance and to evaluate congestion control schemes. One of the techniques in congestion control is based on estimating the Packet Loss Probability (PLP) of an IP/MPLS/RPR node. This method allows us to obtain a good estimate of the QoS and ability to accept or refuse a connection or service request. As shown in Figure 3-1 (b), there is a relationship between the source level, buffer size and PLP. For smaller buffer sizes, the packet loss is dominated more by packet level fluctuations. While for larger buffer sizes, the packet loss is more attributed to source burst and connection level traffic, since connections can be modeled as burst level with larger periods.

Each of the broadband services discussed in the previous section has its unique attributes and can be best represented by certain traffic models, which can be classified into two classes: short-range dependent (SRD) and long-range dependent (LRD) models. SRD models have a correlation structure that decays at least at an exponential rate. Examples of SRD models are Renewal models, Markov processes, Markov-modulated processes and Regression models. LRD traffic models have a correlation structure that decays slower than exponential rate. Fractional and Superposition of high variability On-Off sources are examples of LRD models [51, 57, 58, 59, 60, 61, 62].

3.2.1 Short-Range and Long-Range Dependence

In this section, we examine stationary traffic models, which are classified into two-classes: *short-range dependent* (SRD) models, which have a correlation structure that decays at an

37

exponential rate, and *long-range dependent* (LRD) or self-similarity models with a correlation structure that decays slower than exponential rate. Figure 3-2 compares the auto-correlation functions of an SRD and LRD processes. The auto-correlation function of a wide-sense stationary stochastic process $\{X_t\}$ at lag k, where $t = 0, 1, \dots$ is defined as ζ_k :

$$\zeta_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{E[(X_t - \mu)^2]}$$

(3-1)

where μ is the stationary mean is $E[X_t]$ and ζ_k depends only on k and not on t.
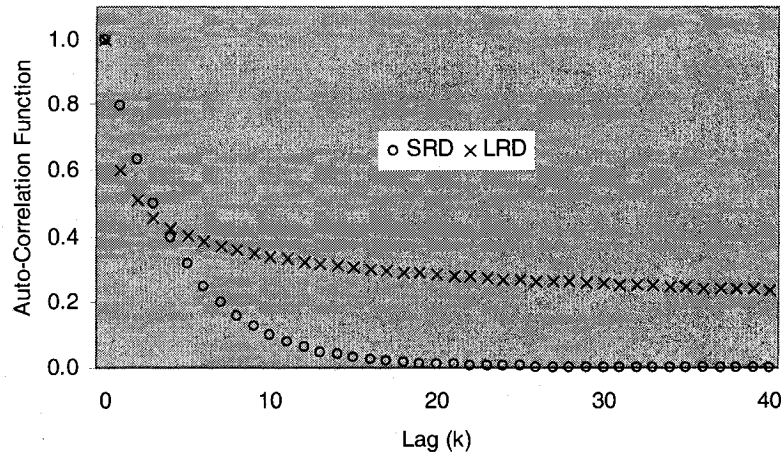
Figure 3-2: Autocorrelation functions for SRD and LRD processes

Also, for each m, let $\{X_j^{(m)}\}$ represent a new time series obtained by averaging the original series $\{X_t\}$ over non-overlapping blocks of size m. The process $\{X_t\}$ is said to be <u>exactly (2nd order) self-similar</u> to its auto-correlation function, $\zeta_k^{(m)}$ for all values of m and k, is preserved across different time scales, and <u>asymptotically self-similar</u> if , $\zeta_k^{(m)} \rightarrow \zeta_k$ for large m and k. Stochastic self-similar processes retain the same statistics over a range of scales, they satisfy the following relationship:

$$\{X_{at}\} \underline{\underline{D}} \ a^H \{X_t\}$$

(3-2)

where $\underline{D}$ denotes equality in distribution, and H is the Hurst parameter which indicates the degree of self-similarity. This relationship means that sample paths appear to be the same irrespective of the time scale.

We will cover two types of SRD models: Markov processes and Regression models. Markov processes include renewal process, alternating renewal process, and Markov Modulated models. Regression models cover Autoregressive (AR) models, AR Moving Average models and Transform Expand Sample (TES) models. As for LRD, we will examine Pareto Distributed Process, Fractional Brownian Motion (FBM), Fractional Autoregressive Integrated Moving Average (F-ARIMA) and Superposition of High On-Off sources [52, 57, 58, 51].

3.2.2 Markov Property and Models

Given a state space $S = \{s_1, s_2, ..., s_M\}$, let X_n be a random variable (RV) that defines S at time n. The set of RVs $\{X_n\}$ form a discrete *Markov chain* if the Pr $\{X_{n+1} = s_j\}$ depends only on the current state. If the state transitions occur at integer points, the Markov chain is discrete time; otherwise, it is continuous time. Hence, *Markov property* implies that the future state depends on current state only, and neither on time spent in current state nor any previous states. As a result, the time spent in a Markov chain state can be modeled as a geometric distribution in the discrete case, and exponential distribution in the continuous case. In a simple Markov traffic model, each state transition represents a new arrival. Hence, inter-arrival times are exponentially distributed in the continuous case, and their rates depend on the state from which the transition occurs. A *semi-Markov process* is obtained by allowing the time between state transitions to follow an arbitrary probability distribution. Ignoring the time distribution between state transitions, the sequence of states visited by the semi-Markov process is known as a *discrete-time Markov chain* and referred to as an *embedded Markov chain* [58, 63].

3.2.2.1 Renewal Process

A renewal process has inter-arrival times that are statistically independent and identically distributed. Since there is no correlation between inter-arrival times, their distribution

39

completely defines the process. Bernoulli and Poisson processes are special cases of renewal processes where the inter-arrival times are geometrically and exponentially distributed, respectively. The Bernoulli (Geometric) process provides a good approximation for short-term and packet level behavior, while the Poisson process has been used for jitter analysis in Constant Bit Rate (CBR) ATM node. Nevertheless, the Poisson process does not model well the IP/MPLS based traffic [52, 59].

3.2.2.2 Alternating State Renewal Process

The alternating state renewal process is a two state process s_0 and s_1 with no self transition. As a result, the embedded Markov chain alternates between these two states. There is no traffic generated in s_0, while traffic is generated during s_1. The mean sojourn time in s_0 is d_0 and for s_1 is d_1. The steady state probabilities of being in s_0 and s_1 are:

$$P(s_0) = d_0 \Big/ (d_0 + d_1) \text{ and } P(s_1) = d_1 \Big/ (d_0 + d_1)$$

(3-3)

Note that the superposition of identical and independent alternating state renewal processes has a binomial distribution.

On-Off model and Interrupted Poisson Process (IPP) are special cases of alternating renewal processes. The On-Off model, also known as Talkspurt-Silence model or Burst-Silence model, is widely used for modeling voice traffic and provides good approximation for burst level congestion, but not suitable for packet-level congestion. During an On period traffic is generated with a fixed inter-arrival time, and an Off period where no packets are generated. The time spent in each of the states is exponentially distributed with mean $1/\alpha$ for the On state, and $1/\beta$ for the Off state. Hence using (3-3), the probabilities of being in each of the states are:

$$P(s_{On}) = \beta \Big/ (\alpha + \beta) \text{ and } P(s_{Off}) = \alpha \Big/ (\alpha + \beta)$$

(3-4)

IPP is the same as the On-Off model except during the On state, the inter-arrivals times are exponentially distributed and arrivals are Poisson distributed with rate λ [52, 58, 64].

3.2.2.3 Markov Modulated Poisson Process (MMPP)

A Markov modulated process, also known as doubly stochastic process, uses an auxiliary Markov process in which the current state of the Markov process modulates the probability distribution of the traffic. Note that the On-Off model is a special case of the Markov Modulated Deterministic Process (MMDP) with two states.

MMPP is a versatile class of traffic models, part of a class of processes known as Switched Poisson Processes (SPP), which can model time-varying sources while keeping the analytical queueing performance tractable. Figure 3-3 shows an n-state MMPP which uses Poisson as the modulated mechanism, while in state s_i, the arrivals occur according to a Poisson process with λ_i rate. Also, the transition rate from state s_i to state s_j is denoted as $r_{i,j}$.

Note that an n-state MMPP can be obtained by the superposition of $n-1$ identical independent IPP sources. As a result of the Poisson additive property, MMPP can effectively model a mixture of different services. Given state s_i and the arrival rates active of λ_{t1}, λ_{t2} and λ_{t3} in the same state, the resulting arrival rate of state s_i is $\lambda_{t1}+\lambda_{t2}+\lambda_{t3}$. Note that as the number of states increases, it becomes more difficult analytically to fit real data using this model. On the other hand, MMPP parameters such as state rates and transition rates can be easily estimated from data traffic [51, 52, 58, 59, 61, 64].
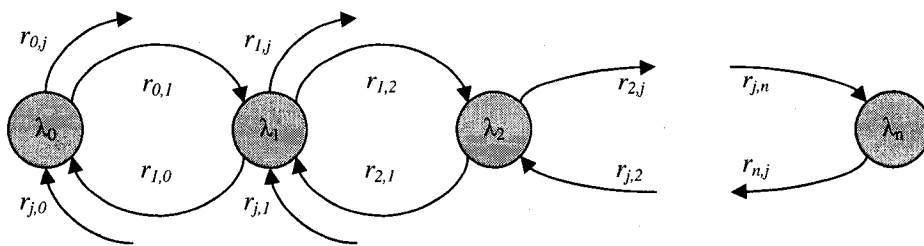
Figure 3-3: n-state Markov Modulated Poisson Process (MMPP)

3.2.2.4 Markov Modulated Fluid Models

Fluid models characterize a discrete packet stream by a continuous flow of information. This abstraction covers only the long-term variations of the arrival modeled process. Thus,

41

this model is appropriate when packet short-term increases have little impact on network performance. Fluid models are conceptually simpler as they react to flow rate changes as opposed to per packet arrivals, which results in significant computing savings. In Markov Modulated Fluid model, the current state s_i of the Markov chain determines the constant flow rate λ_i. This model is used to model VBR video sources [52, 58].

3.2.3 Autoregressive Models

Regression models define explicitly the next RV in the sequence by its previous ones within a certain window of time and a moving average of white noise. An Autoregressive model of order p is denoted as AR(p). AR models are used to model the output bit rate of VBR encoders as successive video frames do not vary much, specifically the following AR(1) model is used where the bit rate within a frame is constant and changes per frame.

$$b_n = \varepsilon b_{n-1} + \gamma Y_n \qquad\qquad (3\text{-}5)$$

where b_n is the bit-rate during frame n, Y_n is independent of the bit rate and assumed to be a Gaussian white noise, ε and γ are real numbers. This model does not capture abrupt changes in the frame due to scene changes, for example. Hence, a viable approach is to model the bit rate of frames within a scene as an AR process, and abrupt scene changes by an underlying Markov chain. An AR process has an exponentially decaying auto-correlation structure which makes it an SRD model; therefore, it is not suitable to model self-similar traffic [57, 58].

3.2.3.1 Discrete Autoregressive Models

A discrete AR model of order p is denoted as DAR(p). A DAR(p) generates a stationary sequence of discrete RV's with an arbitrary probability distribution and auto-correlation structure similar to AR(p). DAR(1) is a special case of the DAR process and can be defined as follows:

$$X_n = V_n X_{n-1} + (1 - V_n) Y_n \qquad\qquad (3\text{-}6)$$

where $\{X_n\}$, $\{V_n\}$ and $\{Y_n\}$ are sequences of RVs. The RV V_n takes two values 0 or 1, with a probability of 1-p and p, respectively. The RV Y_n has a discrete state space S and $P\{Y_n = i\}$ $= \pi(i)$. This results a discrete Markov chain with discrete space S and a transition matrix. DAR(1) is used to model the number of packets per frame of teleconferencing VBR video. Its analytical queueing performance is tractable and has much less parameters than general Markov chains. However, its fast decaying auto-correlation function makes it less suitable to model self-similar traffic.

3.2.3.2 Autoregressive Moving Average Models

An Autoregressive Moving Average model of order (p, q) is denoted as ARMA (p, q). It is used to model VBR traffic. The duration of a video frame is equally divided into v time intervals. The number of packets in the nth time interval is modulated as follows:

$$X_n = \varepsilon X_{n-v} + \sum_{k=0}^{v-1} \theta_k Y_{n-k} \qquad (3\text{-}7)$$

The auto-correlation function has peaks at all lags which are integer multiples of v. The parameter estimation of ARMA models are more involved than that of AR models. Also, the analytical models are more involved. Another type of AR model is the Autoregressive Integrated Moving Average model of order (p, d, q) known as ARIMA (p, d, q) which is an extension of the ARMA (p, q) model. It is used to model homogeneous non-stationary time series.

3.2.3.3 Transform-Expand-Sample (TES) Models

TES models are non-linear regression models with modulo-1 arithmetic. Their objective is to capture both auto-correlation and marginal distribution of empirical data. A combination of TES processes can be used to better fit the auto-correlation function. The empirical distribution is matched using the distribution inversion methods. Note that the auto-correlation function of TES decays exponentially; hence, and it is an SRD model.

43

3.2.4 LRD & Self-Similarity Models

In section 3.2.1, we discussed short-range and long-range dependent models and their relationship to the auto-correlation function. In this section, we focus on the key LRD models, namely: Pareto Distributed Process, Fractional Brownian Motion (FBM), Fractional F-ARIMA and Superposition of High On-Off sources [51, 57, 58, 74, 65].

3.2.4.1 Pareto Distributed Process

The Pareto distribution is a distribution with memory, heavy tail, and strong burstiness. It is suitable for modeling self-similar arrival in packet traffic. The Pareto distribution produces independently identically distributed inter-arrival times. It has the following probability density function:

$$f(t) = \frac{ab^a}{(t+b)^{a+1}} \qquad (3-8)$$

where a is the shape parameter or tail index, and b is the scaling parameter. In the region $0 < a < 2$, the distribution is heavy tailed with infinite variance. With $a < 1$, expected value is also infinite. Output of the distribution is 'conventional', when $a > 2$, because expected value and variance are both finite. One method to calculate a is from the Hurst parameter, as $a = 3 - 2H$, where H can be derived from the Index of Dispersion for Counts (IDC) or Block Packet Count (BPC) which is the number of packet arrivals in a fixed time interval. Also, we can obtain b using the mean of the measured traffic E[T] and a, as $b = (a - 1)$ E[T].

3.2.4.2 Fractional Autoregressive Integrated Moving Average (F-ARIMA)

The fractional ARIMA process, F-ARIMA (p, d, q) with $0 < d < \frac{1}{2}$, is a stationary process with LRD. A special case is when $p = q = 0$, the resulting auto-correlation function for F-ARIMA (0, d, 0) is shown below and plotted in Figure 3-2

$$\zeta_k = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \approx \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \text{ as } k \to \infty \qquad (3-9)$$

where $\Gamma(x)$ denotes the gamma function, $d = H - 0.5$ for the range of d values, and H is the Hurst parameter. Note that F-ARIMA can model short-range and long-range dependence. F-ARIMA is used to model VBR video traffic.

3.2.4.3 Fractional Brownian Motion (FBM)

Brownian motion is a stochastic process denoted by $\{B_t\}$ for $t \geq 0$. Fractional Brownian motion $\{fB_t\}$ is a Gaussian self-similar process with Hurst parameter $H \in [0.5, 1)$. FBM process characterizes the number of arrivals N_t in the time interval (0, t) as:

$$N_t = mt + \sqrt{\phi m} Z_t, \quad t \in (-\infty, \infty)$$
$$m = E[N_t] \text{ and } \phi = \frac{Var[N_t]}{mt^{2H}}$$

(3-10)

where Z_t is a normalized FBM, m denotes the mean arrival rate, and φ is the variance coefficient of the process. Note that for $H = 0.5$, Z_t is reduced to an ordinary Brownian motion. In addition, the auto-correlation of this process can be expressed as follows:

$$\zeta_k \approx k^{-2(1-H)}, \text{ as } k \to \infty$$

(3-11)

For large H, increasing the utilization requires a significant amount of storage. Also, the packet loss probability decreases algebraically with buffer size and exponentially as with Markovian and ARMA models.

3.2.4.4 Superposition of High On-Off Sources

Superposition of High On-Off sources method is an extension to the traditional On-Off model which assumes finite variance distributions for the sojourn time in each of the two periods. As a result, the aggregation of a high number of sources in the On-Off model will not have significant correlation. Instead, the superposition method allows the sources to have infinite variance which results in producing aggregate traffic with long-range dependence. During the On states, sources transmit at a constant rate, and the time spent in these states is an independent identically distributed RV T with a Pareto-type distribution

$P\{T > t\} \sim t^{-b}$, where $1 < b < 2$. This results in an aggregate process which is asymptotically self-similar with auto-correlation and Hurst parameters as follows:

$$\zeta_k \approx k^{1-b} \text{ for large } k, \text{ and } H = (3-b)/2 > 0.5 \qquad (3\text{-}12)$$

Similar to FBM, the packet loss probability decreases algebraically with buffer size. This method has been used to model Ethernet LAN traffic.

3.3   Traffic Modeling Research

The research community and industry have many studies in the literature dealing with the statistical performance of multimedia networks with a focus on IP-based services. Most of the IP research conducted has been centered around the Core of the network due to the focus on the Internet [61, 66, 67, 68, 69]. It is well known now that IP-based traffic exhibits packets burstiness and correlation in inter-arrival times [70, 71]. Moreover, a significant portion of the performance analysis has been carried assuming an ATM infrastructure underpinning the IP layer [58, 51, 72, 73, 74, 75, 76].

Many of these studies are based on the work done on core data networks in the 80's, evolving to ATM based networks in the 90's, which are currently migrating to IP/MPLS/RPR infrastructure. Obviously, the underlying assumptions are similar; nevertheless, there are key differences with these networks such as bandwidth, delay and QoS requirements. It is also recognized that the Poisson model, used to model voice traffic, is not suitable for modeling today's data traffic which exhibits burstiness and self-similarity. In addition, service convergence of multimedia traffic at the customer premises is now becoming technologically feasible and cost-effective across the network. As a result, the need to study the impact of convergence across the network, starting from the Access to the Edge and Core, is more critical nowadays to enable successful and cost-effective service delivery. In this section, we focus on the key research which has been carried out on packet based networks with multimedia traffic.

3.3.1 Compound Poisson Process (CPP) with GE Distribution Service

CPP and Markov Modulated CPP (MM CPP) are models suited for producing traffic burstiness in which inter-arrival times have a Generalized Exponential (GE) probability distribution. The *MM CPP/GE/m/K* G-queue is capable of accommodating large batch sizes of arrivals and services with an efficient solution. Nonetheless, it suffers from the restriction on the batch size distributions. The model provides a homogeneous multi-server queue with m servers, General Exponential services times and K buffer length. The $MM \sum_{i=1}^{N} CPP_i / GE / m / K$ G-queue overcomes this restriction by making the model vastly flexible and also enables the superposition of multiple arrivals streams, where the superposition of N independent positive and independent negative customer arrival streams each as a CPP (with batch arrivals of geometrically distributed batch size). This means that the inter-arrival times of each of the *(N+1)* arrival streams are also independent GE RVs. This model is mathematically tractable with efficient analytical solution for the steady state probabilities which models well the burstiness (batch size distribution) of the traffic but not the auto-correlations of the inter-arrival times.

This model has been proposed for use in MPLS networks. The arrival of packets can be modeled by the $MM \sum_{i=1}^{N} CPP_i$ process. Multi-path LSPs are modeled as multi-servers with GE service times, a maximum of m servers (or LSPs) and the capacity of the node as K. The use of negative customers represented by G- is used to model packet losses, corrupted packets, and breakdown of LSPs etc. This specific application is more representative of the network Core, and can be extended to the Edge [77, 78, 79, 80].

3.3.2 MMPP with General Distribution Service

Markov Modulated Poisson Process (MMPP) is a doubly stochastic Poisson process where the Markov Chain (MC) governs the transition of the process between states, and at each state the number of arrivals is determined by a Poisson process. The most popular and simple model of MMPP is the 2-state MMPP which has been successfully used to model aggregate traffic from voice, data and variable-bit rate video sources. MMPP model is capable of catching the inter-frame dependency between consecutive frames. A 2-state

MMPP, referred to as MMPP-2, process has four parameters: $\lambda_1, \lambda_2, r_1, r_2$ which are the packet arrival rates in states 1 & 2 and the mean state durations in states 1 & 2, respectively. MMPP-2 can be parameterized by the infinitesimal generator matrix R and the arrival rate matrix Λ, as follows:

$$R = \begin{bmatrix} -r_1 & r_1 \\ r_2 & -r_2 \end{bmatrix} \text{ and } \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad (3\text{-}13)$$
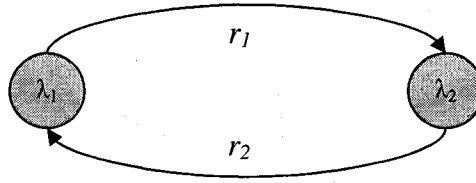
Figure 3-4: 2-State MMPP (MMPP2) and 2-state SPP (SPP-2)

Let π be the equilibrium probability vector $= (r_2/(r_1+r_2),\ r_1/(r_1+r_2))$, where $\pi R = 0$. Also, $\pi e = 1$, where e is the unit column $= (1, 1)^T$, and the traffic intensity or average traffic generation rate can be expressed as:

$$\rho = \bar{\lambda} = \pi \Lambda e = \frac{\lambda_1 r_2 + \lambda_2 r_1}{r_1 + r_2} \qquad (3\text{-}14)$$

This work demonstrates that a superposition on N on-off sources can be modeled by an MMPP-2. Several methods have been proposed for matching the MMPP-2 four parameters to the parameters of the aggregated multimedia traffic of on-off sources such as, Moment-based matching, Σ-Matching technique and IDC matching technique [59, 62, 81, 82, 83].

3.3.2.1 Moment-Based Matching

Moment-based matching technique by Heffes and Lucantoni computes the four MMPP-2 parameters using the first moment (mean arrival rate), two IDC points representing the 2nd moment (variance-to-mean ratio and asymptotic value of IDC i.e. IDC(∞)), and one point of the third central moment of the superposition of N on-off sources. In their work, they

modeled aggregated voice and data traffic multiplexed over a high-speed link using MMPP/G/1 queue. The result of this model is extremely well matched with IDC curve of the aggregate traffic. However, it requires a lengthy calculation of the third moment using inverse Laplace transforms [59].

3.3.2.2 Σ-Matching

Using Σ-Matching technique by Kim and Lee, the MC is divided into two parts: an under-load (UL) and over-load (OL) representing each of the MMPP-2 states. The arrival rates λ_{UL} and λ_{OL} are computed based on the on-off sources mean sojourn times and their respective stationary probabilities. The MMPP-2 transition rates, r_{UL} and r_{OL}, are determined by equating the mean sojourn time in each of the states to that in all of the corresponding states in MC using a recursive equation. The advantage of this technique is that it is applicable to a heterogeneous case, and it provides a simple way to estimate the MMPP-2 parameters. Nonetheless, it does not match the IDC curve; hence, it fails to effectively capture the inter-frame correlation of the aggregate traffic under heavy loads [82].

3.3.2.3 IDC Matching

In IDC matching technique uses the same under/over-load approach as mentioned the previous sub-section, but instead of the time process, LeNgoc, Huang and Hayes propose using a counting process of the number of arrivals of over a fixed time interval. Thus, the sojourn time of the MMPP-2 model are calculated by matching the values of the IDC at large lags i.e. IDC(∞). This calculation is very simple as the values of IDC(∞) are known for both the on-off sources and the MMPP-2 model. The results of this method are the closest to the moment matching technique; however, it is not as accurate. On the other hand, it provides a simpler but accurate method to estimate the parameters and analyze the queueing system [62, 83].

3.3.2.4 Superposition of MMPPs

One of the key characteristics of MMPP processes is that the superposition and splitting of MMPPs give rise again to an MMPP. The superposition of N MMPP processes can be represented, as follows:

$$R = R_1 \oplus R_2 \oplus \Lambda \oplus R_N \text{ and } \Lambda = \Lambda_1 \oplus \Lambda_2 \oplus \Lambda \oplus \Lambda_N \quad (3\text{-}15)$$

where R_i and Λ_i are $k_i \times k_i$ matrices, and R and Λ are $k \times k$ matrices with $k = \prod_{i=1}^{N} k_i$. Also, $\oplus$ denotes the Kronecker sum which is defined as:

$$R_1 \oplus R_2 = \left(R_1 \otimes I_{R_2} \right) + \left(I_{R_1} \otimes R_2 \right) \quad (3\text{-}16)$$

I_{Ri} is an identity matrix of the same order as matrix R_i, and $\otimes$ denotes the Kronecker product, which is defined for two matrices $C = \{c_{ij}\}$ and $D = \{d_{ij}\}$ as:

$$C \otimes D = \begin{bmatrix} c_{11}D & c_{12}D & \Lambda & c_{1m}D \\ M & M & M & M \\ c_{n1}D & c_{n2}D & \Lambda & c_{nm}D \end{bmatrix} \quad (3\text{-}17)$$

Andersen and Nielsen proposed an approach where a number of MMPP-2's are superposed to model traffic self-similarity over many time scales. The parameters of each of the MMPP-2's are determined so as to match the mean and auto-correlation function of the self-similar traffic over several different time scales. An algorithm is also provided for fitting the covariance function of the Markovian model to that of second order self-similar processes over several time scales, and demonstrated that it works well over the entire range of the Hurst parameter. The superposition of three to four MMPP-2's is sufficient to model self-similar behavior reasonably accurate over four to five time scales. The advantage of this method is its capability to model LRD using well-developed queueing systems [84, 85].

50

Other recent research applied the superposing of MMPP-2's to model traffic in Wireless ATM/TDMA queue system and to model message arrivals in Dimension-Ordered routing in multi-computer networks [86, 87].

In summary, MMPP models, which are SRD processes by definition, are widely used and adopted for queueing systems due to its simplicity and analytical tractability. MMPP models are suited to capture the auto-correlation of the inter-arrival times especially for lower loads and smaller buffer applications. In some cases, it is unable to capture the burstiness of the aggregate multimedia traffic. This downfall could be partly addressed by increasing the number of MMPP states; however, this results in increasing the analytical complexity and the time to convergence/solution. In addition, recent research demonstrated that a superposition of MMPP processes can model certain LRD traffic behavior of several different time scales.

3.3.3 Pareto Distributed Inter-Arrival Times

Pareto Modulated Poisson Process (PMPP) is a doubly stochastic process with a heavy-tailed inter-arrival time Pareto distribution, which can be used to capture self-similarity and LRD behavior in traffic. The mathematical representation presented in 3.3.2 for MMPP is also applicable to PMPP as they are both SPPs, and instead of the four MMPP-2 parameters, PMPP-2 requires three parameters, namely: λ_1, λ_2, a which are the arrival rates in states 1 & 2 and the Pareto model shape parameter, respectively. Hence, when the process is in state i, the packets arrive at a Poisson rate of λ_i and with Pareto distributed inter-arrival times with shape parameter a. In other words, the sojourn times in each of the states are independent and identically distributed with Pareto distribution. The Pareto shape parameter has three ranges of interest: (i) $0 < a < 1$, (ii) $1 < a < 2$ and (iii) $a > 2$. In case (iii), the mean and the variance of the inter-arrival distribution are finite. Thus, the process is not self-similar in this region. In case (ii), the mean is finite, but the variance is infinite. While in case (i), both the mean and the variance are infinite; hence, the utilization is zero for this region. This means that for case (i) and (ii), the arrival process is self-similar. Hence, case (ii) is the region of interest to generate self-similar traffic with utilization > 0.

51

Several researchers have studied the Pareto model specifically: Gordon analyzed Pareto/M/1 queue delay, Koh and Kim derived loss probability in Pareto/M/1/K, Huebner compared Pareto's loss probability and queue lengths against other models, Subramanian and Le-Ngoc presented an aggregated model of PMPP and MMPP for data and real-time traffic, respectively. The findings demonstrated that Pareto distributed inter-arrival can capture self-similarity and performs better than other models for smaller to medium buffer size LRD applications [57, 88, 89, 90].

3.3.4 FBM with Deterministic Distribution Service

The Fractional Brownian Motion (FBM) process, described in section 3.2.4.3, is used to model connectionless traffic with self-similarity as presented by Norros. He demonstrated that FBM accurately describes internal LAN traffic where many sources could be sending traffic at the same time. However, external and WAN traffic, which has less intensity, is not well described by FBM, despite the traffic's 2nd order self-similarity. He showed that the queue length for an FBM is simpler and more tractable than other typical LRD models which are very complex and difficult to study [91, 92, 93, 94].

Moreover, Huebner presented two comparative studies for FBM. He modeled an FBM/D/1 infinite queue and used it to approximate the loss probability in finite queues. The result for the approximation did not work well under the LRD FBM process and was a difference of a few orders of magnitude [65]. Nonetheless, in another study, he compared various modeling techniques including FBM to see which model captures better the statistical properties of data traffic. The result indicated that FBM modeled well the effect of long-range dependence in data traffic, especially for applications with larger queue size; however, the FBM model failed to predict the queueing performance for small buffer size [57].

3.4 Ring and RPR Performance Research

Ring technologies have received a lot of positive endorsement within the research and industry communities. One of the key technologies is Buffer Insertion Ring (BIR), which

is the basis that RPR was built on. RPR introduced several key properties on top of BIR such as resilient failover (50ms protection), fairness algorithm, and class of service hierarchy. As a result, RPR is receiving strong focus from Service Providers for delivering a converged infrastructure. Nonetheless, the current RPR standard suffers from: (a) slow convergence of the fairness algorithm, (b) permanent oscillations of the fairness algorithm, (c) long ring access delay, (d) total (i.e. dual-ring) bandwidth inefficient utilization, and (e) inconsistent class of service traffic handling. The first three items have been researched quite extensively in recent years. Several proposals demonstrate alternative solutions to address these limitations. However, items (d) and (e) have not been resolved yet, which we attempt to address in this research. A summary of the key contributions to BIR and RPR research is provided next.

3.4.1 Buffer Insertion Ring

Buffer Insertion Ring (BIR) was proposed by Hafner, Nenadal and Tschanz as an unslotted ring allowing for uncoordinated access to the ring and use of variable sized packets [95]. BIR has the maximum spatial reuse on a ring allowing stations to transmit packets concurrently. A node may enter a state where it has packets ready for transmission; however, it is not able to transmit on the ring due to upstream node traffic. If this state lasts for a long period of time, this state is called congestion or starvation. Therefore, to avoid packet loss, each node requires an infinite buffer also known as insertion buffer, as shown in Figure 3-5. Nonetheless, an infinite or large buffer is not practical and results in unacceptable long packet delays [96, 97].
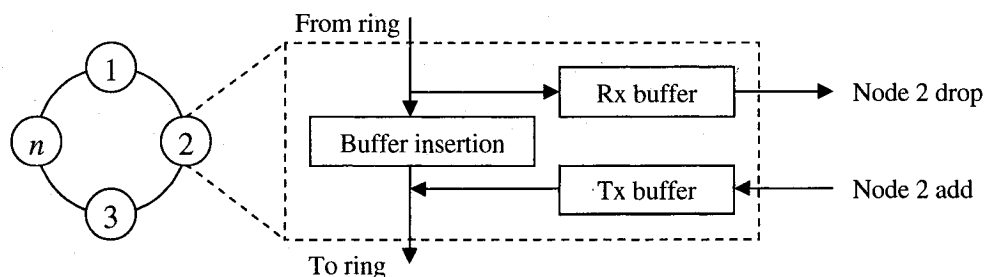
Figure 3-5: Buffer Insertion Ring and Node Architecture

As illustrated in Figure 3-5, if the traffic generated from node 1 and destined to node 3 is equal or larger than the ring bandwidth, and without any access control or fairness mechanism, node 2 will not be able to insert any of its traffic onto the ring, which results in node 2 starvation.

In order to limit the size of the insertion buffer, eliminate starvation and minimize packet delay, many researchers introduced new techniques to provide fair access to all nodes on the ring while minimizing delay and maximizing throughput. In [98], Head-Of-Line (HOL) delay is used as a performance measure for access delay. HOL is defined as the amount of time the first packet in the transmission buffer of a node has to wait before it is inserted on the ring. HOL delay or the range of its variations is a major component in the transmission jitter that the ring introduces. This research introduced two algorithms which demonstrate a reasonable tradeoff between packet delay and ring throughput with an upper bound for HOL delay. The delay bound derived for the first algorithm is:

$$6\frac{Q_{max}}{B_w} + 3T_{RD} + 2n\frac{L_{max}}{B_w} \qquad (3\text{-}18)$$

Similarly, the HOL delay bound presented for the second algorithm is:

$$4(\ln n + 1)\frac{Q_{max}}{B_w} + 3T_{RD} + 2n\frac{L_{max}}{B_w} \qquad (3\text{-}19)$$

where L_{max} is the maximum size of a packet, B_w is the link bandwidth, n is the number of nodes on the ring, T_{RD} is the round trip delay and Q_{max} is the maximum quota allowed for any node to transmit on the ring.

In [96], Buffer Insertion/Self-Token (BIST) protocol was introduced to prevent starvation and maintain fairness on the ring. BIST assigns each node individual tokens, called self-tokens. A node can transmit packets only if it holds its own tokens. Hence, each packet transmitted by a node is attached a self-token belonging to the node. This results in regulating packet flow and providing fair access to the ring. Two modes of self-tokens are

54

supported: asynchronous and synchronous transmissions. The asynchronous traffic is associated with station priority transmission mode, while the synchronous traffic is related to ring priority transmission mode.

Another BIR research is presented in [97]. Two BIR architectures are studied, namely Mono Transit Buffer (MTB) and Dual Transit Buffer (DTB). DTB architecture allows the ring to differentiate between low priority and high priority traffic, which is important during congestion periods. This research introduced congestion control mechanisms which are rate based and proposed to regulate traffic using leaky buckets. The research demonstrated through simulation that MTB architecture oscillated due to over-reaction of rate estimations, while DTB oscillated due to buffer threshold settings. To achieve fair throughput for all nodes, the research showed that by correctly setting parameters, oscillations can be dampened.

3.4.2 ABA Scheme

Adaptive Bandwidth Allocation (ABA) scheme is a distributed bandwidth allocation scheme based on classical control theory. It is proposed in 2005 by L. Tan, H. Wang and M. Zukerman [8]. ABA's objective is to (a) achieve General Weighted (GW) fairness [99] amongst various flows sharing a common link, (b) maintain stable queueing delay, and (c) maximize the link utilization.

This scheme assigns the Minimum Rate (MR) for each flow, and then fairly allocates a portion of the excess bandwidth in accordance with each flow's weight. Note that a portion of the excess bandwidth is reserved for congestion avoidance. Then, ABA applies a Proportional Derivative (PD) control approach to re-adjust the sending rates of these flows at the relevant sources. This results in maximizing the link utilization while maintaining stable buffer occupancy. The following PD-controller is used to control the sending rate of one flow which passes through a given link:

$$R(t) = R_{max} - a(Q(t - \tau_b) - Q_0) - b(Q(t - \tau_b) - Q(t - \tau_b - 1) \qquad (3\text{-}20)$$

where $R(t)$ is the adjusted sending rate of the flow at its source, R_{max} is the maximum allowed sending rate of this flow, τ_b is the feedback delay from the link to the source, $Q(t)$ is the buffer occupancy, Q_0 is the target value of $Q(t)$, lastly a and b are control gain parameters. These two parameters are determined by solving the following characteristic equation (CE) of the closed-loop system, using nth-order Taylor expansion of the exponential functions with the condition in the control gains under which the network is asymptotically stable in terms of buffer occupancy:

$$s + a\sum_{i=1}^{N}\omega_i e^{(-\tau_f^i - \tau_b)s} + b\sum_{i=1}^{N}\omega_i e^{(-\tau_f^i - \tau_b)s} - b\sum_{i=1}^{N}\omega_i e^{(-1-\tau_f^i - \tau_b)s} = 0 \qquad (3\text{-}21)$$

note that τ_f^i is the forward time delays from the source node i, and ω_i is the weight of flow from node i.

The selection of the control parameters is important to achieving system stability. Also, the simulation and the theoretical analysis demonstrate that the ABA scheme address the RPR limitations (a) and (b) while maximizing on network utilization.

3.4.3   Moderate Fairness Mode Algorithm

F. Davik, A. Kvalbein and S. Gjessing introduced in 2005 another fairness algorithm to improve two of the major performance deficiencies in the RPR fair algorithm. The first deficiency is caused by the method used by the node closest to the congested link to calculate a fair rate estimate. The second deficiency is in the method used for distributing the calculated rate values [9].

The goals of the moderate fairness mode algorithm are as follows. First, eliminate the oscillation behavior in the presence of a head with a modest sending behavior. Second, retain the RPR standard behavior in the presence of a head with greedy sending behavior. Third, minimize the changes to current algorithm for faster implementation. Lastly, fit the modifications within the guidance of the RPR standard.

56

The basic idea of this algorithm is to monitor the TB occupancy of the congested node. If the occupancy increases, the fair rate estimate is too high and must be decreased. Alternatively, if the occupancy decreases, the fair rate is too low and must be increased. The occupancy must be bound to minimum and maximum allowed rates. This process results in improving the rate calculation.

In order to enhance the rate distribution, the algorithm proposes to alter the responsibility of the tail node (during congestion), so that it is no longer its responsibility to stop the propagation of the fair rate estimates received from the head (where the congestion is detected). However, to support the establishment of more than one congestion domain, the algorithm allows the removal of the fairness message once it reaches a new congestion domain head. To allow for a congested node to decide whether a fairness message has passed through a downstream tail or not, this algorithm proposes to use one of the 13 reserved and unused bits as a "passed_tail" bit. When the fairness message reaches the tail, the tail sets this bit to one; hence, indicating that the message has been propagated beyond a congestion tail. Fairness messages with this bit set to one are terminated once they reach a head node.

3.4.4   DVSR Algorithm

Distributed Virtual-time Scheduling in Rings (DVSR) is a new dynamic bandwidth allocation algorithm introduced in 2004 by V. Gambiroza et al [6]. The concept is for nodes to compute a simple lower bound of temporally and spatially aggregated virtual time using per-ingress counters of byte arrivals without any complex operations such as fair queueing. As this information gets propagated along the ring, each node can approximate its fair rate for its own traffic.

This proposed algorithm addresses the following RPR standard limitations: severe and permanent oscillations, lack of achieving full spatial reuse and maximizing the ring utilization, and the large number of messages required for the fairness algorithm to converge.

In order to address these limitations, a reference model termed Ring Ingress Aggregated with Spatial reuse (RIAS) fairness is provided. RIAS fairness achieves maximum spatial reuse subject to providing fair rates to each ingress-aggregated flow at each link. DVSR uses measurements available at an RPR node to compute the minimum cumulative change in virtual time since the receipt of the last control message, as if the node was performing weighted fair queueing at the granularity of ingress-aggregated traffic. By distributing such control information upstream, nodes can perform simple operations on the collected information and throttle their ingress flows to their ring-wide RAIS fair rate.

The theoretical and simulation results provided demonstrate that DVSR provides an order of magnitude faster convergence than the RPR standard. In addition, DVSR mitigates the fairness algorithm oscillation while maximizing link utilization whereas the fairness algorithm suffers from utilization losses of up to 33% (for scenarios with unequal input rates). Some of the key RPR drawbacks that DVSR does not address are long ring access delays, dual ringlet utilization and end-to-end consistent CoS handling.

3.4.5 VSQ Access Control Scheme

H. Tang et al proposed in 2004 a new fair access control scheme called Virtual Source Queueing (VSQ) focused on metropolitan packet ring networks. It also proposes a simple feedback control scheme for VSQ to avoid packet loss on the rings [10].

Similar to DVSR, VSQ resolves the slow convergence and permanent oscillation in the RPR fairness algorithm. However, both RPR and DVSR have not addressed the long access delays for the node traffic due to the feedback rate control scheme. VSQ uses fair scheduling to provide guaranteed fair medium access and eliminates also the access delay problems.

The main idea of VSQ is to have a fair share of the bandwidth to all competing nodes based on the weighted source-based fairness definition in the RPR standard; where on any given segment on the ring, the available bandwidth is allocated to each node using the segment in proportion to its relative weight. VSQ provides a virtual queue Q_{ij} for traffic from each

possible source to its destination node on the ringlet, for a maximum of N-1 virtual queues per node, where N is the number of nodes on the ring and $1 \leq i, j \leq N$.

The key difference between the VSQ and RPR MAC architecture is that VSQ eliminates the need for: the fair rate estimation/feedback module, traffic monitor, rate controller and virtual-destination queues. The reason is that VSQ uses the fair scheduler to serve all the queues fairly so that every node can have guaranteed fair share of the ring bandwidth without fair rate estimation and feedback as needed by RPR.

VSQ introduces One-Hop-Backpressure which allows a very simple and effective feedback control as it checks the length of the queue upon reception and transmission of packets. On the other hand, the RPR feedback control provides hop-by-hop feedback to upstream nodes and is much more complex.

Another key advantage of VSQ is the isolation of the low priority traffic of different sources from each other. Hence, a greedy source that demands excessive bandwidth may suffer delay or loss while a well-behaved source has better access to its guaranteed fair share.

VSQ assumes two classes of traffic priorities: high and low. In real RPR networks, the three major CoS are used and often with their subclasses. Hence, limiting the analysis and simulation to a minimum subset does not address the SP RPR requirements for a deeper CoS and QoS requirements.

A similar research was introduced in 2006 by F. Alharbi et al [100]. This research introduced the use of virtual destination queues (VDQs) to avoid head-of-line blocking. Different bandwidth allocation policies were discussed to assign rates to VDQs. Also, a bandwidth allocation policy is proposed which was shown to achieve maximum utilization at low complexity.

3.4.6 Flow-based Fairness Algorithm

M. Chen et al proposed in 2006 a flow-based fairness algorithm. The analytical and simulation results indicated that the throughput of an RPR can reach up to twice the conventional rings in balanced uniform traffic pattern using node-based fairness. In addition, these results also demonstrated that the throughput of flow-based fairness will be approximately twice higher than the node-based fairness. Nonetheless, the complexity is higher with the flow-based algorithm [101].

The bandwidth utilization factor of the ring can be represented as follows:

$$\rho = \frac{1}{C} \sum_{i=0}^{N-1} S_{traffic,i} \qquad\qquad (3\text{-}22)$$

where $S_{traffic,i}$ is the total send traffic out of node i.

The node-based fairness utilization factor can be expressed as:

$$\rho = \frac{1}{C} \sum_{i=0}^{N-1} \frac{2C}{N} = 2 \qquad\qquad (3\text{-}23)$$

Also, the flow-based fairness utilization factor can be derived to:

$$\rho = \frac{4N}{N+2} \qquad\qquad (3\text{-}24)$$

This approaches 4 as N approach infinity.

Another key contribution by M. Chen in 2006 is the investigation of the RPR queue length, as the buffer size configuration is very important to the RPR performance. In this study, a relationship was developed between the throughput and number of nodes. The simulation and analytical results were carried out, and confirmed the accuracy. This study considered there classes of services: low, medium and high. It was shown that the queue length does not increase linearly with the number of nodes [102].

3.4.7 Additional RPR Fairness Improvement Research

C. Liu and J. Li proposed in 2004 another scheme to address the fairness algorithm drawbacks and reduce the convergence iterations beyond what DVSR has achieved for the case when the total input rate is less than the link capacity. This scheme provides fairness using estimated number of unbounded flows in each link [11]. The proposed fair-rate iteration algorithm is:

$$F_i(n+1) = \min\{C, C/K_i(n)\}$$

$$K_i(n) = \begin{cases} K_i(n-1) + \Delta N_i, if\,(\Delta N_i \neq 0) \\ [K_i(n-1) + N_i]/2, if\,(P_s = 0 \,\&\, T_t > T_u) \\ R_i(n)/F_i(n), otherwise \end{cases} \tag{3-25}$$

where $F_i(n)$ is the estimated fair rate for link i during the nth iterations, C is the link capacity, $K_i(n)$ is the estimated number of unbounded flows in link i during the nth iterations, ΔN_i is the change of active flow number in link i, P_s is the ratio of time in an iteration used by station traffic, T_t is the number of iterations, T_u is a threshold of iterations with $P_s = 0$. This proposal addresses the oscillation and convergence drawbacks of RPR and converges faster than DVSR algorithm.

Another proposal is called Low Complexity Fairness Algorithm (LCFA) proposed by F. Alharbi and N. Ansari in 2004 [12]. LCFA adopts the RIAS fairness concept, where the level of traffic granularity at a link is defined as an ingress-aggregated (IA) flow which is the aggregate of all flows originated from the same node but destined to different nodes. During congestion, all nodes should be able to send the same amount of data on the congested link relative to all other nodes. LCFA computes the advertised fair rate F_k in bytes by dividing the available bandwidth by the number of active nodes at node k, as follows:

$$F_k = \frac{C}{\max(1, \sum_{i=0, i \neq (k+1) \bmod N}^{N-1} \min(1, \frac{B_i}{F_k}))} \tag{3-26}$$

where C is the available bandwidth in bytes and is equal to the link capacity minus the reserved bandwidth for higher priority class traffic. Also, B_i is the IA traffic demand at node i, given by:

$$B_i = \sum_{j=(k+1)\bmod N}^{(i-1)\bmod N} b_{i,j} \qquad\qquad (3\text{-}27)$$

where $b_{i,j}$ is the flow originated from node i destined to node j. This scheme results in simple hardware design, fast convergence time and addresses the oscillation problem. However, it assumes that all nodes are treated equally during congestion and does not differentiate between the various levels of CoS required by SP's.

Another recent bandwidth allocation scheme, called Distributed Bandwidth Reallocated in Rings (DBRR), is introduced by P. Yue et al in [103]. It is based on approximate virtual time. It uses a simple FIFO scheduling scheme between transit and transmit buffers. Strict priority of transit traffic over local traffic is not considered in this research. In DBRR, each node uses its per-destination byte counters to construct a lower bound of the fair rate based on the evolution of the virtual time. Combining local calculated and remote information, the fair rate of per-destination ingress nodes can be adjusted dynamically. This algorithm reduces the fairness algorithm oscillations and convergence time, it also increases the throughput. The drawback of this algorithm is the ring access delays due to the lack of strict priority support.

Lastly, X. Zhou et al also presented an improvement to the RPR fairness algorithm in 2003. This proposal is based on the Node Ingress Aggregate (NIA) fairness in which the IA congested flows fairly share the bandwidth of the congested link that they pass through [104]. This proposal provides enhancements to RPR-AM fairness algorithm and enhances its performance.

3.4.8   Additional Ring Access Delay Improvement Research

D. Schupke and A. Riedl investigated packet transfer delay of 1-TB and 2-TB designs for low and high priority traffic in [13]. This research presented an analytical approximation

of the average packet delay. Also, 1-TB and 2-TB design delays were derived, presented and compared. However, only two classes of service: high and low priority traffic were considered. This work demonstrated that high priority traffic has higher delays with 1-TB design, whereas low priority traffic performs almost the same for both designs. In addition, it focused on the delay and utilization on a single ringlet with the minimum hop route assumption.

A follow on to this research by S. Zhou et al. modeled incoming traffic with three classes of services as opposed to only high and low priorities. This is obviously closer to the RPR CoS standard; nevertheless, it does not capture the subclass CoS support and only considered a single ringlet minimum hop routing. Also, this research demonstrated the relationship between the maximum throughput, delay and the number of nodes for a single ringlet [14].

Another contribution to the delay analysis and its steady state bound for RPR 1-TB design are analyzed in [15]. The work proposed a rate-based fairness algorithm vs. quota-based. This proposal demonstrated that this enhancement with 1-TB design significantly reduces access delays and allows RPR to be a scalable MAN/WAN solution.

A follow on research focused on simulating access delay for bursty traffic such as streaming video and interactive on-line gaming. This work considered both 1-TB and 2-TB designs. In order to minimize the access and end-to-end delay of Class B traffic, the shaper needs to be set to the committed information rate (CIR) and the excess information rate (EIR) instead of CIR alone [16].

Finally, in 2005, ring access delay bound was proven analytically under steady state. This approach demonstrated that the delay does not depend on the ring size; hence, it is a scalable solutions for the MAN/WAN applications [17].

3.4.9 Unaddressed RPR Drawbacks

The research surveyed in this section delivers innovative schemes to address the limitation of the current RPR fairness algorithm of severe oscillation and lengthy convergence time as well as long ring access delay, specifically drawbacks (a), (b) and (c). However, the limitations (d) and (e) remain unresolved. The limitations are critical to the success of RPR to deliver on the requirements for SP convergence in broadband networks. Bandwidth efficiency and service differentiation are fundamental elements to any SP. Hence, addressing these two limitations is a must for SP's to adopt RPR and make it a key building block in delivering convergence while leveraging the existing infrastructure. In the next two chapters, we propose two enhancements that address these two limitations.

3.5 Multi-Service RPR Model and Utilization

3.5.1 RPR Reference Model and Ringlet Selection

One of the key objectives of ring technologies is to ensure fair and high bandwidth utilization. Since RPR consists of dual counter-rotating rings, as shown in Figure 2-10, it is important to achieve this objective on both rings. IEEE 802.17 standard outlines the RPR layer model and its relationship to the open systems interconnect (OSI) reference model, as illustrated in Figure 3-6.
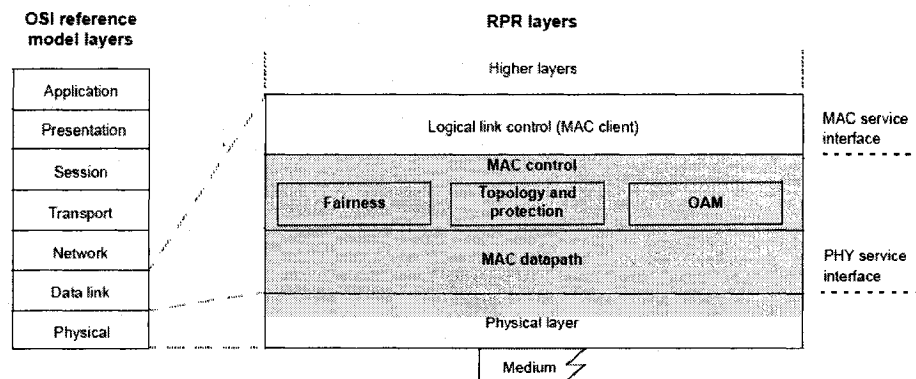

Figure 3-6: RPR Service and Reference Model [42]

The *MAC service interface* provides service primitives used by MAC clients to exchange data with other peer clients, or to transfer local control information between the MAC and

64

the MAC client. The *MAC control sublayer* controls the datapath sub-layer, maintains the MAC state and coordination with the MAC control sub-layer of other MACs, and controls the transfer of data between the MAC and its client. The control activities are distributed among stations on the ring in order to survive any single point of failure. Control entities play an essential role triggering the fairness algorithm and ringlet selection. The ringlet selection entity uses *destination_address*, *source_address*, and topology and protection database to determine: a) the ringlet to use to transmit the frame, b) whether and how to flood the frame, and c) type of RPR frame format to use. The *MAC datapath* sublayer provides data transfer functions for each ringlet. The *PHY service interface* is used by the MAC to transmit and receive frames on the physical media, shown in Figure 3-7 [42].
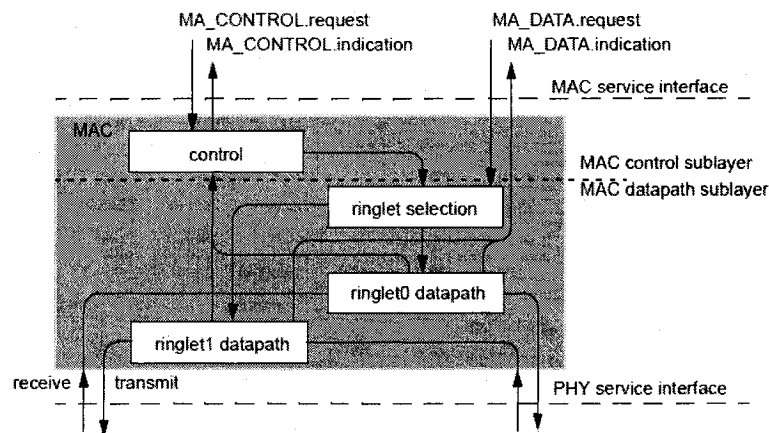
Figure 3-7: RPR MAC Architecture [42]

3.5.2 RPR Ringlet Utilization

Recent research conducted in studying and analyzing ring technologies especially RPR has focused on optimizing one ringlet based on per node traffic (vs. per flow on either ringlet). Ringlet selection control is typically used in the default mode. As a result, ringlet selection is done based on minimum-hop-count around the ring, and fairness is only invoked when congestion is detected (or threshold is reached). This option (or assumption) simplifies both the analysis and implementation for an RPR node. Nonetheless, it results in inefficient use of the total bandwidth on both ringlets, and an avoidable congestion management control lowering available allocated bandwidth.

For example, consider the RPR in Figure 3-8, and assume that each ringlet has 10Gbps link capacity. Let $f(cos, node\_i, node\_j)$ indicate the required bandwidth for a flow f with class of service cos starting at $node\_i$ and ending at $node\_j$. Given flows $f0(A0, 0, 3) = 5$Gbps, $f1(B\text{-}CIR, 1, 3) = 4$ Gbps and $f2(C, 2, 3) = 3$Gbps. All three flows will be on ringlet0, since it provides the least number of hops from node 0, 1 and 2 to node 3. However, the three flows require 12Gbps forcing congestion on the link between nodes 2 and 3. This results in invoking the fairness algorithm to share the bandwidth amongst these flows. In the meantime, ringlet1 has 10Gbps without utilization. Our research introduces a new method to enhance the existing standard to deliver higher bandwidth utilization and provide congestion avoidance; hence lowering the need to run the fairness algorithm. In the above example, also known as parking lot scenario, this can be accomplished by placing flows $f0$ and $f1$ on ringlet0 and $f2$ on ringlet1. We also carry out the performance analysis for packet delay to understand the impact on various traffic prioritization and ringlet selection.
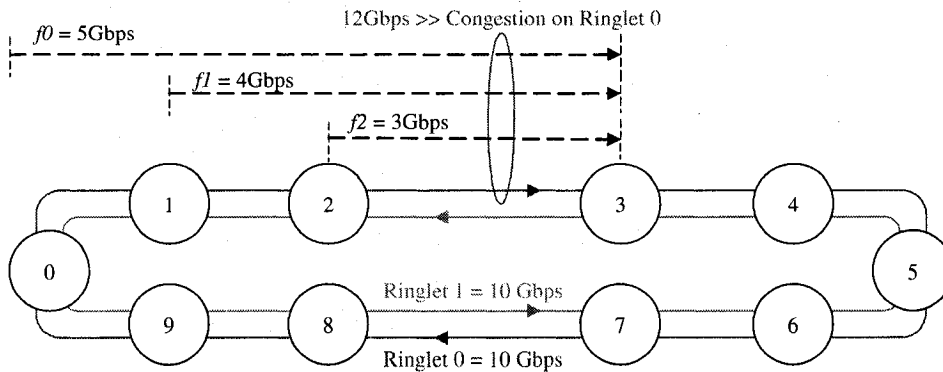
Figure 3-8: A 10-node RPR Example with Flows to Node 3

3.6 Summary

We classified IP services and its characteristics into four classes: Voice, Video, low-speed data and high-speed data. We outlined IP service sources according to their short and long-term dependence properties and models. For the SRD models, we covered both Markov and Regression models. As for the LRD models, we discussed Pareto process, FBM and superposition of high on-off sources.

Then, we summarized traffic modeling research that is applicable to IP traffic and services. First, we discussed how to model an MPLS node with MM CPP and model packet arrival and loss appropriately. Then, we discussed MMPP and how it can model an aggregate traffic of voice, video and data. Third, we presented a PMPP model and how it can model self-similar traffic. Finally, we showed the relevant work done on FBM to model LAN traffic.

Also, we discussed recent ring and RPR performance research covering BIR, ABA, DVSR, VSQ, VSD and other contributions to address the fairness algorithm drawbacks. We highlighted how recent research addressed the ring delay drawbacks. Lastly, we discussed how RPR enables multi-service support, and we highlighted how it is inefficient in optimizing the network utilization and does not maximize service fulfillment.

CHAPTER 4 – Flow-based Priority Distribution (RPR-FPD)

4.1 RPR Utilization and Service Fulfillment Enhancement Scheme

In CHAPTER 3, we presented the characteristics of IP services and the behavior at the packet level. We also surveyed the various traffic and source models applicable to IP services and IP/MPLS traffic. Then, we discussed traffic modeling and presented the key models researched. We summarized recent RPR research that focused on addressing current RPR drawbacks. Finally, we demonstrated one of the key limitations that recent research has not addressed, which is the consideration of the total bandwidth on both ringlets. In this chapter, we devise and analyze an algorithm which takes the total bandwidth available into account. Our algorithm results in enhancing RPR utilization and service fulfillment. In order to maximize the utilization on an RPR ring and avoid the scenarios described in the previous chapter, it is important that the RPR MAC Control sub-layer participates in ringlet selection on a per flow basis.

4.1.1 RPR-FPD Algorithm

Our proposed algorithm is called Flow-based Priority Distribution (FPD) algorithm. Figure 4-1 shows an N-node RPR ringlet, the alternate ringlet is counter-clockwise and not shown. A node i is connected to its neighbor using links i on one ringlet and link i-1 on the other ringlet.
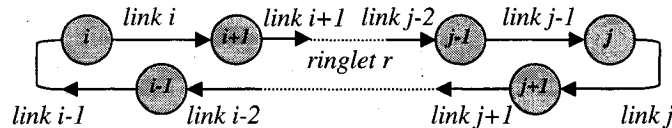

Figure 4-1: N-Node RPR Ringlet

Our proposed algorithm is composed of two major steps. In the first step, the node must establish the Lowest Cost Path (LCP) for every new flow. Then, the second step checks for Priority-based Congestion Prevention (PCP). RPR-FPD is shown in Figure 4-2. The LCP step starts by establishing the lowest cost ringlet (or path). We propose to use the

following parameters when ringlet selection mode is on: lowest path delay (link delays), if delay for both paths is equal, then use lowest number of hops; otherwise, use default ringlet. Based on these parameters, LCP is assigned.

Step 1: LCP

```
//Calculate each ringlet delay for flow_f(cos,i,j)
ringlet0_delay = Σ(x=i to j, ringlet0)(link_delay(x))
ringlet1_delay = Σ(x=i to, ringlet1j)(link_delay(x))
if ringlet0_delay < ringlet1_delay
    LCP = ringlet0
elseif ringlet0_delay > ringlet1_delay
    LCP = ringlet1
else
    //Calculate each ringlet number of hops for flow_f(cos,i,j)
    ringlet0_hops = Σ(x=i to j, ringlet0)
    ringlet1_hops = Σ(x=i to j, ringlet1)
    if ringlet0_hops < ringlet1_hops
        LCP = ringlet0
    elseif ringlet0_hops > ringlet1_hops
        LCP = ringlet1
    else
        LCP = default ringlet
if LCP = ringlet0
    alternate = ringlet1
else
    alternate = ringlet0
```

Step 2: PCP

```
//Check bandwidth for congestion and ensure that CoS requirement is satisfied
LCP_utilization = MAX(x=i to j, LCP)(link_bw_used(x)+flow_f)/link_capacity
alternate_utilization = MAX(x=i to j, alternate)(link_bw_used(x)+flow_f)/link_capacity
if LCP_utilization < congestion_threshold
    ringlet_selection flow_f, LCP
elseif (alternate_delay < flow_f cos_delay) AND
       (alternate_utilization < congestion_threshold)
    ringlet_selection flow_f, alternate
else
    //Congestion is unavoidable … fairness algorithm will be invoked
    ringlet_selection flow_f, LCP
    fairness LCP
```

Figure 4-2: Flow-based Priority Distribution (FPD) Algorithm

Then in step 2, PCP checks the bandwidth utilization of the LCP. If the LCP bandwidth and the requested flow bandwidth is less than the congestion threshold, the LCP ringlet is selected. If the alternative path bandwidth and the requested flow bandwidth is less than the congestion threshold, and the alternative path delay satisfies the CoS flow requirements, the alternative path ringlet is selected. Otherwise, congestion is unavoidable and the LCP ringlet is selected, which results in running the fairness algorithm.

4.1.2 Simulation Parameters

We define the following initial parameters to simulate an RPR network:

- N is number of RPR nodes on the ring
- C is the ringlet capacity in Mbps
- T_{link} is the link latency in μsec
- ρ_{Th} is the ringlet congestion threshold
- Δ defines length of a timeslot in msec
- Flow start-time, duration, end-time, source-node and destination-node
- Flow required CoS and bandwidth (Mbps)
- d_f is total flow link delay calculated in usec from source to destination node
- ρ_r is the utilization for ringlet r

Our simulation environment, shown in Figure 4-3 with initial flow demands, is composed of two steps: first, it takes the above parameters and applies them to an RPR 802.17 network, and then it applies the same parameters to an RPR-FPD network.

Each of these two steps is divided into equal timeslots of Δ msec. During each timeslot, we compute the following network parameters: ρ_{r0}, ρ_{r1}, the amount of bandwidth provided to each flow and the delay inhibited from the source to destination per flow. Each step status is detailed in the "Flow Demand and Flow/Link Allocation" box, and the flow bandwidth allocation is shown in the "Service Fulfillment" box.

We repeat this process to obtain statistical variability and track the number of iterations, as shown in the figure. We calculate the mean, standard deviation and obtain the confidence interval of the flow demands and fulfilled bandwidth results. The confidence interval is calculated based on the mean $\bar{x}$, variance σ^2, number of samples n, and significance level of 0.05 which yields a confidence level of 95%. Hence, the confidence interval is:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \qquad (4\text{-}1)$$

70

In the next section, we compare the results of the above steps for current RPR standard and our proposed FPD algorithm. Also, we discuss the advantages and disadvantages for each method.

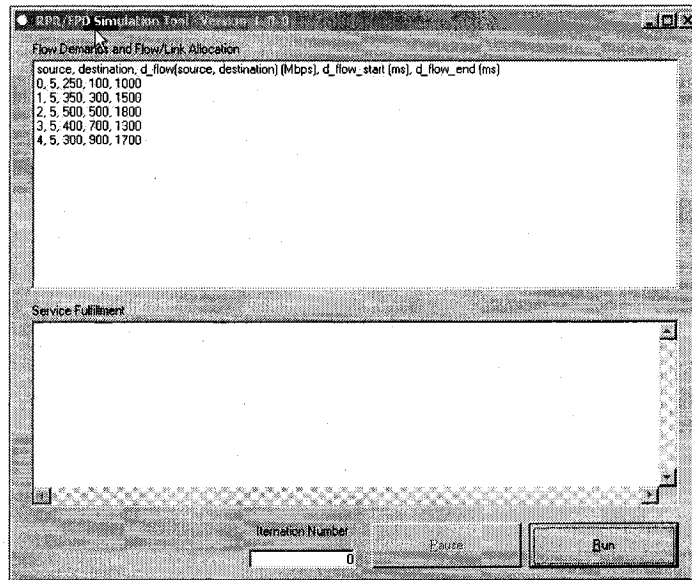Figure 4-3: RPR Network Simulation Tool

4.2 Results of RPR-FPD versus Current Methods

4.2.1 Ring & Network Initial Parameters

Consider an RPR network with $N = 10$ and $C = 1$ Gbps (1000 Mbps), as shown in Figure 4-4. In addition, we assume that $T_{link} = 2$ msec (2000 μsec) and $\rho_{Th} = 95\% \times C$. We also set Δ to 100 msec.
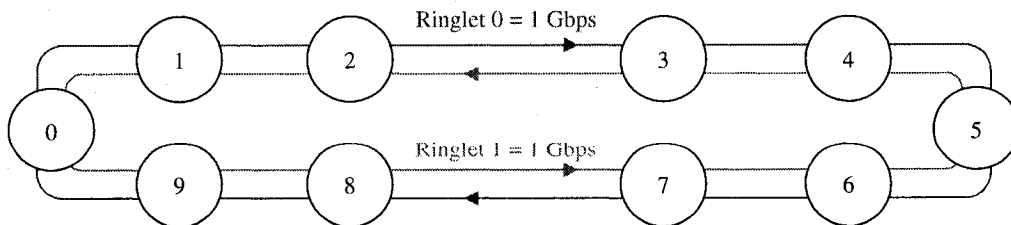

Figure 4-4: A 10-node RPR Network with 1 Gbps Ringlet

We consider five flows with a parking-slot scenario to node 5. The flow parameters are detailed in Table 4-1, Table 4-2and Figure 4-5. Note that the mean required bandwidth per flow is calculated based on 500 iterations.

Table 4-1: Flow 0 to Flow 4 Mean Required Bandwidth in Mbps

Time (ms)	flow 0	flow 1	flow 2	flow 3	flow 4
0	0	0	0	0	0
100	248	0	0	0	0
200	248	0	0	0	0
300	253	348	0	0	0
400	247	355	0	0	0
500	251	346	515	0	0
600	250	347	498	0	0
700	252	354	505	398	0
800	256	346	498	397	0
900	255	359	498	398	298
1000	250	342	500	391	300
1100	0	354	504	397	302
1200	0	351	499	404	297
1300	0	347	497	405	306
1400	0	348	492	0	289
1500	0	363	500	0	296
1600	0	0	499	0	300
1700	0	0	507	0	299
1800	0	0	491	0	0
1900	0	0	0	0	0
2000	0	0	0	0	0

Table 4-2: Flow 0 to Flow 4 Parameters and Maximum Confidence Interval

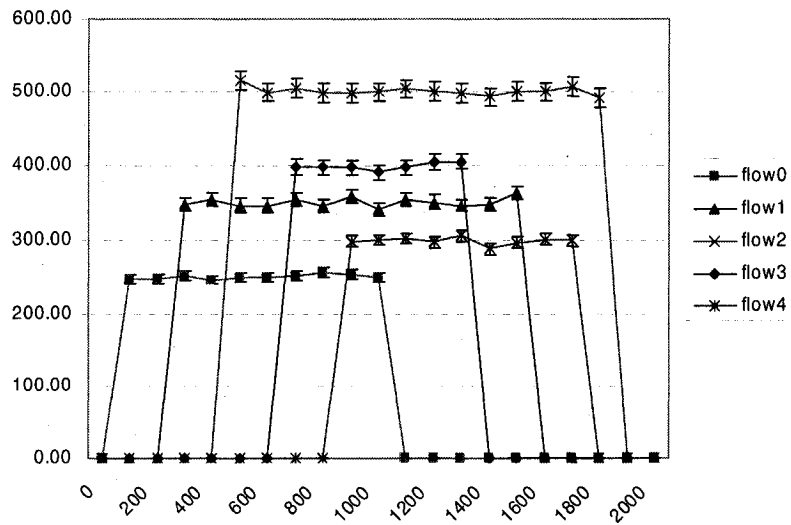| Flow Name	f(CoS, source, destination)	Max. Confidence Interval
flow 0	f0(A0, 0, 5)	6.94
flow 1	f1(A1, 1, 5)	9.16
flow 2	f2(B_CIR, 2, 5)	13.16
flow 3	f3(B_EIR, 3, 5)	10.19
flow 4	f4(C, 4, 5)	7.77

Figure 4-5: Flow 0 to Flow 4 Bandwidth Demands

4.2.2 RPR Simulation Results

Using 802.17, all flows are routed on ringlet 0 since it provides the shortest path to node 5. The RPR Simulation Tool output is illustrated in Figure 4-6. The flow propagation delay is listed in Table 4-3. The results of the simulation are detailed in Figure 4-7 and Table 4-4.
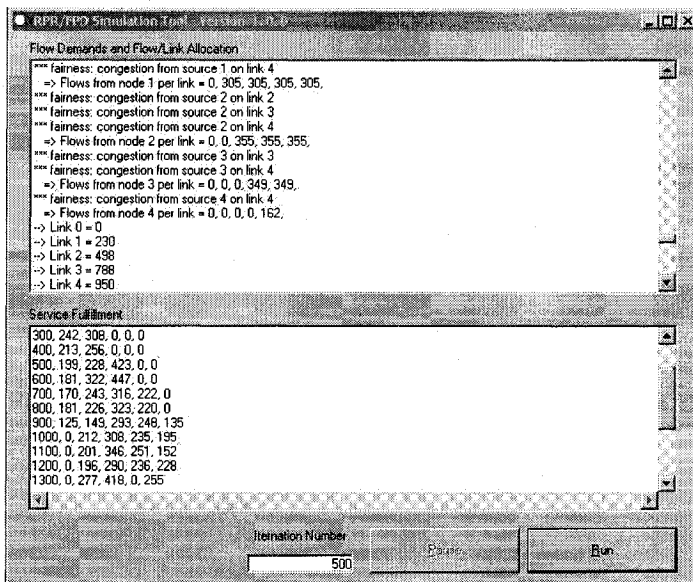
Figure 4-6: RPR Network Simulation Tool Results

Table 4-3: RPR Flow Propagation Delay & Fulfillment

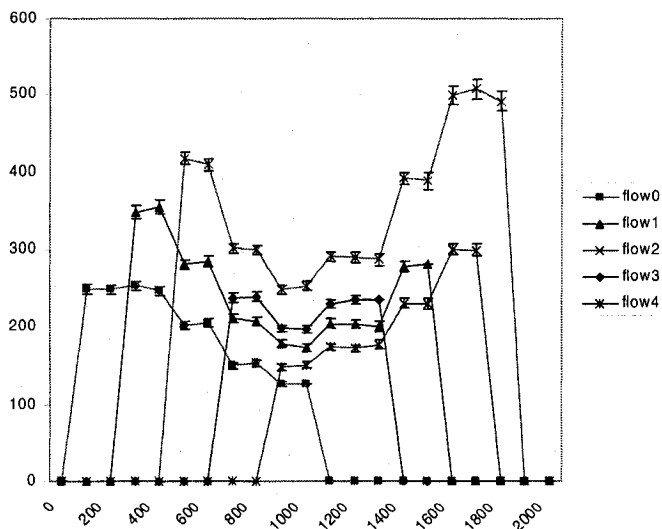| Flow Name	Flow 0	Flow 1	Flow 2	Flow 3	Flow 4
Propagation Delay	10 ms	8 ms	6 ms	4 ms	2 ms

Figure 4-7: RPR Flow 0 to Flow 4 Fulfillment

74

Table 4-4: Flow 0 to Flow 4 Mean RPR Fulfilled Bandwidth in Mbps

Time (ms)	flow 0	flow 1	flow 2	flow 3	flow 4
0	0	0	0	0	0
100	0	0	0	0	248
200	0	0	0	0	248
300	348	0	0	0	253
400	355	0	0	0	247
500	280	417	0	0	203
600	285	409	0	0	206
700	211	301	237	0	150
800	208	299	239	0	154
900	179	248	198	148	127
1000	172	252	197	152	126
1100	205	291	229	175	0
1200	204	290	234	172	0
1300	201	288	235	177	0
1400	277	392	0	230	0
1500	282	388	0	230	0
1600	0	499	0	300	0
1700	0	507	0	299	0
1800	0	491	0	0	0
1900	0	0	0	0	0
2000	0	0	0	0	0

We note that congestion is detected when $f2$ is initiated, and results in lower flow bandwidth fulfillment starting at $t = 0.5$ s. We track both bandwidth fulfillment and total ring utilization as shown in Figure 4-8.
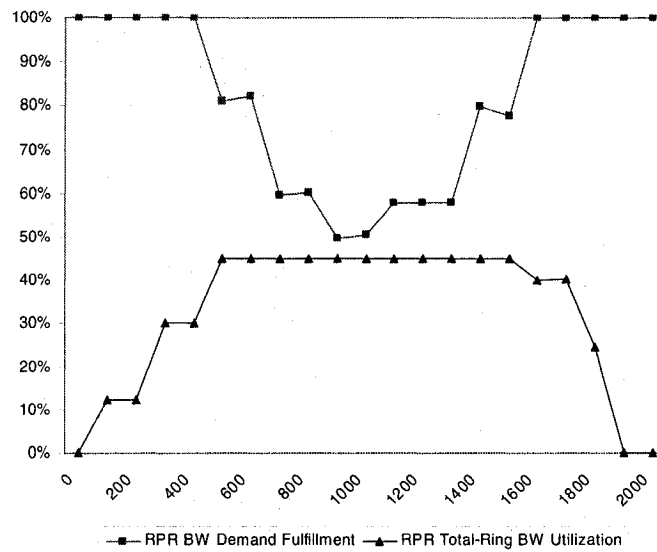
Figure 4-8: RPR Fulfillment and Utilization

4.2.3   RPR-FPD Simulation Results

Using our proposed RPR-FPD algorithm, we use the same network parameters to simulate the flow fulfillment, flow delay and ringlet utilization. FPD routes flows on both ringlets depending on algorithm parameters. The flow propagation delay is listed in Table 4-5. The results of the simulation are shown in Table 4-6 and Figure 4-9.

Table 4-5: RPR-FPD Flow Link Delay and Fulfillment

Flow Name	Flow 0	Flow 1	Flow 2	Flow 3	Flow 4
Propagation Delay	10 ms	8 ms	14 ms	16 ms	2 ms

76

Table 4-6: Flow 0 to Flow 4 Mean FPD Fulfilled Bandwidth in Mbps

Time (ms)	flow 0	flow 1	flow 2	flow 3	flow 4
0	0	0	0	0	0
100	248	0	0	0	0
200	248	0	0	0	0
300	253	348	0	0	0
400	247	355	0	0	0
500	251	346	515	0	0
600	250	347	498	0	0
700	252	354	503	397	0
800	256	346	498	397	0
900	252	354	498	398	294
1000	250	342	500	391	300
1100	0	354	503	397	302
1200	0	351	497	403	297
1300	0	347	496	404	306
1400	0	348	492	0	289
1500	0	363	500	0	296
1600	0	0	499	0	300
1700	0	0	507	0	299
1800	0	0	491	0	0
1900	0	0	0	0	0
2000	0	0	0	0	0

We note that RPR-FPD avoids congestion in this simulation and provides full bandwidth fulfillment by routing flows 2 and 3 to ringlet 1. However, this results in a delay penalty for these two flows as they take the longer path on the alternate ringlet. Figure 4-10 illustrates the bandwidth fulfillment and ring utilization.
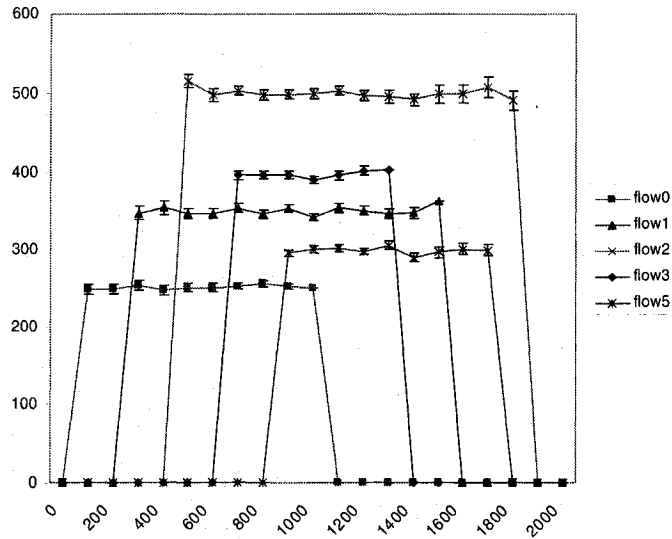
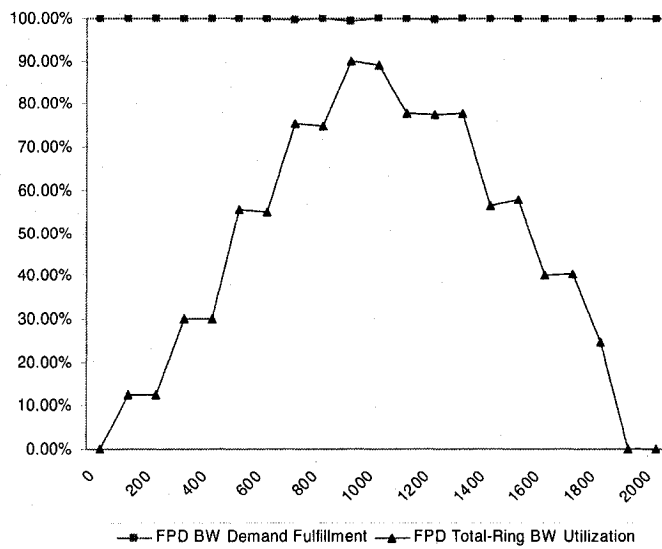Figure 4-9: RPR-FPD Flow 0 to Flow 4 Fulfillment

Figure 4-10: RPR-FPD Fulfillment and Utilization

4.2.4 Results Comparison

In our scenario above, RPR allocates all flows on a single ringlet since it used nodal distance and lowest cost path as a criterion for ringlet selection. As demonstrated in our simulation, this resulted in congestion when flow 2 was initiated on the ring, which in turn

lowers the service and flow fulfillment as the bandwidth is shared amongst the various active flows.

On the other hand, RPR-FPD distributed the flows on both ringlets based on LCP and PCP steps and parameters. In our specific case, congestion was completely avoided which results in providing 100% bandwidth fulfillment and higher total ring utilization.

It is important to note that this significant increase in fulfilling bandwidth demand and total ring bandwidth utilization comes at the expense of the packet delay. Consider flow 2 where RPR-FPD fulfills all of its demands on the alternate ringlet, but introduces delay of 3 hops with RPR vs. 7 hops with RPR-FPD. This is why it is critical to perform the PCP step in FPD; therefore, we ensure that the CoS and delay requirements are met. In order to make sure that we respect delay requirements, we derive and analyze RPR and FPD packet delay performance for 1-TB and 2-TB designs.

4.2.5   RPR-FPD Impact on Resiliency

One of the key benefits of RPR is its resiliency. During protection, RPR dual-ring bandwidth is reduced. As we demonstrated above, FPD enables higher utilization on both ringlets while respecting CoS requirements and avoiding congestion in certain scenarios. Once FPD is enabled and a link failure occurs, it is very likely that congestion will occur. Consequently, lower priority traffic will be impacted. We will consider the impact of FPD on resiliency in the following cases:

First, since 1-TB design does not require buffering more than a single packet, station ingress (client add) queues will have to buffer traffic during congestion. Hence, station ingress low-priority queues will suffer longer delays and potentially face dropping packets.

Second, 2-TB STQ is capable of buffering CoS A1 to C traffic. With higher utilization and under ring protection condition, STQ buffered packets will face longer delays due to Control and A0 CoS traffic in pass-through and station ingress queues. In this case, A1 and C CoS traffic admitted onto the RPR ring will face undesirable delays with a 2-TB design

due to the high utilization enabled by FPD and higher CoS getting priority access onto the protected ring.

Finally, given that FPD allows flows to use longer paths to increase utilization and avoid congestion. When a link failure occurs, and a flow using a longer path due to FPD is now switched to the shorter path, it is possible that a new transmitted packet on the shorter path can arrive before the last transmitted packet on the longer path. As shown in Figure 4-11, a 5-node RPR ring with 1Gbps per ringlet has two flows: $f1$ from node 0 to 1 and $f2$ from node 0 to 2. Figure 4-11 (a) demonstrates normal operation with FPD enabled for $f2$, utilization is maximized and congestion is avoided. When a link failure occurs between nodes 0 and 4, as illustrated in Figure 4-11 (b), $f2$ will be routed on the shorter path which results in congestion. However, node 4 might have received packets from node 0 prior to the failure, and they might be in transit to node 2. Hence, it is important for node 2 to ensure that any received packets from either ringlet are delivered in order and not out-of-sequence to the egress station queues.

(a) Normal Mode with FPD

$f1(0,1,A0)$ = 600Mbps on ringlet 0

$f2(0,2,A1)$ = 500Mbps on ringlet 1 (FPD)
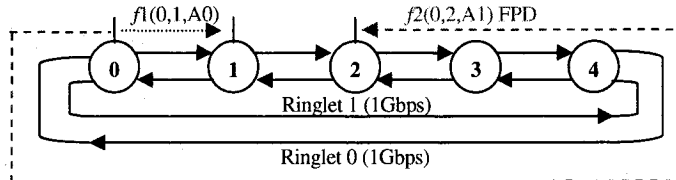
(b) Protection Mode with FPD

$f1(0,1,A0)$ = 600Mbps on ringlet 0 (congestion)

$f2(0,2,A1)$ = 500Mbps on ringlet 0 (congestion)
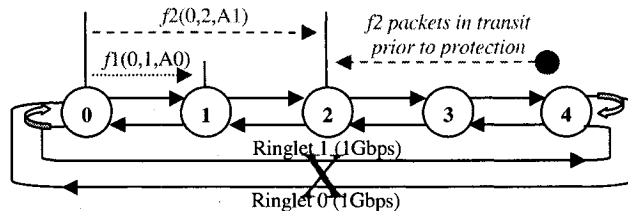
Node 2 must ensure that ringlet 1 in transit packets from node 4 to 2 are either delivered in sequence or discarded vs. packets arriving ringlet 0.

Figure 4-11: FPD Impact on RPR Resiliency

4.3 RPR-FPD Packet Delay Performance Analysis

4.3.1 RPR Node Model

We consider both 1-TB and 2-TB designs in our analysis. We model an RPR node as shown in Figure 4-12. The RPR standard specifies the classes of service (CoS) listed in Figure 2-9. In our research, we model all CoS queues supported by RPR, whereas previous RPR research only modeled 2 or 3 CoS. As a result, the control and subclass classes were not modeled appropriately. In our modeling, we assume the following:

- The number of RPR nodes on the ring is N, and an RPR node $\in \{0, 1, ..., N-1\}$

- Inter-arrival and services times are independent

- The latency per link (i.e. between two nodes) is constant $= T_{link}$

- Buffer capacities are infinite

- Total number of client traffic and control queues is P_T

- The probability of a node transmitting to node x traffic with priority P_i is $q_{x,Pi}$.

- Arrival process for each client CoS priority P_i transmission traffic is Poisson with mean arrival rate $\lambda_{client,Pi}$ for all nodes

- Packet length distributions are the same for all nodes, with packet sending times 1st and 2nd moments $\overline{S_{P_i}}, \overline{S_{P_i}^2}$, respectively.
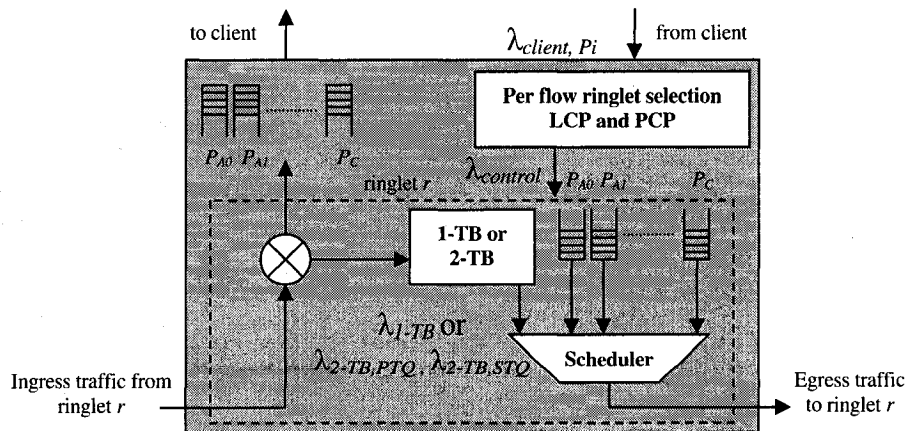

Figure 4-12: RPR-FPD Model

The packet arrival rates for the 1-TB (PTQ) is $\lambda_{1\text{-}TB}$, and 2-TB (PTQ, STQ) are ($\lambda_{2\text{-}TB,PTQ}$, $\lambda_{2\text{-}TB,STQ}$), respectively. Let α_{Pi} be the average number of TB's a packet traverses for a CoS P_i, from a given node to any other node on the ring:

$$\alpha_{P_i} = \sum_{x=2}^{N-1} (x-1) q_{x,P_i} \tag{4-2}$$

Now, the TB arrival rates and the client arrival rates can be related as follows:

$$\lambda_{1\text{-}TB} = \alpha_{P_{control}} \lambda_{P_{control}} + \sum_{i=2}^{P_T} \alpha_{P_i} \lambda_{client,P_i}$$

$$\lambda_{2\text{-}TB,PTQ} = \alpha_{P_{control}} \lambda_{P_{control}} + \alpha_{P_{A0}} \lambda_{client,P_{A0}} \tag{4-3}$$

$$\lambda_{2\text{-}TB,STQ} = \sum_{i=3}^{P_T} \alpha_{P_i} \lambda_{client,P_i}$$

4.3.2 CoS Packet Queueing Delay

This model can be regarded as head-of-line non-preemptive priority M/G/1 queueing system as derived in [105]. The general formula for waiting time of a packet with priority P_i:

$$W_i = \frac{R}{\left(1 - \sum_{j=1}^{i-1} \rho_j\right)\left(1 - \sum_{j=1}^{i} \rho_j\right)}$$

$$\text{where } R = \frac{1}{2} \sum_{k=1}^{P_T} \lambda_k \overline{S_k^2} \tag{4-4}$$

where R is the mean residual service time, and ρ_j is the throughput for priority queue j.

Note that the throughput per node for traffic priority P_j is $\dfrac{\rho_{P_j}}{N} = \lambda_{P_j} \overline{S_{P_j}}$.

Note that there is an approximation that the packets leaving node i are assumed to be arriving at node $i+1$ as a Poisson arrival process, in reality the Poisson property is not maintained after the service completion.

We start with the 1-TB design, the throughput parameters are:

$$\rho_1 = \rho_{1-TB} = \frac{1}{N}\left(\alpha_{P_{control}}\rho_{P_{control}} + \sum_{i=2}^{P_T}\alpha_{P_i}\rho_{client,P_i}\right)$$ (4-5)

$$\rho_2 = \rho_{control} = \frac{\rho_{P_{control}}}{N}$$ (4-6)

$$\rho_3 = \rho_{client,P_{A0}} = \frac{\rho_{P_{A0}}}{N}$$ (4-7)

$$\rho_4 = \rho_{client,P_{A1}} = \frac{\rho_{P_{A1}}}{N}$$ (4-8)

$$\rho_5 = \rho_{client,P_{B-CIR}} = \frac{\rho_{P_{B-CIR}}}{N}$$ (4-9)

$$\rho_6 = \rho_{client,P_{B-EIR}} = \frac{\rho_{P_{B-EIR}}}{N}$$ (4-10)

$$\rho_7 = \rho_{client,P_C} = \frac{\rho_{P_C}}{N}$$ (4-11)

The average packet queueing delay for CoS Pi is $T_{1-TB,Pi}$ can be represented as:

$$T_{1-TB,P_{Control}} = W_2 + \alpha_{P_{Control}}W_{1-TB}$$ (4-12)

$$T_{1-TB,P_{A0}} = W_3 + \alpha_{P_{A0}}W_{1-TB}$$ (4-13)

$$T_{1-TB,P_{A1}} = W_4 + \alpha_{P_{A1}}W_{1-TB}$$ (4-14)

$$T_{1-TB,P_{B-CIR}} = W_5 + \alpha_{P_{B-CIR}}W_{1-TB}$$ (4-15)

$$T_{1-TB,P_{B-EIR}} = W_6 + \alpha_{P_{B-EIR}}W_{1-TB}$$ (4-16)

$$T_{1-TB,P_C} = W_7 + \alpha_{P_C}W_{1-TB}$$ (4-17)

As for the 2-TB design, the throughput parameters are:

$$\rho_1 = \rho_{2-TB,PTQ} = \frac{1}{N}\left(\alpha_{P_{control}}\rho_{P_{control}} + \alpha_{P_{P_{A0}}}\rho_{client,P_{A0}}\right) \qquad (4\text{-}18)$$

$$\rho_2 = \rho_{control} = \frac{\rho_{P_{control}}}{N} \qquad (4\text{-}19)$$

$$\rho_3 = \rho_{client,P_{A0}} = \frac{\rho_{P_{A0}}}{N} \qquad (4\text{-}20)$$

$$\rho_4 = \rho_{client,P_{A1}} = \frac{\rho_{P_{A1}}}{N} \qquad (4\text{-}21)$$

$$\rho_5 = \rho_{client,P_{B-CIR}} = \frac{\rho_{P_{B-CIR}}}{N} \qquad (4\text{-}22)$$

$$\rho_6 = \rho_{client,P_{B-EIR}} = \frac{\rho_{P_{B-EIR}}}{N} \qquad (4\text{-}23)$$

$$\rho_7 = \rho_{client,P_C} = \frac{\rho_{P_C}}{N} \qquad (4\text{-}24)$$

$$\rho_8 = \rho_{2-TB,STQ} = \frac{1}{N}\left(\sum\nolimits_{P_i=P_{A1}}^{P_T} \alpha_{P_{P_i}}\rho_{client,Pi}\right) \qquad (4\text{-}25)$$

The average packet queueing delay for CoS Pi is $T_{2\text{-}TB,Pi}$ can be represented as:

$$T_{2-TB,P_{Control}} = W_2 + \alpha_{P_{Control}} W_{2-TB,PTQ} \qquad (4\text{-}26)$$

$$T_{2-TB,P_{A0}} = W_3 + \alpha_{P_{A0}} W_{2-TB,PTQ} \qquad (4\text{-}27)$$

$$T_{2-TB,P_{A1}} = W_4 + \alpha_{P_{A1}} W_{2-TB,STQ} \qquad (4\text{-}28)$$

$$T_{2-TB,P_{B-CIR}} = W_5 + \alpha_{P_{B-CIR}} W_{2-TB,STQ} \qquad (4\text{-}29)$$

$$T_{2-TB,P_{B-EIR}} = W_6 + \alpha_{P_{B-EIR}} W_{2-TB,STQ} \qquad (4\text{-}30)$$

$$T_{2-TB,P_C} = W_7 + \alpha_{P_C} W_{2-TB,STQ} \qquad (4\text{-}31)$$

4.3.3   Packet Queueing Delay Parameters

Based on the current RPR algorithm and the analysis, the probability of a node transmitting traffic priority P_i to node x is:

$$q_{x,P_i} = \begin{cases} \dfrac{2}{N-1}, & x = 1,2,\ldots,\dfrac{N-1}{2}, \; N \text{ is odd} \\[2ex] \dfrac{2}{N}, & x = 1,2,\ldots,\dfrac{N}{2}, \; N \text{ is even} \\[2ex] 0, & \text{otherwise} \end{cases} \qquad (4\text{-}32)$$

Therefore, α_{P_i} can be represented as:

$$\alpha_{P_i} = \begin{cases} \dfrac{N-3}{4}, & N \text{ is odd} \\[2ex] \dfrac{N-2}{4}, & N \text{ is even} \\[2ex] 0, & \textit{otherwise} \end{cases} \qquad (4\text{-}33)$$

Considering our proposed RPR-FPD algorithm, the probability that a node transmits traffic priority P_i to node x can be represented as follows:

$$q_{x,P_i} = \begin{cases} \dfrac{2}{N-1}, & x = 1,2,\ldots,\dfrac{N-1}{2}, \; N \text{ is odd and } C_{x,P_i}^{LCP} \\[2ex] \dfrac{2}{N}, & x = 1,2,\ldots,\dfrac{N}{2}, \; N \text{ is even and } C_{x,P_i}^{LCP} \\[2ex] \dfrac{1}{N-1}, & x = 1,2,\ldots,N-1 \text{ and } C_{x,P_i}^{alternate} \\[2ex] 0, & \textit{otherwise} \end{cases} \qquad (4\text{-}34)$$

where $C_{x,P_i}^{ringlet}$ is the ringlet condition that traffic priority P_i destined to node x.

Given $d_{Pi,Th}$ denotes the delay threshold for traffic priority P_i, ρ_{Th} as the ringlet throughput threshold, also $\land$ and $\lor$ are the AND and OR operators, respectively, we can represent $C_{x,P_i}^{ringlet}$ as:

$$C_{x,P_i}^{LCP} = \left(\rho_{x,P_i}^{LCP} < \rho_{Th}\right) \vee \left\{\left(\rho_{x,P_i}^{LCP} \geq \rho_{Th}\right) \wedge \left[\left(\rho_{x,P_i}^{alternate} \geq \rho_{Th}\right) \vee \left(d_{x,P_i}^{alternate} \geq d_{P_i,Th}\right)\right]\right\}$$

$$C_{x,P_i}^{alternate} = \left(\rho_{x,P_i}^{LCP} \geq \rho_{Th}\right) \wedge \left(\rho_{x,P_i}^{alternate} < \rho_{Th}\right) \wedge \left(d_{x,P_i}^{alternate} < d_{P_i,Th}\right) \qquad (4\text{-}35)$$

Therefore, we can derive α_{P_i} as:

$$\alpha_{P_i} = \begin{cases} \dfrac{N-3}{4}, & N \text{ is odd and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{N-2}{4}, & N \text{ is even and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{N-2}{2}, & C_{x,P_i}^{alternate} \\[2mm] 0, & otherwise \end{cases} \qquad (4\text{-}36)$$

4.3.4   End-to-End Packet Queueing Delay on RPR

We now consider the end-to-end delay for a packet as it enters and exits an RPR ring. We first consider 1-TB design; we can represent the average end-to-end delay as follows:

$$T_{E2E,1\text{-}TB,P_i} = \sum_{j=1}^{N-1} (W_{P_i} + (j-1)W_{1\text{-}TB})q_{j,P_i} \qquad (4\text{-}37)$$

where P_i represents CoS Control, A0, A1, B-CIR, B-EIR or C.

As for the 2-TB design, the end-to-end packet delay is expressed as,

$$T_{E2E,2\text{-}TB,P_{control}} = \sum_{j=1}^{N-1} (W_{P_{control}} + (j-1)W_{2\text{-}TB,PTQ})q_{j,P_{control}} \qquad (4\text{-}38)$$

$$T_{E2E,2\text{-}TB,P_{A0}} = \sum_{j=1}^{N-1} (W_{P_{A0}} + (j-1)W_{2\text{-}TB,PTQ})q_{j,P_{A0}} \qquad (4\text{-}39)$$

$$T_{E2E,2\text{-}TB,P_i} = \sum_{j=1}^{N-1} (W_{P_i} + (j-1)W_{2\text{-}TB,STQ})q_{j,P_i} \qquad (4\text{-}40)$$

where P_i represents CoS A1, B-CIR, B-EIR or C.

It is important to note that the above equations are applicable to both RPR and FPD algorithms.

4.4 RPR-FPD Packet Delay Results

4.4.1 Packet Delay Analysis (PDA) Tool & Parameters

We developed a tool to compute the RPR packet delay for current RPR and RPR-FPD based networks. Our Packet Delay Analysis (PDA) tool provides 1-TB and 2-TB delay for each case. In addition, our tool has the option to enable RPR-FPD for each CoS.

Using our PDA tool, we compare the results of our packet delay analysis for RPR and RPR-FPD based networks. Our analysis includes both designs 1-TB and 2-TB for all RPR supported CoS traffic: control, A0, A1, B-CIR, B-EIR and C, whereas previous research focused on two or three CoS. We also consider the impact of various RPR parameters on the packet delay. Our PDA tool is shown in Figure 4-13. This figure illustrates the input parameters which can be set for each scenario considered: number of RPR nodes, ring capacity (Mbps), each CoS average arrival rate (Mbps) and each CoS average packet length (Bytes). Our PDA tool also provides an option to increment each parameter in order to study its impact on the network packet delay.

We assume that packet length has exponential distribution; hence, packet sending time is exponentially distributed. Let $\overline{S_{P_i}} = \frac{1}{\mu_{P_i}}$ and $\overline{S_{P_i}^2} = \frac{2}{\mu_{P_i}^2}$. The results of our tool, for the given parameters, are shown in Figure 4-13, with the scenario to increase the number of nodes from 10 to 50 nodes in increments of 10. The tool computes the packet delay for each scenario using our analysis from the previous section.
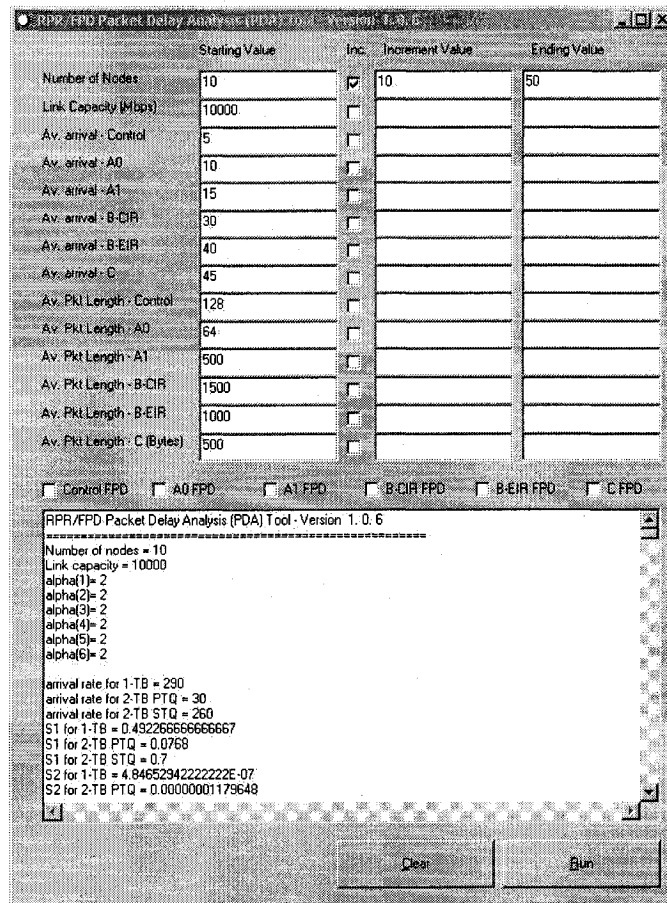
Figure 4-13: Packet Delay Analysis (PDA) Tool

4.4.2 Number of Nodes Impact on Packet Delay

We consider the input parameters in Figure 4-13 and increase the number of nodes from 10 to 50 in increments of 10, as illustrated in the figure. Using our PDA tool, we compute the packet delay for each CoS and consider the following cases: 1-TB with RPR and 1-TB with FPD as illustrated in Figure 4-14, 2-TB with RPR and 2-TB with FPD as in Figure 4-15.
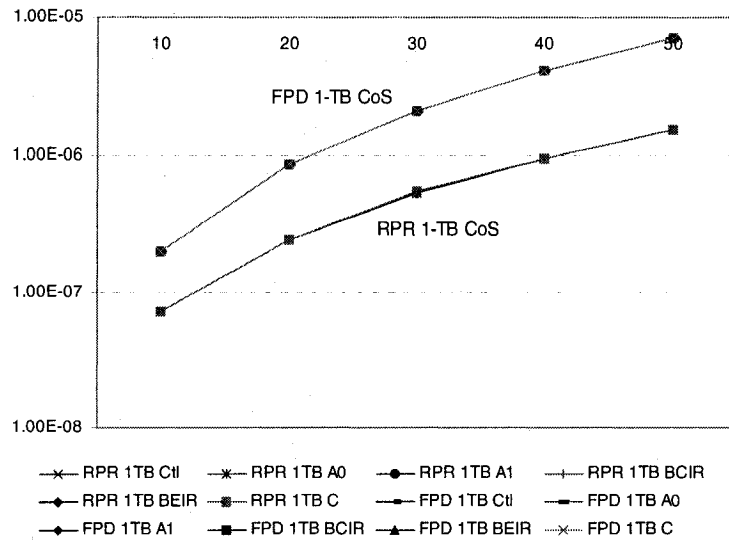
Figure 4-14: Queueing Delay vs. No. of Nodes for a 1-TB RPR and FPD

Our analysis demonstrates that in the case of 1-TB design the packet delay increases as the number of nodes increase. Also, the 1-TB RPR packet delay is lower than the 1-TB FPD due to the longer route utilized, which results in higher ringlet utilization with FPD. We also note that in the 1-TB design, all CoS's follow closely in packet delay and there is no differentiation between the various classes, as they appear as a single line on the graph.
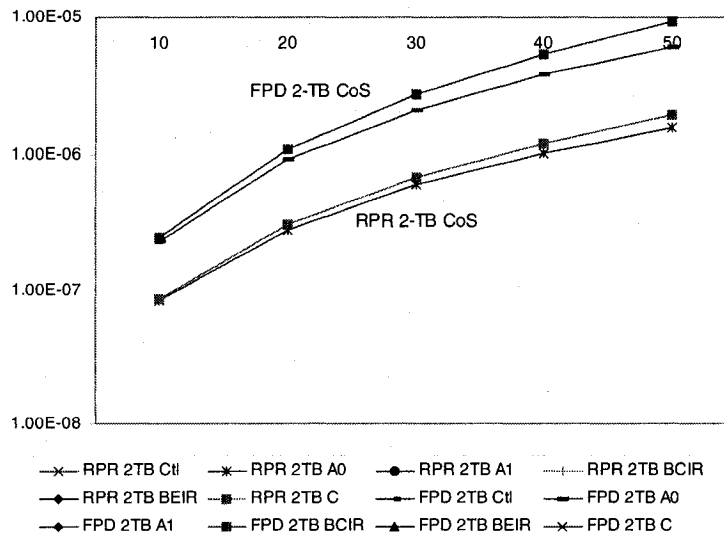

Figure 4-15: Queueing Delay vs. No. of Nodes for a 2-TB RPR and FPD

89

Similarly, with the 2-TB designs, the packet delay is higher as we increase the number of nodes on the ring. Also, FPD provides higher delay due to the utilization of the alternate routes. However, we notice that with 2-TB design, there is a clear differentiation between two categories of services due to the two queueing mechanism that this design supports. This differentiation becomes more pronounced as we increase the number of nodes; hence, increasing the ringlet utilization. The analysis demonstrates that 2-TB design provides lower delays for PTQ traffic, and performs as well as for traffic in the STQ. The relative performance of 1-TB and 2-TB is the same. However, we notice a higher delay impact on all traffic in the case of 1-TB and 2-TB STQ traffic, which is contributed to the higher number of TB's (on the alternate/longer ringlet path).

4.4.3   Ringlet Capacity Impact on Packet Delay

Similarly, we analyze the impact of ringlet capacity on the various CoS, TB designs and RPR algorithms. We start with the 1-TB design with RPR and 1-TB design with FPD as in Figure 4-16. Then, we analyze 2-TB design with RPR and 2-TB design with FPD as shown in Figure 4-17.
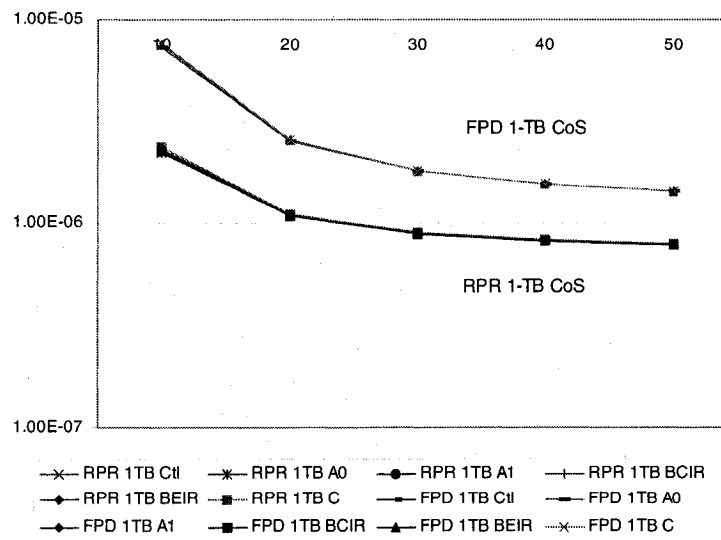

Figure 4-16: Queueing Delay vs. Ringlet Capacity for a 1-TB RPR and FPD

The results confirm the throughput impact on packet queueing delay. Also, we expected higher delay when using FPD which provides higher demand fulfillment and throughput in exchange for higher delay. Since these two cases cover the 1-TB design, the packet delays for all CoS's track closely, which means that there is little service differentiation during congestion time.
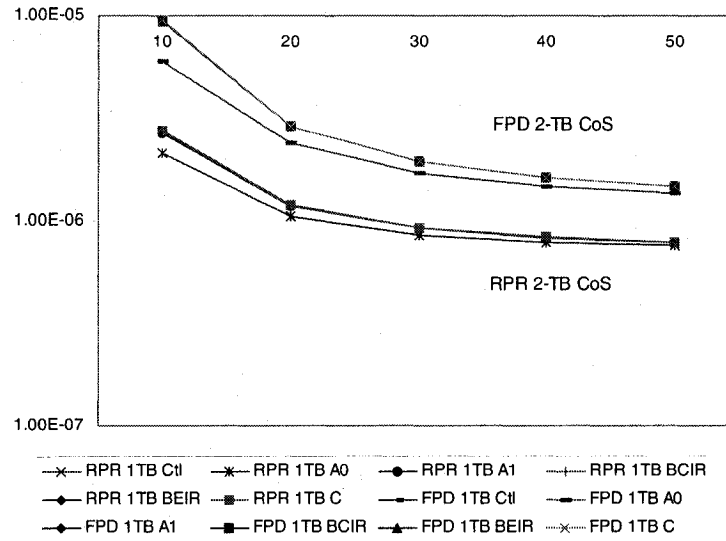

Figure 4-17: Queueing Delay vs. Ringlet Capacity for a 2-TB RPR and FPD

Similarly, with the 2-TB designs, the packet delay decreases as we increase the ringlet capacity. FPD provides higher delays as it uses the alternate ringlet. Also, we notice that with the 2-TB design, there are two classes of service due to the two queueing mechanisms namely, PTQ and STQ, whereas the 1-TB design has only PTQ with little differentiation between the CoS's.

4.4.4 Average Arrival Rate Impact on Packet Delay

Let us now consider the impact of the average arrival rate on the packet queueing delay. We consider the same parameters as in Figure 4-13, while incrementing each CoS average arrival rate by 200 Mbps. In Figure 4-18, we demonstrate the impact of average arrival rate for each CoS on packet delay with 1-TB design on an RPR ring. The results indicate that the packet delay increases as we increase the average arrival rate. However, this is also

dependent on the CoS average packet length. For example, we notice that the packet delay with B-CIR is higher than that of class C for the same average arrival rate. This is due to the fact that the average packet length of B-CIR is three times larger than that of class C. Moreover, we notice that increase in a certain CoS average arrival rate results in noticeable increase in its packet delay as well as all lower priority CoS's.
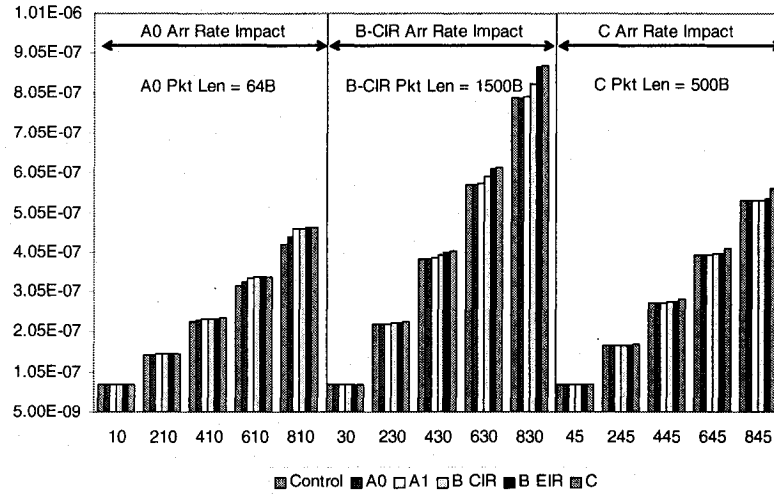

Figure 4-18: Queueing Delay vs. Average Arrival Rate with 1-TB RPR

We consider the 2-TB design and vary the same average arrival rates as before to study the impact on packet delay, as shown in Figure 4-19. The interesting result with the 2-TB design is that an increase in the PTQ CoS's such as A0 has lower delay than that of 1-TB, whereas any increase in the STQ CoS's such as B-CIR or C has a higher delay than 1-TB respective increase. This is attributed to the priority queueing mechanism in the 2-TB design.
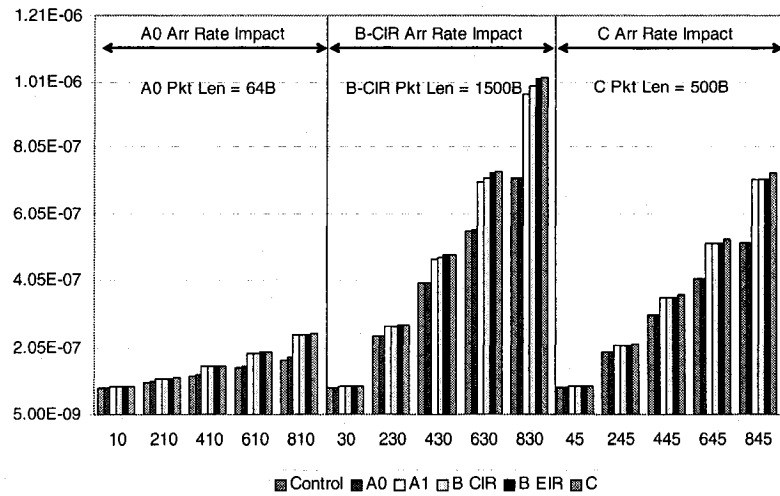
Figure 4-19: Queueing Delay vs. Average Arrival Rate with 2-TB RPR

Finally, we notice that the 2-TB design clearly differentiates between two distinct classes of service, whereas 1-TB provides little differentiation.

4.4.5 Average Packet Length Impact on Packet Delay

Finally, we consider the last set of input parameters; namely, average packet length for each of the CoS's. Using the same data as in Figure 4-13, with the exception that we set all average packet lengths to 100 bytes, we increment each of the packet length variables by 200 bytes until 1300 bytes average packet size is reached. As shown in Figure 4-20, we consider the impact of varying average packet length for three key classes: A0, B-CIR and C. We notice that the variation on the packet length with constant arrival rate has minimal impact on the delay. It also provides slightly lower packet delay as we increase lower priority packet length. In other words, as we increase packet lengths of the highest priority traffic, the overall delay is increased due to the accumulation of scheduling and servicing.
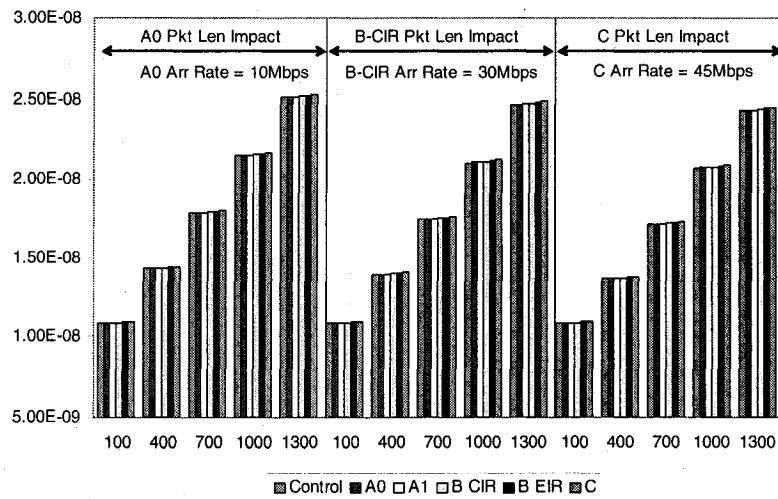
Figure 4-20: Queueing Delay vs. Average Packet Length with 1-TB RPR

We also consider the above packet length variability with 2-TB design, as illustrated in Figure 4-21. Similarly, packet delay increases as we increase packet length. However, an important observation is the impact of A0 packet length variability on all CoS's, whereas the impact of STQ CoS's packet length variability results in a higher increase for the STQ CoS's packet delay. Also, the differentiation of two categories of services is obvious in the 2-TB vs. the 1-TB case. Moreover, the packet delay in the case of 2-TB (vs. 1-TB) is overall higher when STQ variability exists, and lower when PTQ variability is considered.
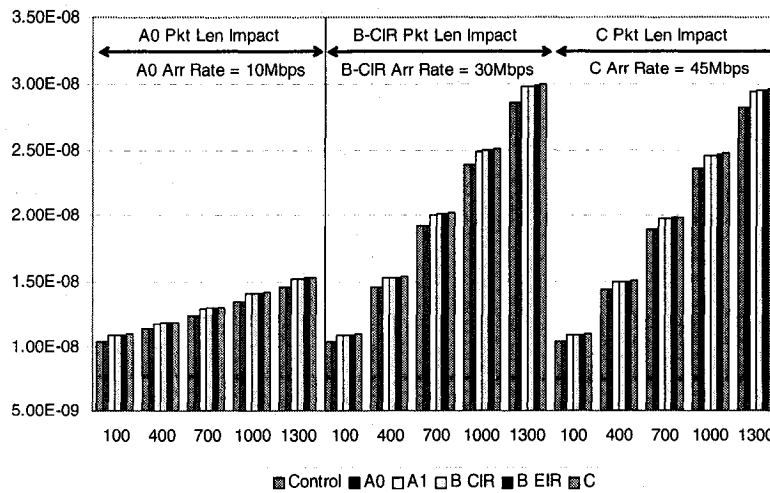

Figure 4-21: Queueing Delay vs. Average Packet Length with 2-TB RPR

In summary, we have demonstrated the impact on packet queueing delay for all CoS's with variability on all input parameters considered. The 2-TB demonstrated better differentiation for the PTQ CoS's; hence, better service fulfillment. However, the 1-TB provides the lowest packet delay across all CoS's when variability is in the lower CoS's. In addition, we showed that FPD results in higher packet delays in all cases against the current methods, which comes at the expense of the higher ring utilization provided by FPD.

4.5 Summary

We introduced our new algorithm Flow-based Priority Distribution (FPD) in this chapter. The algorithm and simulation environment were described. We demonstrated through our results the advantage of FPD over the current RPR standard for both 1-TB and 2-TB designs, specifically in maximizing ring utilization and increasing service fulfillment. Nevertheless, this increase was shown to be at the expense of packet delay.

We derived the packet delay for 1-TB and 2-TB designs, and we analyzed the results for RPR and FPD with both designs considering various input parameters such as: number of nodes on a ring, ring capacity, average arrival rate per CoS and average packet length per CoS. It is important to note that we considered all CoS's supported by the standard in our algorithm, simulation and analysis.

CHAPTER 5 - RPR 3-TB Design & Performance Analysis

5.1 End-to-End CoS Consistency

As discussed in CHAPTER 2, RPR supports three main CoS's, which makes it suitable for Service Provider deployments for current and next generation services. However, the current RPR standard supports two types of transit buffer designs: 1-TB and 2-TB. As a result, the three main CoS's will need to be shared with such designs, which will force at least one class to be compromised across the ring with lower CoS's.

5.1.1 1-TB CoS Differentiation and Drawbacks

The 1-TB system is shown in Figure 5-1, where incoming ringlet traffic gets dropped locally to the appropriate CoS queue or passed-through to the 1-TB PTQ. This queue has the highest priority (P1) access to the outgoing ringlet. The local add control and client traffic has lower access priority per CoS: Control, A0, A1, B-CIR, B-EIR and C, with priorities P2 to P7, respectively.

Consider an incoming packet with CoS C from the ringlet, where it is destined to pass-through this RPR node with 1-TB design. This packet will have a higher priority access to the ringlet than any other add traffic. Now, assume that this RPR node has Control traffic to add to the outgoing ringlet. Even though the add traffic CoS is higher than the transit traffic, in 1-TB design, the pass-through traffic will always have higher priority. As a result, the Control add traffic will be queued until the 1-TB PTQ is fully served; hence, adding undesirable delay to this CoS. It is important to note that with the 1-TB design, PTQ has maximum one packet in its queue. Hence, the maximum delay is the time required to transmit a packet with the maximum length.

As this pass-through packet traverses multiple 1-TB's, the impact on downstream nodes adding higher CoS traffic becomes higher. Hence, all nodes that pass-through this packet will not be able to add any traffic until their PTQ queues are served or the fairness algorithm is invoked. In both cases, the higher priority CoS will not receive its appropriate

96

CoS handling. In exchange, the 1-TB design is the simplest and requires the smallest queue length of a single packet. As a result, 1-TB provides a simple queueing design with minimal and upper-bound packet pass-through delay; however, it results in higher delays for add traffic, both Control and client.
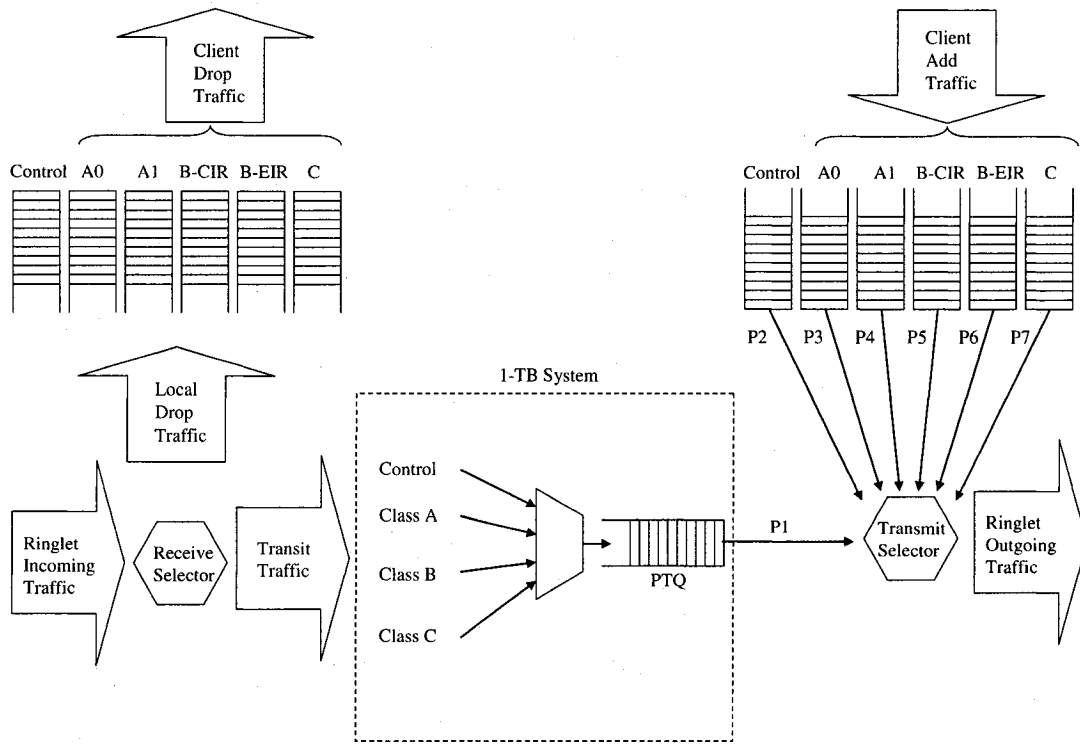

Figure 5-1: RPR Node 1-TB Design and CoS Scheduling

5.1.2   2-TB CoS Differentiation and Drawbacks

Similarly, the 2-TB system is provided in Figure 5-2, where incoming ringlet traffic gets either dropped locally to the appropriate CoS queue or passed-through to the 2-TB system. The 2-TB system has PTQ and STQ. PTQ serves the Control and A0 CoS's, whereas STQ serves all remaining CoS's. PTQ has the highest priority (P1) access to the outgoing ringlet, while STQ has the lowest priority (P8). The local add control and client traffic has access priority P2 to P7 for Control, A0, A1, B-CIR, B-EIR and C, respectively.

In the case of 2-TB, STQ provides some CoS differentiation between lower priority pass-through traffic and higher priority add Control of A0 traffic. However, the number of CoS's merged in STQ is too high resulting in insufficient service differentiation. As a

97

result, classes A1, B and C have the same level of service. Consider a C CoS client add packet which has P7 priority, and let us assume that there is a pass-through packet in STQ with CoS A1, which has access priority P8. The 2-TB system will provide access to the C add packet over A1 pass-through packet.
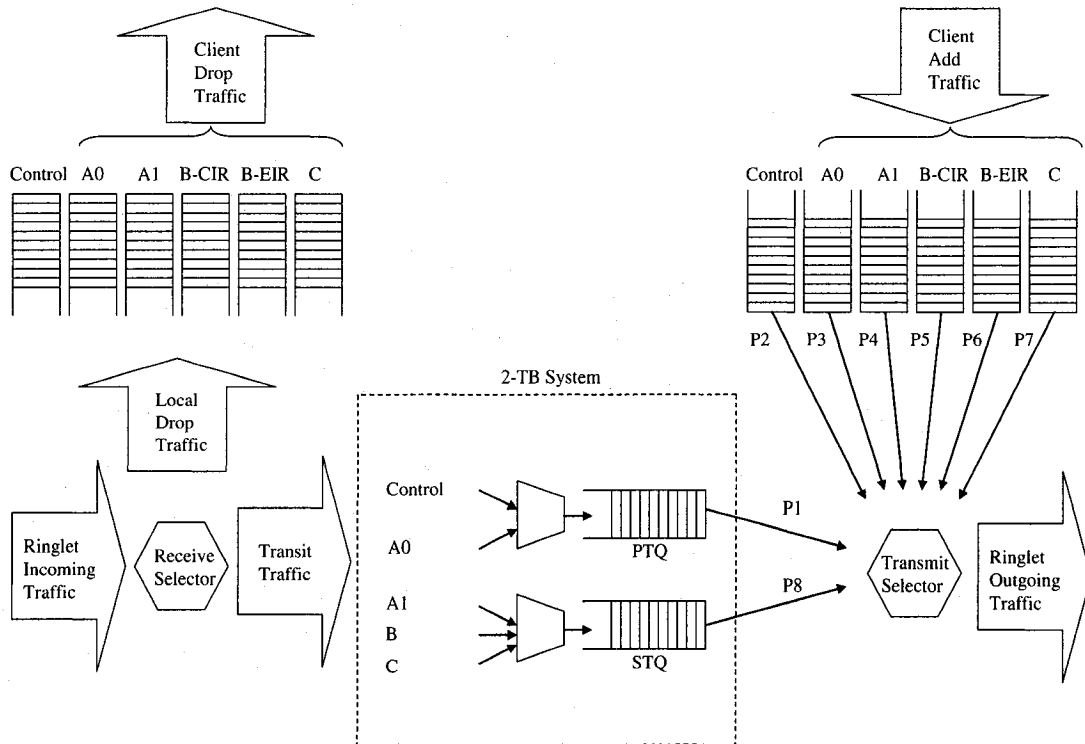
Figure 5-2: RPR Node 2-TB Design and CoS Scheduling

At the network level, higher CoS pass-through traffic, especially A1 and B-CIR, will experience undesirable delays in order to serve lower CoS's B-EIR and C. For example, a node adding CoS C will severely impact A1 CoS traffic from all upstream nodes. Therefore, the 2-TB design provides better CoS differentiation than 1-TB, but it is still insufficient for CoS A1 and B-CIR. However, the 2-TB has a more expensive queueing system requiring larger buffers for STQ than the simple single queue 1-TB design.

5.2 Proposed 3-TB Design

5.2.1 3-TB Design Overview

In order to address the short-comings of the 1-TB and 2-TB and provide consistent CoS differentiation across RPR, we propose a new TB design, called 3-TB design as illustrated in Figure 5-3.
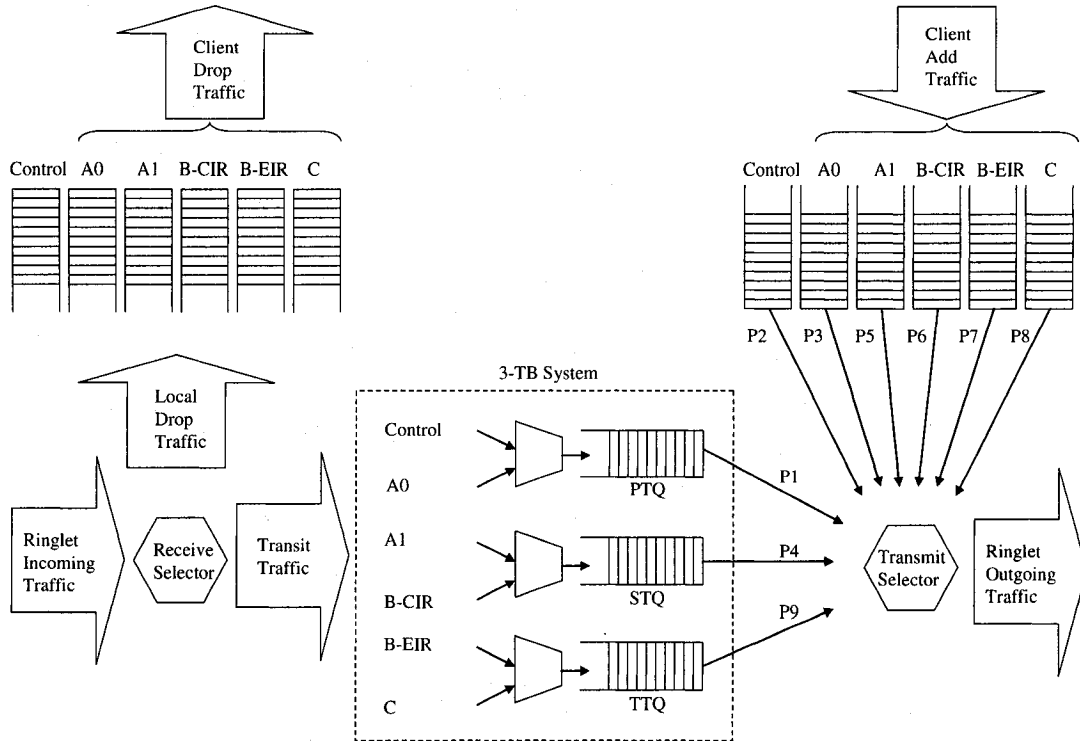
Figure 5-3: Proposed RPR Node 3-TB Design and CoS Scheduling

Our new approach operates in a similar fashion to the 1-TB and 2-TB designs. This is an important aspect as it allows implementation portability and compatibility. As demonstrated in Figure 5-3, incoming ringlet traffic gets either dropped locally to the appropriate CoS queue or passed-through to the 3-TB system. The 3-TB system has three queues: PTQ, STQ and Tertiary Transit Queue (TTQ). PTQ serves the Control and A0 CoS's, whereas STQ serves A1 and B-CIR CoS's, while TTQ serves B-EIR and C CoS's. PTQ has the highest priority (P1) access to the outgoing ringlet, followed by local-add control traffic as P2 and local-add client A0 traffic as P3. Then, STQ has priority P4, A1 as

P5, B-CIR as P6, B-EIR as P7 and C as P8. Finally, TTQ is given the lowest priority with P9.

5.2.2 3-TB Design Operation

As packets arrive from the incoming ringlet, the receiver selector decides whether it is locally destined or not. Consequently, locally destined packets are dropped in their respective queue based on their CoS, whereas transit packets are passed to the 3-TB system. The 3-TB system then classifies all transit packets into three queues based on the scheme described above. Finally, the transmit selector uses this scheme to schedule the outgoing packet for transmission on the ringlet.

Our three-class approach provides a more efficient and consistent, but not an ideal scheme for delivering end-to-end services. The ideal case would be to consider a transit buffer with as many queues as there are CoS's plus one for the control traffic, which would result in a very expensive RPR-node implementation. Hence, providing as many TB's as the main CoS's is the right balance which provides an efficient CoS handling. In addition, having PTQ serviced by the transmit selector at a higher priority than local add Control and A0 traffic provides a priority to admitted and transit RPR traffic, similarly with STQ over A1 and B-CIR. Finally, we propose to provide higher priority to locally added class B-EIR and C traffic than TTQ transit traffic so that locally-added traffic does not wait for undesired periods, as it is the least time-sensitive traffic.

We also propose to allow the use of our FPD algorithm on a per class basis as provisioned or automatically determined based on the services and its associated CoS requirements as well as the network load and status.

5.3 3-TB Packet Delay Analysis

5.3.1 3-TB RPR Node Model

In our proposed 3-TB model, we assume the following:

- The number of RPR nodes on the ring is N, and an RPR node ϵ $\{0, 1, ..., N-1\}$

100

- Inter-arrival and services times are independent

- The latency per link (i.e. between two nodes) is constant $= T_{link}$

- Buffer capacities are infinite

- Total number of client traffic and control queues is P_T

- The probability of a node transmitting to node x traffic with priority P_i is $q_{x,Pi}$.

- Arrival process for each client CoS priority P_i transmission traffic is Poisson with mean arrival rate $\lambda_{client,Pi}$ for all nodes

- Packet length distributions are the same for all nodes, with packet sending times 1st and 2nd moments $\overline{S_{P_i}}, \overline{S_{P_i}^2}$, respectively.

Figure 5-4: RPR 3-TB Model

The packet arrival rates for the 3-TB (PTQ, STQ, TTQ) are $(\lambda_{3\text{-}TB,PTQ}, \lambda_{3\text{-}TB,STQ}, \lambda_{3\text{-}TB,TTQ})$, respectively.

5.3.2 Throughput Analysis

Let α_{Pi} be the average number of TB's a packet traverses for a CoS P_i, from a given node to any other node on the ring:

$$\alpha_{P_i} = \sum_{x=2}^{N-1} (x-1) q_{x,P_i} \qquad (5\text{-}1)$$

The TB arrival rates and the client arrival rates can be related as follows:

$$\lambda_{3-TB,PTQ} = \alpha_{P_{control}}\lambda_{P_{control}} + \alpha_{P_{A0}}\lambda_{client,P_{A0}}$$

$$\lambda_{3-TB,STQ} = \alpha_{P_{A1}}\lambda_{P_{A1}} + \alpha_{P_{B-CIR}}\lambda_{client,P_{B-CIR}} \qquad (5\text{-}2)$$

$$\lambda_{3-TB,TTQ} = \alpha_{P_{B-EIR}}\lambda_{P_{B-EIR}} + \alpha_{P_C}\lambda_{client,P_C}$$

The throughput per node for traffic priority P_j is $\dfrac{\rho_{P_j}}{N} = \lambda_{P_j}\overline{S_{P_j}}$. The throughput for each of

the priority queues can be represented as follows:

$$\rho_1 = \rho_{3-TB,PTQ} = \frac{1}{N}\left(\alpha_{P_{control}}\rho_{P_{control}} + \alpha_{P_{A0}}\rho_{client,P_{A0}}\right) \qquad (5\text{-}3)$$

$$\rho_2 = \rho_{control} = \frac{\rho_{P_{control}}}{N} \qquad (5\text{-}4)$$

$$\rho_3 = \rho_{A0} = \frac{\rho_{P_{A0}}}{N} \qquad (5\text{-}5)$$

$$\rho_4 = \rho_{3-TB,STQ} = \frac{1}{N}\left(\alpha_{P_{A1}}\rho_{P_{A1}} + \alpha_{P_{B\_CIR}}\rho_{client,P_{B\_CIR}}\right) \qquad (5\text{-}6)$$

$$\rho_5 = \rho_{A1} = \frac{\rho_{P_{A1}}}{N} \qquad (5\text{-}7)$$

$$\rho_6 = \rho_{B\_CIR} = \frac{\rho_{P_{B\_CIR}}}{N} \qquad (5\text{-}8)$$

$$\rho_7 = \rho_{B\_EIR} = \frac{\rho_{P_{B\_EIR}}}{N} \qquad (5\text{-}9)$$

$$\rho_8 = \rho_C = \frac{\rho_{P_C}}{N} \qquad (5\text{-}10)$$

$$\rho_9 = \rho_{3-TB,TTQ} = \frac{1}{N}\left(\alpha_{P_{B\_EIR}}\rho_{P_{B\_EIR}} + \alpha_{P_{P_C}}\rho_{client,P_C}\right) \qquad (5\text{-}11)$$

5.3.3   Waiting Times and Packet Queueing Delay

As discussed in the previous chapter, this model can be regarded as head-of-line non-preemptive priority M/G/1 queueing system as derived in [105]. The waiting times for each of the priority queues can be described as:

$$W_1 = W_{3-TB,PTQ} = \frac{R}{1-\rho_1} = \frac{R}{1 - \frac{1}{N}\left(\alpha_{P_{control}} \rho_{P_{control}} + \alpha_{P_{P_{A0}}} \rho_{client,P_{A0}}\right)} \tag{5-12}$$

$$W_2 = W_{Control} = \frac{R}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{5-13}$$

$$W_3 = W_{A0} = \frac{R}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2-\rho_3)} \tag{5-14}$$

$$W_4 = W_{3-TB,STQ} = \frac{R}{(1-\rho_1-\rho_2-\rho_3)(1-\rho_1-\Lambda-\rho_4)} \tag{5-15}$$

$$W_5 = W_{A1} = \frac{R}{(1-\rho_1-\Lambda-\rho_4)(1-\rho_1-\Lambda-\rho_5)} \tag{5-16}$$

$$W_6 = W_{B\_CIR} = \frac{R}{(1-\rho_1-\Lambda-\rho_5)(1-\rho_1-\Lambda-\rho_6)} \tag{5-17}$$

$$W_7 = W_{B\_EIR} = \frac{R}{(1-\rho_1-\Lambda-\rho_6)(1-\rho_1-\Lambda-\rho_7)} \tag{5-18}$$

$$W_8 = W_C = \frac{R}{(1-\rho_1-\Lambda-\rho_7)(1-\rho_1-\Lambda-\rho_8)} \tag{5-19}$$

$$W_9 = W_{3-TB,TTQ} = \frac{R}{(1-\rho_1-\Lambda-\rho_8)(1-\rho_1-\Lambda-\rho_9)} \tag{5-20}$$

Where R is the mean residual service time and can be represented as:

$$R = \frac{1}{2} \left(\lambda_{3-TB,PTQ} \overline{S^2_{3-TB,PTQ}} + \lambda_{Control} \overline{S^2_{Control}} + \lambda_{A0} \overline{S^2_{A0}} + \right.$$

$$\lambda_{3-TB,STQ} \overline{S^2_{3-TB,STQ}} + \lambda_{A1} \overline{S^2_{A1}} + \lambda_{B\_CIR} \overline{S^2_{B\_CIR}} + \qquad (5\text{-}21)$$

$$\left. \lambda_{B\_EIR} \overline{S^2_{B\_EIR}} + \lambda_C \overline{S^2_C} + \lambda_{3-TB,TTQ} \overline{S^2_{3-TB,TTQ}} \right)$$

The average packet queueing delay $T_{3\text{-}TB,Pi}$ for priority Pi can be represented as:

$$T_{3-TB,P_{A0}} = W_{A0} + \alpha_{P_{A0}} W_{3-TB,PTQ} \qquad (5\text{-}22)$$

$$T_{3-TB,P_{A1}} = W_{A1} + \alpha_{P_{A1}} W_{3-TB,STQ} \qquad (5\text{-}23)$$

$$T_{3-TB,P_{B\_CIR}} = W_{B\_CIR} + \alpha_{P_{B\_CIR}} W_{3-TB,STQ} \qquad (5\text{-}24)$$

$$T_{3-TB,P_{B\_EIR}} = W_{B\_EIR} + \alpha_{P_{B\_EIR}} W_{3-TB,TTQ} \qquad (5\text{-}25)$$

$$T_{3-TB,P_C} = W_C + \alpha_{P_C} W_{3-TB,TTQ} \qquad (5\text{-}26)$$

5.3.4 RPR Packet Queueing Delay

Considering an RPR network using shortest path routing, any RPR node can transmit to another node which is less than half a ringlet away from it; otherwise, the alternate path becomes the shortest. As a result, the probability that a node transmits traffic priority P_i to node x can be represented as follows:

$$q_{x,P_i} = \begin{cases} \dfrac{2}{N-1}, & x = 1,2,\dots,\dfrac{N-1}{2}, \ N \text{ is odd} \\ \dfrac{2}{N}, & x = 1,2,\dots,\dfrac{N}{2}, \ N \text{ is even} \\ 0, & \text{otherwise} \end{cases} \qquad (5\text{-}27)$$

Therefore, we can simplify α_{Pi} to:

$$\alpha_{P_i} = \begin{cases} \dfrac{N-3}{4}, & N \text{ is odd} \\[2mm] \dfrac{N-2}{4}, & N \text{ is even} \\[2mm] 0, & \text{otherwise} \end{cases} \qquad (5\text{-}28)$$

5.3.5 RPR-FPD Packet Queueing Delay

Based on RPR-FPD algorithm and our analysis above, the probability that a node transmits traffic priority P_i to node x can be represented as follows:

$$q_{x,P_i} = \begin{cases} \dfrac{2}{N-1}, & x = 1,2,\dots,\dfrac{N-1}{2}, \, N \text{ is odd and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{2}{N}, & x = 1,2,\dots,\dfrac{N}{2}, \, N \text{ is even and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{1}{N-1}, & x = 1,2,\dots,N-1 \text{ and } C_{x,P_i}^{alternate} \\[2mm] 0, & \textit{otherwise} \end{cases} \qquad (5\text{-}29)$$

where $C_{x,P_i}^{ringlet}$ is the ringlet condition that traffic priority P_i destined to node x.

Given $d_{Pi,Th}$ denotes the delay threshold for traffic priority P_i, ρ_{Th} as the ringlet throughput threshold, also $\wedge$ and $\vee$ are the AND and OR operators, respectively, we can represent $C_{x,P_i}^{ringlet}$ as:

$$C_{x,P_i}^{LCP} = \left(\rho_{x,P_i}^{LCP} < \rho_{Th}\right) \vee \left\{\left(\rho_{x,P_i}^{LCP} \geq \rho_{Th}\right) \wedge \left[\left(\rho_{x,P_i}^{alternate} \geq \rho_{Th}\right) \vee \left(d_{x,P_i}^{alternate} \geq d_{P_i,Th}\right)\right]\right\}$$
$$C_{x,P_i}^{alternate} = \left(\rho_{x,P_i}^{LCP} \geq \rho_{Th}\right) \wedge \left(\rho_{x,P_i}^{alternate} < \rho_{Th}\right) \wedge \left(d_{x,P_i}^{alternate} < d_{P_i,Th}\right) \qquad (5\text{-}30)$$

Therefore, we can represent α_{Pi} as:

$$\alpha_{P_i} = \begin{cases} \dfrac{N-3}{4}, & N \text{ is odd and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{N-2}{4}, & N \text{ is even and } C_{x,P_i}^{LCP} \\[2mm] \dfrac{N-2}{2}, & C_{x,P_i}^{alternate} \\[2mm] 0, & \textit{otherwise} \end{cases} \qquad (5\text{-}31)$$

5.3.6 End-to-End Packet Queueing Delay for 3-TB Design

We now consider the end-to-end delay for a packet as it enters and exits an RPR ring. We represent the average end-to-end delay as follows:

$$T_{E2E,3-TB,P_{control}} = \sum_{j=1}^{N-1}(W_{P_{control}} + (j-1)W_{3-TB,PTQ})q_{j,P_{control}} \qquad (5\text{-}32)$$

$$T_{E2E,3-TB,P_{A0}} = \sum_{j=1}^{N-1}(W_{P_{Ao}} + (j-1)W_{3-TB,PTQ})q_{j,P_{A0}} \qquad (5\text{-}33)$$

$$T_{E2E,3-TB,P_{A1}} = \sum_{j=1}^{N-1}(W_{P_{A1}} + (j-1)W_{3-TB,STQ})q_{j,P_{A1}} \qquad (5\text{-}34)$$

$$T_{E2E,3-TB,P_{B-CIR}} = \sum_{j=1}^{N-1}(W_{P_{B-CIR}} + (j-1)W_{3-TB,STQ})q_{j,P_{B-CIR}} \qquad (5\text{-}35)$$

$$T_{E2E,3-TB,P_{B-EIR}} = \sum_{j=1}^{N-1}(W_{P_{B-EIR}} + (j-1)W_{3-TB,STQ})q_{j,P_{B-EIR}} \qquad (5\text{-}36)$$

$$T_{E2E,3-TB,P_C} = \sum_{j=1}^{N-1}(W_{P_C} + (j-1)W_{3-TB,STQ})q_{j,P_C} \qquad (5\text{-}37)$$

It is important to note that the above equations are applicable to both RPR and FPD algorithms.

5.4 Packet Delay Results for 3-TB

5.4.1 3-TB Packet Delay Parameters

In order to calculate the delay parameters for the 3-TB design, we modified our Packet Delay Analysis (PDA) tool to include the new analysis. The 3-TB PDA tool network and CoS parameters are the same as the ones used in our analysis in section 4.3.4. However, the TB arrival rates, throughput and waiting times are different due to our 3-TB design. For example, consider the network and CoS input parameters in Figure 5-5, specifying the number of nodes on an RPR ringlet, its capacity (Mbps), average arrival rates per CoS

(Mbps) and average packet length per CoS (Bytes). Using our PDA tool, as shown in Figure 5-6, we provide the following parameters: average number of TB's per CoS, 3-TB average arrival rates, 1st moment and 2nd moment packet sending times for each of the three TB's, throughput per CoS and all three TB's, residual time for 3-TB design, waiting times per CoS and all three TB's, and packet delay for each CoS.

Number of Nodes	10
Link Capacity (Mbps)	10000
Av. arrival - Control	5
Av. arrival - A0	10
Av. arrival - A1	15
Av. arrival - B-CIR	30
Av. arrival - B-EIR	40
Av. arrival - C	45
Av. Pkt Length - Control	128
Av. Pkt Length - A0	64
Av. Pkt Length - A1	500
Av. Pkt Length - B-CIR	1500
Av. Pkt Length - B-EIR	1000
Av. Pkt Length - C (Bytes)	500

Figure 5-5: PDA Tool Input Parameters Example

```
RPR/FPD Packet Delay Analysis (PDA) Tool - Version  1. 0. 7
======================================================
Number of nodes = 10
Link capacity = 10000
alpha(1)= 2
alpha(2)= 2
alpha(3)= 2
alpha(4)= 2
alpha(5)= 2
alpha(6)= 2

arrival rate for 3-TB PTQ = 30
arrival rate for 3-TB STQ = 90
arrival rate for 3-TB TTQ = 170
S1 for 3-TB PTQ = 0.0768
S1 for 3-TB STQ = 0.8
S1 for 3-TB TTQ = 0.6
S2 for 3-TB PTQ = 0.00000001179648
S2 for 3-TB STQ = 0.00000128
S2 for 3-TB TTQ = 0.00000072

throughput per node for CoS 1 = 0.0005
throughput per node for CoS 2 = 0.001
throughput per node for CoS 3 = 0.0015
throughput per node for CoS 4 = 0.003
throughput per node for CoS 5 = 0.004
throughput per node for CoS 6 = 0.0045
throughput for 3-TB PTQ = 0.003
throughput for 3-TB STQ = 0.009
throughput for 3-TB TTQ = 0.017

residual time for 3-TB = 0.0000000269328

3-TB waiting time for W1 = 2.78162487462387E-08
3-TB waiting time for W2 = 2.79139475627082E-08
3-TB waiting time for W3 = 2.79560077549172E-08
3-TB waiting time for W4 = 2.74247799843340E-08
3-TB waiting time for W5 = 2.77171253547253E-08
3-TB waiting time for W6 = 2.78441386582857E-08
3-TB waiting time for W7 = 2.80434320842652E-08
3-TB waiting time for W8 = 2.82882899915238E-08
3-TB waiting time for W9 = 2.89241480519711E-08


3-TB queueing delay for CoS Control, A0, A1, B-CIR, B-EIR and C
8.11364051009023E-08
8.11772519950034E-08
8.25666853233933E-08
8.26936986269536E-08
8.58917281882074E-08
8.61365860954660E-08
```

Figure 5-6: PDA Tool Results

5.4.2   Number of Nodes Impact on 3-TB

We consider the input parameters in Figure 5-5 and vary the number of nodes from 10 to 50 nodes in increments of 10. Using our PDA tool, we analyze the CoS packet delay results for 1-TB, 2-TB and 3-TB, as illustrated in Figure 5-7, Figure 5-8 and Figure 5-9, respectively.
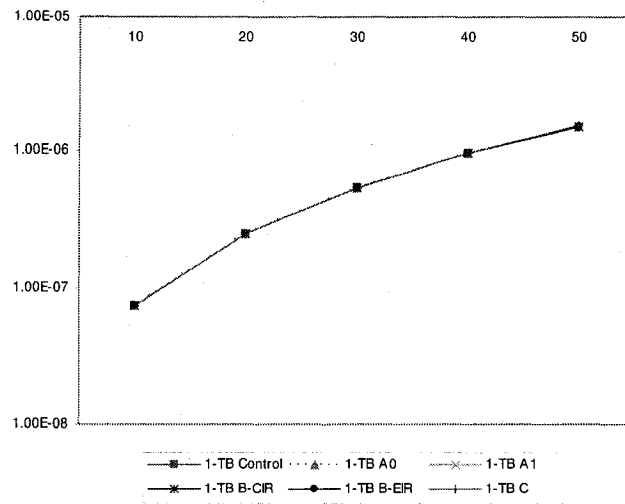
108

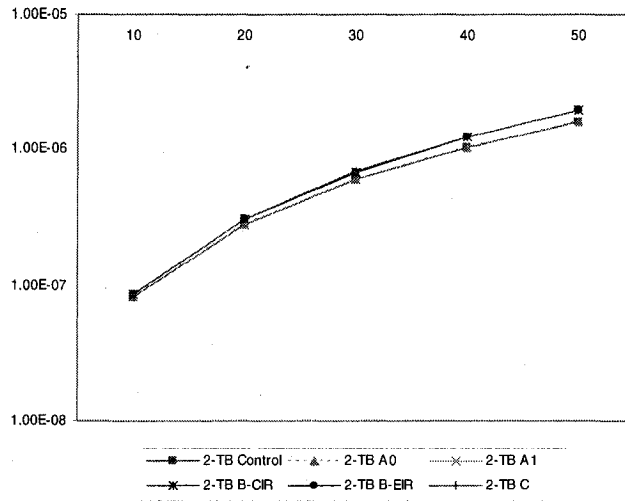Figure 5-7: No. of Nodes Impact on 1-TB Packet Delay per CoS

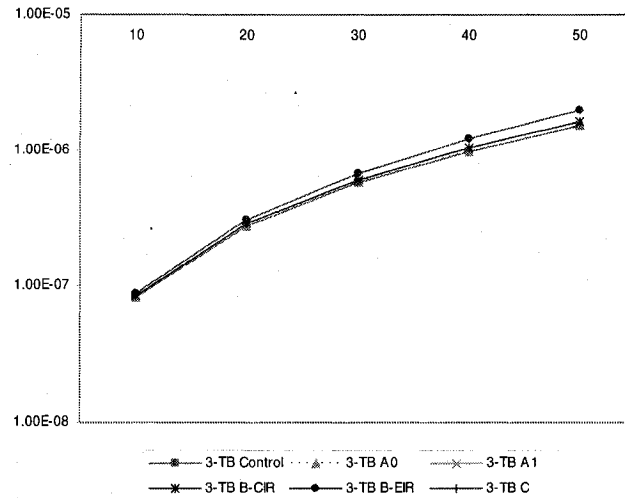Figure 5-8: No. of Nodes Impact on 2-TB Packet Delay per CoS

Figure 5-9: No. of Nodes Impact on 3-TB Packet Delay per CoS

The figures above demonstrate that the packet delay increases as the number of nodes increase on the ring. In the 1-TB case, we notice that all CoS's follow closely the same delay line with almost no CoS differentiation. Similarly, in the 2-TB, we notice two separate lines, the first for control and A0 CoS, while the remaining CoS's are on the 2nd line. As a result, there is no CoS differentiation between A1, B-CIR, B-EIR and C, which is not acceptable for SP's to enable new and next generation services. However, in the 3-TB case shown in Figure 5-9, we demonstrate that our proposed design can differentiate between the three main CoS's; hence, providing a mechanism for SP's to provide new and differentiated services especially as the number of nodes increases on the ring. In order to clearly demonstrate this, we considere the case where N=50, as illustrated in Figure 5-10. We show the values of each of the CoS's for 1-TB, 2-TB and 3-TB designs. In all three designs, the priority of CoS is maintained, i.e. control traffic has the lowest delay, whereas C CoS has the highest delay. However, in the 1-TB case, there is almost no differentiation. With the 2-TB, the control and A0 CoS have higher delay than 1-TB, and the only differentiated service is A0. With the 3-TB, there is a clear differentiation between the three classes of service. We also note that with our 3-TB design, the delay for control and A0 CoS is comparable to the 1-TB design.
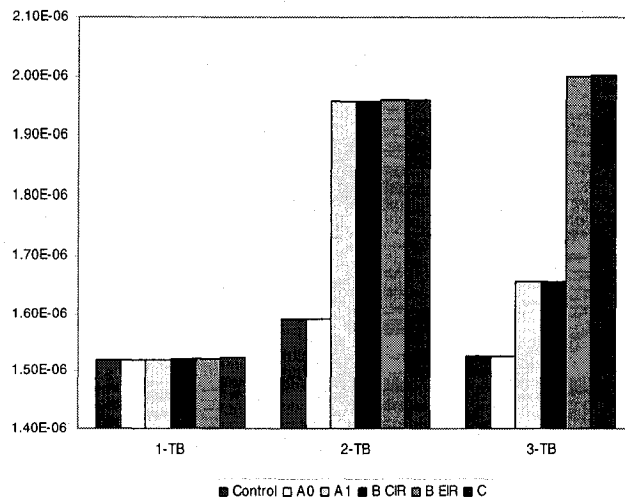
110

Figure 5-10: Packet Delay for N = 50 for: 1-TB, 2-TB and 3-TB Designs

5.4.3   Ringlet Capacity Impact on 3-TB

Given the input parameters in Figure 5-5, we now vary the capacity of the ringlet from 2Gbps to 10Gbps in increments of 2Gbps. Our PDA tool runs these inputs parameters and calculates the packet delay for each CoS and for 1-TB design shown in Figure 5-11, 2-TB design as in Figure 5-12 and 3-TB design provided in Figure 5-13.
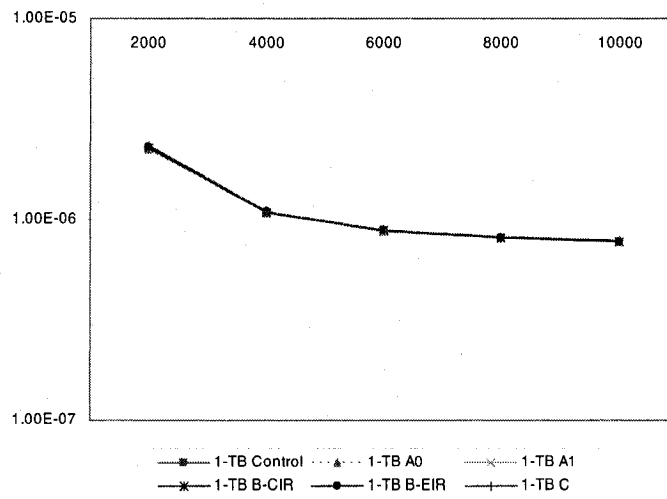

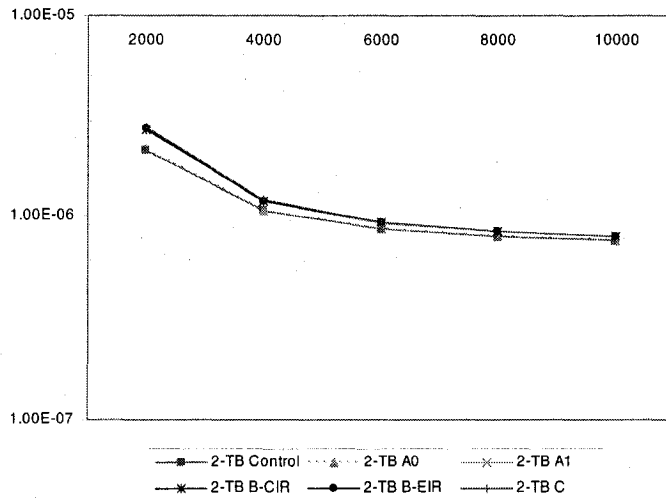Figure 5-11: Capacity Impact on 1-TB Packet Delay per CoS

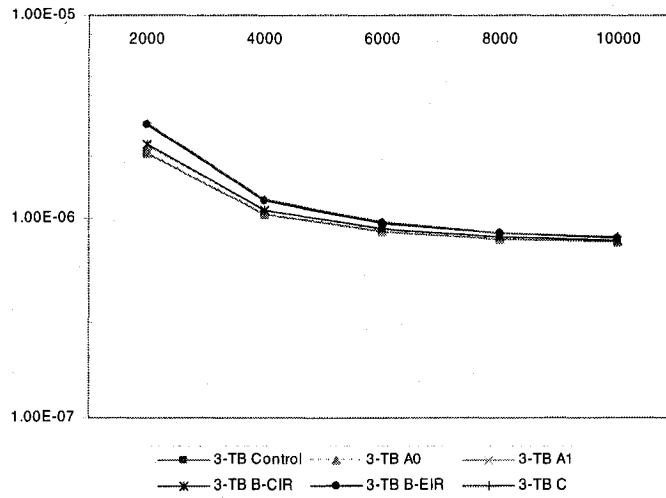Figure 5-12: Capacity Impact on 2-TB Packet Delay per CoS

Figure 5-13: Capacity Impact on 3-TB Packet Delay per CoS

The packet delay decreases as the capacity of the ring increases, irrespective of the TB design, as shown in the three figures above. We also note that the 3-TB design provides better differentiation than 1-TB and 2-TB, with three classes of service, as shown in Figure 5-14 with 2Gbps ringlet capacity. In the 1-TB case, there is little differentiation amongst the CoS's. With the 2-TB design, we have two distinct types of services. However, the 3-TB design provides solid differentiation between three classes of service, with the lowest delay for control and A0 CoS.
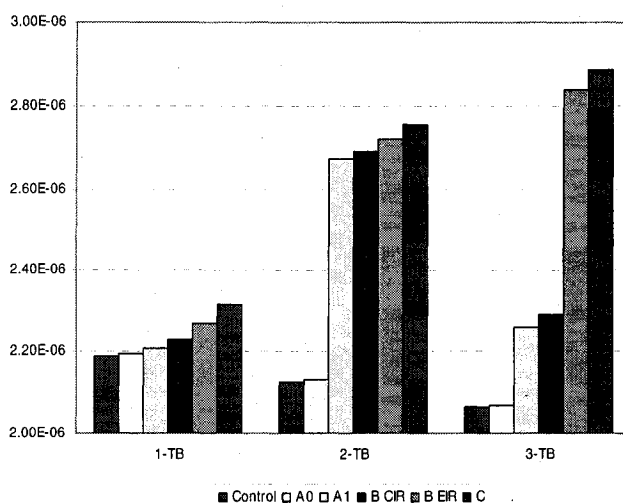
112

Figure 5-14: Packet Delay for 2Gpbs Capacity for: 1-TB, 2-TB and 3-TB Designs

5.4.4 Average Arrival Rate Impact on 3-TB

We now consider the impact of the average arrival rates. Using our PDA tool and the parameters in Figure 5-15, we calculate the packet delay for each of the ranges.

	Starting Value	Inc.	Increment Value	Ending Value
Number of Nodes	10	□		
Link Capacity (Mbps)	10000	□		
Av. arrival - Control	5	□	200	1100
Av. arrival - A0	10	□	200	1100
Av. arrival - A1	15	□	200	1100
Av. arrival - B-CIR	30	□	200	1100
Av. arrival - B-EIR	40	□	200	1100
Av. arrival - C	45	□	200	1100
Av. Pkt Length - Control	128	□		
Av. Pkt Length - A0	64	□		
Av. Pkt Length - A1	500	□		
Av. Pkt Length - B-CIR	1500	□		
Av. Pkt Length - B-EIR	1000	□		
Av. Pkt Length - C (Bytes)	500	□		

Figure 5-15: Ranges for Average Arrival Rates per CoS

Since 3-TB provides three tiers of service, we first consider the highest CoS for each tier, i.e. Control, A1 and B-EIR traffic, as shown in Figure 5-16. Then, we consider the remaining classes A0, B-CIR and C, in Figure 5-17.

113

Figure 5-16: Av. Arrival Rate Impact on 3-TB Design: Control, A1 and B-EIR CoS's
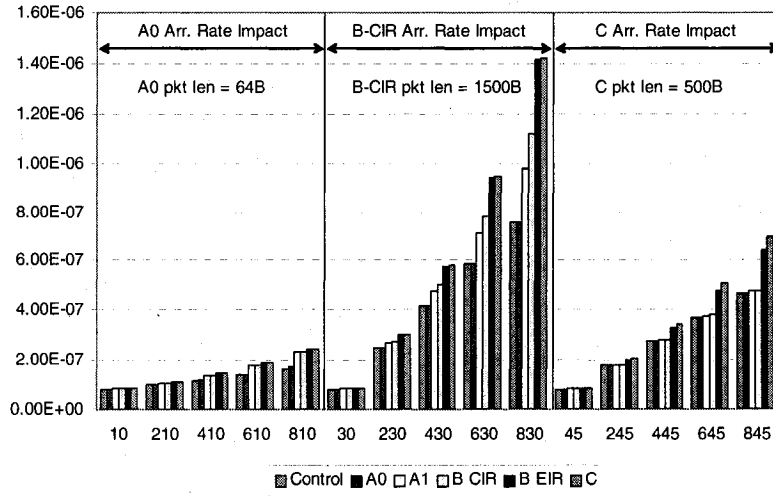
Figure 5-17: Av. Arrival Rate Impact on 3-TB Design: A0, B-CIR and C CoS's

We notice with both figures that increasing the average arrival rate for the Tier 1 classes (Control, A0) results in a consistent increase in the delay across all CoS's. Meanwhile, increasing the average arrival rate in Tier 2 classes (A1, B-CIR) results in higher delays across all CoS's specifically more noticeable in the Tier 2 and 3 classes. Similarly, average arrival rate increases in Tier 3 classes (B-EIR, C) has a higher delay increase on its respective Tier 3 classes.

One of the key parameters to keep in mind is the packet size for each CoS. As the packet size increases, we notice that an increase in the average arrival rate will result in a higher increase in the packet delay. For example, consider Tier 2 services A1 and B-CIR, their packet sizes are 500 bytes and 1500 bytes, respectively. As a result, any increase in arrival rates in this tier of service results in a higher impact on the packet delay. We will research the impact of the packet length in the next section.

5.4.5 *Average Packet Length Impact on 3-TB*

In this section, we consider the impact of the average packet length on packet delay for all CoS's given the ranges in Figure 5-18.

	Starting Value	Inc.	Increment Value	Ending Value
Number of Nodes	10			
Link Capacity [Mbps]	10000			
Av. arrival - Control	5			
Av. arrival - A0	10			
Av. arrival - A1	15			
Av. arrival - B-CIR	30			
Av. arrival - B-EIR	40			
Av. arrival - C	45			
Av. Pkt Length - Control	100		300	1500
Av. Pkt Length - A0	100		300	1500
Av. Pkt Length - A1	100		300	1500
Av. Pkt Length - B-CIR	100		300	1500
Av. Pkt Length - B-EIR	100		300	1500
Av. Pkt Length - C [Bytes]	100		300	1500

Figure 5-18: Ranges for Average Packet Length per CoS

Now, we consider the highest CoS for each tier, i.e. Control, A1 and B-EIR traffic, shown in Figure 5-19, and the remaining classes A0, B-CIR and C, as in Figure 5-20.
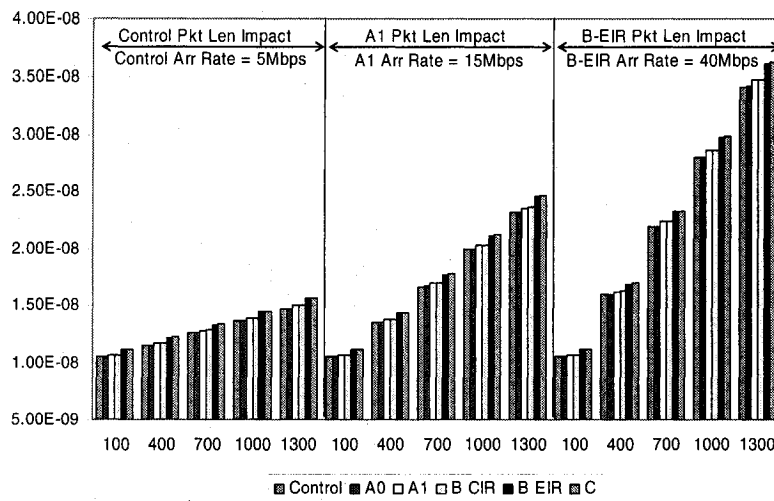
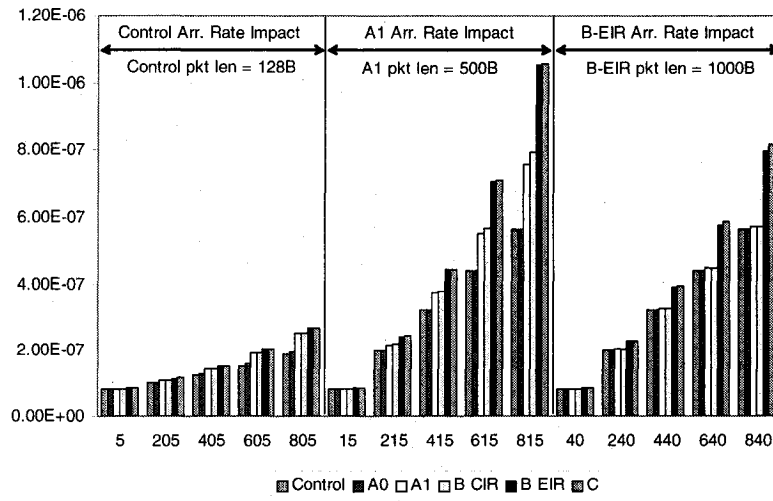Figure 5-19: Av. Packet Length Impact on 3-TB Design: Control, A1 and B-EIR CoS's
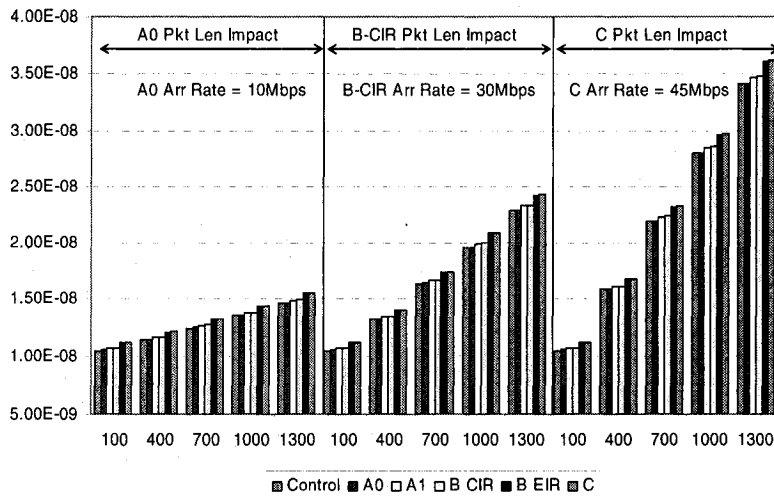
Figure 5-20: Av. Packet Length Impact on 3-TB Design: A0, B-CIR and C CoS's

We notice from both figures that increasing the average packet length results in increasing the packet delay. The increase becomes more pronounced as we move into Tier 2 and even higher into Tier 3 classes. An important impact of the packet length is also within the same Tier of service. Consider an average packet length of 1300 bytes; we notice that the packet delay in Figure 5-19 is higher than the same packet length for the same tier of service in Figure 5-20, even though the arrival rate is higher in the latter. As a result, within the same tier of service, higher CoS such as A1, has a slightly higher packet delay than B-CIR which is a lower CoS in a Tier 2 service.

116

5.4.6   FPD Impact on 3-TB

In our analysis for 3-TB design, we have focused on routing traffic on a single ringlet. As we demonstrated in the previous chapter, our algorithm RPR-FPD results in increasing RPR utilization at the expense of higher packet delay. In this section, we consider the impact of FPD on packet delay for Tier 2 and 3 CoS's, as in Figure 5-21.

	Starting Value	Inc.	Increment Value	Ending Value
Number of Nodes	10	☑	10	50
Link Capacity (Mbps)	10000	☐		
Av. arrival - Control	5	☐		
Av. arrival - A0	10	☐		
Av. arrival - A1	15	☐		
Av. arrival - B-CIR	30	☐		
Av. arrival - B-EIR	40	☐		
Av. arrival - C	45	☐		
Av. Pkt Length - Control	128	☐		
Av. Pkt Length - A0	64	☐		
Av. Pkt Length - A1	500	☐		
Av. Pkt Length - B-CIR	1500	☐		
Av. Pkt Length - B-EIR	1000	☐		
Av. Pkt Length - C (Bytes)	500	☐		

☐ Control FPD   ☐ A0 FPD   ☑ A1 FPD   ☑ B-CIR FPD   ☑ B-EIR FPD   ☑ C FPD

Figure 5-21: PDA Input Parameters with FPD Option for Tier 2 and 3 Services

Our results for enabling FPD on a subset of the services are illustrated in Figure 5-22. As we concluded in the previous chapter, enabling FPD results in increasing the packet delay while leveraging any available bandwidth on the alternate ringlet especially during congestion periods on the primary ringlet. As expected, the classes using FPD have a higher packet delay due to utilizing the alternate ringlet. However, our 3-TB design provides CoS differentiation between the various tiers of services. Hence, using a 3-TB with FPD on Tier 2 and 3 services provides higher utilization on both ringlets while delivering service differentiation, as opposed to the 1-TB or 2-TB designs with FPD option that result in no differentiation between Tier 2 and 3 services, as indicated in Figure 5-23.

117
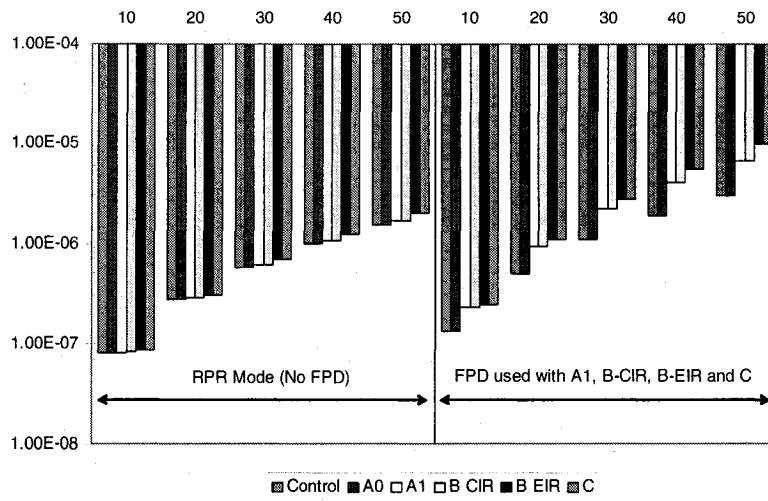
Figure 5-22: RPR-FPD Impact on 3-TB Design with N = 10 to 50
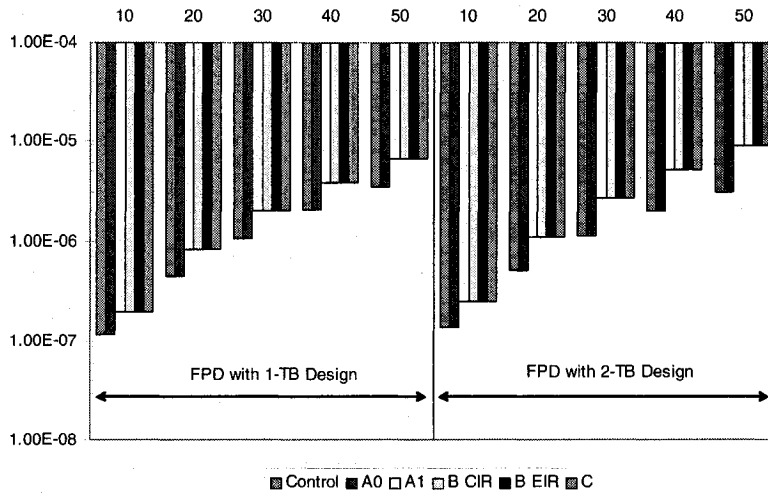
Figure 5-23: RPR-FPD with 1-TB and 2-TB Designs with N = 10 to 50

5.5 Summary

In this chapter, we detailed the 1-TB and 2-TB designs and provided examples of how these design can deliver inconsistent CoS across the network; hence, the current RPR 1-TB and 2-TB designs are unable to provide the required CoS differentiation. As a result, we introduced our proposed 3-TB design and described its operation.

Then, we derived packet delay analysis for our new design. We also updated our PDA tool to include our new analysis. Using our tool, we demonstrated the advantage of our proposed scheme to differentiate between three classes of service, whereas current RPR schemes were unable to achieve this level of required differentiation.

CHAPTER 6 - Conclusion and Future Work

6.1 Thesis Summary

In Chapter 1, we provided an overview of multi-service broadband networks. We discussed the challenges that service providers are facing and how broadband convergence over a packet-based infrastructure is a critical element to address these challenges. Then, we described the key role that IP and MPLS play in the evolution of SP networks to deliver broadband convergence. We reviewed various network topologies and the importance of ring architecture, specifically how RPR was recently standardized to address these challenges at the transport level. Finally, we summarized recent research conducted on RPR and listed areas of contributions which assist service providers to deliver on their objectives.

We described how multi-service broadband networks have to evolve to achieve convergence, in Chapter 2. Also, we outlined the main components of IP protocol and next generation IP-based services. Then, we discussed MPLS protocol, architecture and IP/MPLS requirements for broadband services convergence, namely: Pseudo-Wire Emulation, Traffic Engineering, Quality of Service, resiliency and scalability. Furthermore, we reviewed SP network topologies and transport solutions, where we highlighted the value of ring topologies and namely RPR as a critical transport solution for service providers. We introduced RPR and its role in multi-service broadband convergence and described its architecture.

Chapter 3 classified IP services into four classes: Voice, Video, low-speed data and high-speed data. We outlined IP service sources according to their short and long-term dependence properties and models. Then, we summarized relevant traffic modeling research that is applicable to IP traffic and services. Also, we discussed recent ring and RPR performance research. We highlighted how recent research addressed some of the service provider challenges and RPR drawbacks. Finally, we discussed how RPR enables

120

multi-service support, and we highlighted how it is inefficient in optimizing the network utilization and does not maximize service fulfillment.

We introduced our new algorithm Flow-based Priority Distribution (FPD) in Chapter 4. The algorithm and simulation environment were described. We demonstrated through our results the advantage of FPD over the current RPR standard for both 1-TB and 2-TB designs, specifically in maximizing ring utilization and increasing service fulfillment. Nevertheless, this increase was shown to be at the expense of packet delay. We derived the packet delay for 1-TB and 2-TB designs, and we analyzed the results for RPR and FPD with both designs considering various input parameters such as: number of nodes on a ring, ring capacity, average arrival rate per CoS and average packet length per CoS. It is important to note that we considered all CoS's supported by the standard in our algorithm, simulation and analysis.

In Chapter 5, we discussed delivering consistent CoS across the network, and how the current RPR 1-TB and 2-TB designs are unable to provide this CoS differentiation. As a result, we introduced our proposed 3-TB design and described its operation. Then, we derived packet delay analysis for our new design. Our results demonstrated the advantage of our proposed scheme to differentiate between the main three classes of services proposed by RPR, whereas current RPR scheme was unable to achieve this level of required differentiation.

Finally, Chapter 6 provided a summary of our report. We also discussed our contributions and conclusions. We outlined several areas to enhance and expand our research to enable service providers to migrate to packet-based architecture and enable new IP services.

6.2 Conclusions

In this report, we introduced two new contributions to enhance multi-service broadband ring networks, namely, FPD and 3-TB design. These two new techniques allow service providers to enhance their network utilization, increase service fulfillment and provide

differentiated service level agreements. These three advantages are very important for service providers to migrate the current circuit-based network to a packet based infrastructure.

We demonstrated that FPD addresses the dual-ringlet bandwidth inefficiency in current RPR standard. FPD provides a mechanism to enhance current RPR ringlet selection by enabling per flow routing; hence, it increases bandwidth utilization across both ringlets. As a result, FPD helps in avoiding congestion and reducing the need to invoke the fairness algorithm. Nonetheless, this results in certain flows utilizing the longer path; thus, higher packet delays for certain flows. We confirmed this result for both 1-TB and 2-TB designs. The 1-TB design provided consistent and lower delays across all CoS's, whereas the 2-TB design provided lower PTQ delays, but noticeably higher STQ delay, as desired. Hence, FPD is suitable for CoS's that can tolerate a certain level of increase in packet delays.

In addition, we addressed the fourth limitation of RPR where CoS traffic handling is inconsistent across the network. We introduced a new design called 3-TB which allows for three classes of differentiation. The packet delay analysis and results confirmed the outcome and value of 3-TB over current designs. Another key benefit to our new design is that it reduces the delay on PTQ traffic which is the delay-sensitive class, and outperforms 1-TB and 2-TB designs. However, our 3-TB is more complex to implement and would require additional memory for TTQ. We recommend implementing 3-TB for to offer differentiated services especially with delay-sensitive CoS, as it enables service providers to deliver various levels of services and evolve their TDM traffic and services more reliably.

6.3 Recommendations and Future Work

6.3.1 Recommendations to Enhance Multi-Service Broadband Networks

In order for service providers to achieve their goal of delivering differentiated next generation IP services while leveraging the current next network and migrating to a pure packet-based infrastructure, we put forward the following three recommendations:

First, service providers can benefit significantly from deploying MPLS over RPR as it leverages their current ring topology, provides a more efficient transport mechanism, enables a multi-service (and multi-protocol) model to transport both TDM-based and packet-based services, delivers services reliably across the network, and is a scalable technology from Mbps to 10's of Gbps.

Second, given the current drawbacks of RPR, we propose to enhance the RPR bandwidth utilization and service fulfillment using our FPD algorithm. The benefit of our proposed scheme is that it can co-exist with current RPR implementation. Hence, equipment vendors can enable service providers to use FPD for the CoS's that can tolerate a certain level of packet delay. For example, A0 and A1 classes will not use FPD, whereas Classes B-CIR, B-EIR and C can fully benefit from the higher utilization that FPD enables.

Third, in order to address the inconsistent CoS delivery across the network, we recommend adoption of our new 3-TB design. As per our results and conclusions, the 3-TB design allows the service providers to clearly differentiate between the three classes of service that RPR supports. In addition, 3-TB provides the lowest PTQ packet delay. Hence, we believe that service providers can benefit greatly by combining FPD and 3-TB over RPR transport networks in the following fashion: disable FPD for PTQ traffic, and enable FPD for STQ and TTQ traffic. This configuration enables the lowest packet delay for PTQ, and maximizes the utilization and service fulfillment.

6.3.2 Future Work and Expanded Research

Our research can be extended in several directions. First, the impact of FPD assigning flows to alternate paths should be researched further. For example, an assignment of a flow to the longer path may block newer flows on the alternate ringlet from accessing the shorter paths. A mechanism can be development to override such flows with FPD enabled depending on the flow CoS parameters and priority. A key element to take into account while performing such a mechanism is packet reordering while switching ringlets.

123

Another area of research is the impact of the 3-TB design on RPR node implementation complexity, in terms of both hardware and software designs. At least, the hardware impact can be quantified in costs per node and whether it is prohibitive or not.

A third area for future investigation is to consider the various traffic models presented in Chapter 3. The result on RPR utilization and packet delay can be studied and analyzed for each of the models considered.

A fourth item to research is a real-time RPR model based on a programmable Network Processor (NP) system that can be built to validate the simulation and results for 1-TB, 2-TB and 3-TB designs, as shown in Figure 6-1.
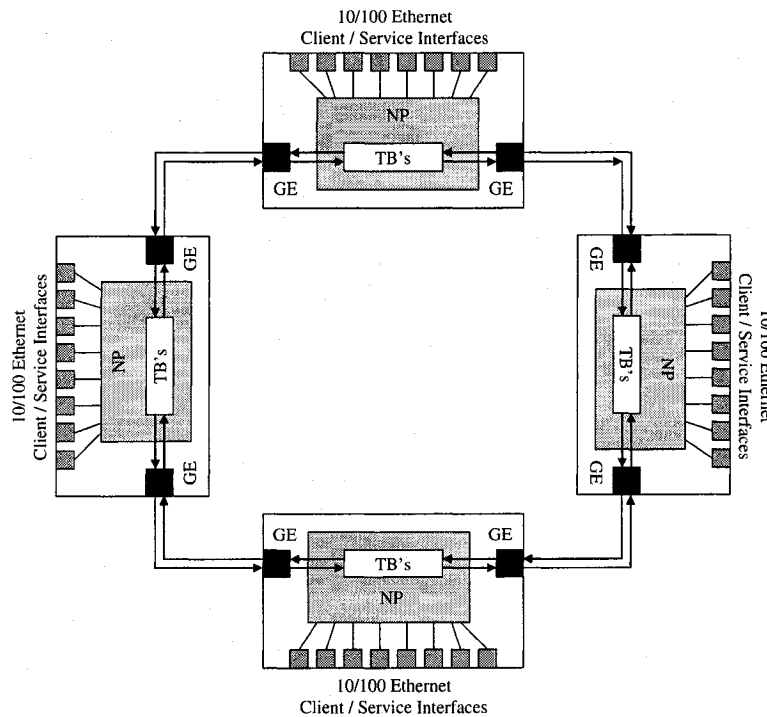
Figure 6-1: Proposed Real-Time RPR System

Each RPR node can be implemented using a system of multiple 10/100 Ethernet ports as client and service ports, and two Gigabit Ethernet ports providing the dual RPR ringlets as the WAN transport connectivity. The NP, at the heart of the system, can be programmed to assign the CoS queues required for each client port as well as the TB's required between

124

the WAN ports. Similarly, the same system can be used to compare the utilization and service fulfillment of FPD on an RPR network. These results can be categorized for each CoS considered.

Furthermore, an extension to this model can adopt real applications and services. Operational tools can be implemented to measure packet delay and other service parameters. Such tools will be used in real life deployments by service providers, and will provide significant enhancements and savings to their deployments and operational costs.

REFERENCES

[1] P. Izzo, *"Gigabit Networks"*, Wiley Computer Publishing, 2000.

[2] J. Wei, K. K. Ramakrishnan, R. Doverspike, J. Pastor, L. Aguirre-Torres, C. Kalmanek, T. Afferton, "Convergence through Packet-Aware Transport", *Journal of Optical Networking*, Vol. 5, No. 4, April 2006.

[3] Frame Relay Forum, The ATM Forum and MPLS Forum, *"Converged Data Networks"*, White Paper, April 2004.

[4] J.P. Fernandez-Palacios, A. Ferreiro, J. F. Lobo, K. Wajda and N. Le Sauze, "Migration Scenarios for Metro Networks", *2006 Proceedings of the Seventh IEEE International Symposium on Computer Networks (ISCN'06)*, pp. 1-6, June 2006.

[5] X. Xiao, "Requirements for Pseudo-Wire Emulation Edge-to-Edge", *IETF Request for Comments*, RFC3916, Sep. 2004.

[6] V. Gambiroza, P. Yuan, L. Balzano, Y. Liu, S. Sheafor and E. Knightly, "Design, Analysis, and Implementation of DVSR: A Fair High-Performance Protocol for Packet Rings", *IEEE/ACM Transactions On Networking*, Vol. 12, No. 1, pp. 85-102, February 2004.

[7] P. Yuan, V. Gambiroza, E. Knightly, "The IEEE 802.17: Media Access Protocol for High-Speed Metropolitan-Area Resilient Packet Rings", *IEEE Network*, Vol. 18, Issue 3, pp. 8-15, May/June 2004.

[8] L. Tan, H. Wang and M. Zukerman, "Adaptive Bandwidth Allocation for Metropolitan and Wide-Area Networks", *IEEE Communications Letters*, Vol. 9, Issue. 6, pp. 561-563, June 2005.

[9] F. Davik, A. Kvalbein and S. Gjessing, "Improvement of Resilient Packet Ring Fairness", *IEEE Global Telecommunications Conference, Globecomm'05*, Vol. 1, November/December 2005.

[10] H. Tang, I. Lambadaris, H. Mehrvar, J. Talim and A. Matrawy, "A New Access Control Scheme for Metropolitan Packet Ring Networks", *IEEE Global Telecommunications Conference Workshops, Globecom Workshops 2004*, pp. 281-287, November/December 2004

[11] C. Liu and J. S. Li, "Improving RPR Fairness Convergence", *The 2004 IEEE Asia-Pacific Conference on Circuits and Systems*, Vol. 1, pp. 469-472, December 2004.

[12] F. Alharbi and N. Ansari, "A Novel Fairness Algorithm for Resilient Packet Ring Networks with Low Computational and Hardware Complexity", *13th IEEE Workshop on Local and Metropolitan Area Networks 2004*, pp. 11-16, April 2004.

[13] D. Schupke and A. Riedl, "Packet Transfer Delay Comparison of a Store-and-Forward and Cut-Through Resilient Packet Ring", *IEEE 2002 International Zurich Seminar on Broadband Communications*, pp.12-1 – 12-5, July 2002.

[14] S. Zhou, F. Ao and M. Chen, "Delay and Throughput Performance Analysis of RPR's Three Priorities Traffic", *2004 International Conference on Communications, Circuits and Systems (ICCCAS 2004)*, Vol. 1, pp. 660-664, June 2004.

127

[15] C. Huang, H. Peng, F. Yuan, J. Hawkins, "A Steady State Bound for Resilient Packet Rings", *IEEE Global Communications Conference, Globecom 2003*, Vo. 7, pp. 4054-4058, December 2003.

[16] Y. Robichaud, C. Huang, J. Yang, H. Peng, "Access Delay Performance of Resilient Packet Ring under Bursty Periodic Class B Traffic Load", *2004 IEEE International Conference on Communications (ICC'04)*, Vol. 2, pp. 1217-1221, 2004.

[17] C. Huang, H. Peng and Y. Fengie, "A Deterministic Bound for the Access Delay of Resilient Packet Rings", *IEEE Communications Letters*, Vol. 9, Issue 1, pp. 87-89, January 2005.

[18] C. B. Kawwas and M. R. Soleymani, "FPD: A Novel Scheme to Enhance RPR Performance in Service Provider Networks", *IEEE Communication Letters*, Vol. 11, Issue 1, January 2007.

[19] C. B. Kawwas and M. R. Soleymani, "Performance Enhancements and Analysis in Next Generation Resilient Packet Ring Networks", *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) 2006*, pp. 738-741, May 2006.

[20] J. Postel, "Internet Protocol", *IETF Request for Comments*, RFC791 , September 1981.

[21] D. Kosiur, *"IP Multicasting"*, Wiley Computer Publishing, 1998.

[22] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", *IETF Request for Comments*, RFC2460, Dec. 1998.

[23] J. Postel, "Internet Control Message Protocol", *IETF Request for Comments*, RFC792, Sep. 1981.

[24] C. Hedrick, "Routing Information Protocol", *IETF Request for Comments*, RFC1058, June 1988.

[25] J. Malkin, "RIP Version 2", *IETF Request for Comments*, RFC2453, Nov. 1998.

[26] J. Moy, "OSPF Version 2", *IETF Request for Comments*, RFC2328, April 1998.

[27] Y. Rekhter, "A Border Gateway Protocol 4 (BGP-4)", *IETF Request for Comments*, RFC1771, March 1995.

[28] W. Fenner, "Internet Group Management Protocol Version 2", *IETF Request for Comments*, RFC2236, November 1997.

[29] D. Watzman, "Distance Vector Multicast Routing Protocol", *IETF Request for Comments*, RFC1075, November 1988.

[30] J. Moy, "Multi-cast Extensions to OSPF", *IETF Request for Comments*, RFC1584, March 1994.

[31] A. Adams, "Protocol Independent Multicast Dense Mode (PIM-DM): Protocol Specification", *IETF Request for Comments*, RFC3973, January 2005.

[32] B. Fenner, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification", *IETF Request for Comments*, RFC4601, August 2006.

[33] L. Aguirre-Torres and G. Rosenfeld, "High-Availability in Multipoint to Multipoint Ethernet for the Delivery of Triple Play Services", Optical *Fiber*

Communication Conference/National Fiber Optics Engineers Conference (OFC/NFOEC'06), 2006.

[34] E. Rosen, "Multi-Protocol Label Switching Architecture", *IETF Request for Comments,* RFC3031, Jan 2001.

[35] L. Martini, "Pseudowire Setup and Maintenance using LDP", *IETF Draft,* draft-ietf-pwe3-control-protocol-14.txt, December 2004.

[36] V. Fineberg, "QoS Support in MPLS Networks", MPLS/Frame Relay Alliance White Paper, *MPLS Forum,* May 2003.

[37] D. Awduche, "RSVP-TE: Extensions to RSVP for LSP Tunnels", *IETF Request for Comments,* RFC3209, December 2001.

[38] F. Le Faucheur, "Multi-Protocol Label Switching (MPLS) support of Differentiated Services", *IETF Request for Comments,* RFC3270, May 2002.

[39] S. Chen, "Evolution of Protection Technologies in Metro Core Optical Networks", *2006 International Conference on Networking and Services (ICNS'06),* pp. 89-94, July 2006.

[40] P. Pan, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", *IETF Draft,* draft-ietf-mpls-rsvp-lsp-fastreroute-00.txt, July 2002.

[41] M. Herzog and M. Maier, "RINGOSTAR: An Evolutionary Performance-Enhancing WDM Upgrade of IEEE 802.17 Resilient Packet Ring", *IEEE Communications Magazine,* Vol. 44, Issue 2, pp. S11-S17, February 2006.

[42] IEEE 802.17, "Resilient Packet Ring: Access Method and Physical Layer Specification", *IEEE Standards*, LAN/MAN Standards Committee, http://www.ieee802.org/17/, 2004.

[43] C. B. Kawwas and M. R. Soleymani, "Resilient Packet Ring Performance Enhancements", *IEEE Infocom 2006*, Student Workshop, April 2006.

[44] H. Tang, I. Lambadaris, H. Mehrvar, J. Talim and A. Matrawy, "A New Access Control Scheme for Metropolitan Packet Ring Networks", *IEEE Global Telecommunications Conference Worshops*, Globecom Workshops 2004, pp. 281-287, 2004.

[45] IEEE 802.17b, "IEEE 802.17b: Spatially Aware Bridging over RPR", *IEEE Standards*, Draft 7.0, http://www.802.org/17, 2007.

[46] P. Setthawong, C. Siriakkarap and S. Tanterdtid, "Inter-Ring Traffic Engineering for Bridged IEEE 802.17 Resilient Packet Rings", *2005 Proceedings 6th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2005)*, pp. 433-438, November 2005.

[47] M. Holeness and G. Parsons, "Mapping of IP/MPLS packets into IEEE 802.17 (Resilient Packet Ring) Networks", draft-ietf-iporpr-basic-02.txt, *IETF Draft*, IETF, http://www.ietf.org/internet-drafts/, June 2006.

[48] H. Schulzrinne, "RTP: A Transport Protocol for Real-Time Applications", *IETF Request for Comments*, RFC1889, January 1996.

[49] J. Rosenberg, "SIP: Session Initiation Protocol", *IETF Request for Comments*, RFC3261, June 2002.

[50] ITU-T, *"Packet-based Multimedia Communications Systems"*, H.323, July 2003.

[51] X. Yuan, M. Ilyas, *"Modeling of Traffic Sources in ATM Networks"*, IEEE Proceedings Southeast Con 2002, pp. 82-87, 2002.

[52] C. B. Kawwas and M. R. Soleymani, "Estimating Cell Loss Probability in an ATM Multiplexer with Heterogeneous Traffic", *CCBR'98*, pp. 66-77, Ottawa, Canada, Jun. 1998.

[53] R. Koenen, *"Overview of the MPEG-4 Standard"*, ISO/IEC JTC1/SC29/WG11 N4668, March 2002.

[54] WAP Forum, *"Wireless Application Protocol (WAP) Multimedia Messaging Service (MMS) Architecture Overview"*, WAP-205-MMSArchOverview-20010425-a, April 2001.

[55] N. Vicari, *"The Dependence of Internet User Traffic Characteristics on Access Speed"*, Report No. 246, University of Würzburg, Germany, January 2000.

[56] T. Ozawa, "Recent advances of the traffic theory for multimedia communication", *The Journal of the IEICE*, vol. 81, pp. 506-515, 1998.

[57] F. Huebner, D. Liu and J.M. Fernandez, "Queueing Performance Comparison of Traffic Models for Internet Traffic", *IEEE Global Telecommunications Conference*, GLOBCOMM98, vol. 1, pp. 471-476, 1998.

[58] A. Adas, "Traffic models in broadband networks", *IEEE Communications Magazine*, pp. 82-89, July 1997.

[59] H. Heffes and D. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance", *IEEE Journal of Select Areas in Communications*, Vol. SAC-4, pp. 856-86, Sept. 1986.

[60] J. Cao, "On the nonstationarity of Internet traffic", *Proceedings of ACM SIGMETRICS*, pp. 102-112, 2001.

[61] D. Heyman and D. Lucantoni, "Modeling IP traffic streams with rate limits", *IEEE/ACM Transactions on Networking*, Vol. 11, No. 6, pp. 948-958, Dec. 2003.

[62] S. Shah-Heydari and T. Le-Ngoc, "Parameter Estimation of Switched Poisson Models for Short and Long Range Dependent Traffic", *IEEE Canadian Conference on Electrical and Computer Engineering*, CCECE 2003, pp. 1035-1038, May 2003.

[63] L. Kleinrock, *"Queueing Systems"*, vol. 1, John Wiley & Sons, 1975.

[64] T. Yang and D. H. Tsang, "A Novel Approach to Estimating the Cell Loss Probability in an ATM Multiplexer Loaded with Homogeneous On-Off Sources", *IEEE Transactions on Communications*, Vol. 43, pp. 117-126, 1993.

[65] F. Huebner, "On the Accuracy of Approximating Loss Probabilities in Finite Queues by Probabilities to Exceed Queue Levels in Infinite Queues", *IEEE Global Telecommunications Conference*, GLOBCOMM98, Vol. 1, pp. 484-489, 1998.

[66] C. Barakat, "Modeling Internet Backbone Traffic at the Flow Level", *IEEE Transactions on Signal Processing*, Vol. 51, No. 8, August 2003.

[67] S. Ata, "Analysis of network traffic and its application to design of high-speed routers", *IEICE Trans. Inform. Sys.*, Vol. E83-D, pp. 988-995, May 2000.

[68] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and S. C. Diot, "Packet-level traffic measurement from the Sprint IP backbone", *IEEE Network*, Vol. 17, Issue 6, December 2003.

[69] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran and C. Diot, "Measurement and Analysis of Single-Hop Delay on an IP Backbone Network", *IEEE Journal on Select. Areas in Communications*, Vol.21, No. 6, pp. 908-921, Aug. 2003.

[70] V. Paxman, "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, 1995.

[71] W. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)", *IEEE/ACM Transactions on Networking*, Vol. 2, pp. 1-15, Feb. 1994.

[72] C. Park, D. H. Han and Y. Lee, "Queueing Analysis of Dynamic Bandwidth Allocation Scheme for IP traffic on ATM networks", *IEEE Proc. of 7th ISCC'02*, pp. 653-658, July 2002.

[73] I. Wldjaja and A. I. Elwalid, "Performance Issues in VC-Merge Capable Switches for Multiprotocol Label Switching", *IEEE Journal on Selected Areas in Communications*, Vol.17, No. 6, pp. 1178-1189, 1999.

[74] C. N. Aduba and M. N. O Sadiku, "Simulation and Analysis of different Traffic Models for ATM Networks", *Proceedings IEEE SoutheastCon 2002*, pp. 73-75, April 2002.

[75] F. Yegenoglu and B. Jabbari, "Performance of MMPP/D/1/K Queues For Aggregate ATM Traffic Models", *IEEE Infocomm'93*, pp. 1314-1319, 1993.

[76] J. Lee, "Design and Analysis of MPLS-based ATM Switching System for Differentiated Services", *IEEE Proc. of 8th ISCC'03*, 2003.

[77] R. Chakka, "Towards a generalized Markovian node model: The MM Σ CPP/GE/c/L G-queue", *International Conference on Stochastic models & 4th International Workshop on Retrial Queues*, Cochin, India, Dec. 2002.

[78] D. Kouvatsos, "Entropy Maximization and Queueing Network Models", *Ann. Of Operations Research*, pp. 63-126, 1994.

[79] D. Gross and C. M. Harris, *"Fundamentals of Queueing Theory"*, Wiley Series in Probability and Mathematics Statistics, 2nd Edition, 1985.

[80] V. S. Frost and B. Melamed, "Traffic Modeling for Telecommunication Networks", *IEEE Communication Magazine*, pp. 70-81, March 1994.

[81] A. Baiocchi, N. B. Melazzi, M. Listanti, A. Roveri and R. Winkler, "Loss Performance Analysis of an ATM Multiplexer Loaded with High-Speed On-Off Sources", *IEEE J. Select. Areas Communications*, Vol. 9, No. 3, pp. 388-393, April 1991.

[82] S. Kim, M. Y. Lee and M. J. Kim, "Σ-Matching Technique for MMPP Modeling of Heterogeneous On-Off Sources", *Globecom '94*, pp.149-171, 1992.

[83] J Huang, T. LeNgoc and J. F. Hayes, "Broadband SATCOM system for multimedia services", *Proc. ICC'96*, pp. 906-909, 1996.

[84] W. Fischer and K. Meier-Hellstern, *"The Markov-Modulated Poisson Process Cookbook"*, Performance Evaluation, pp. 149-171, 1993.

[85] A. T. Andersen and B. F. Nielsen, " A Markovian Approach for Modeling Packet Traffic with Long-Range Dependence", *IEEE J. Selected Areas of Communications*, pp. 719-732, 1998.

[86] T. V. J. Babu, T. Le-Ngoc and J. F. Hayes, "Performance of a Priority-Based Dynamic Capacity Allocation Scheme for Wireless ATM Systems", *IEEE J. Selected Areas in Communications*, Vol. 19, No. 2, Feb. 2001

[87] G. Min, M. Ould-Khaoua, D. D. Kouvatsos and I. U. Awan, "A Queueing Model of Dimen-Ordered Routing under Self-Similar Traffic Loads", *Proc. of the 18th International Parallel and Distributed Processing Symposium*, 2004.

[88] J. Gordon, "Pareto Process as a Model of Self-Similar Packet Traffic", *IEEE Globecom'95*, pp. 2232-2236, Nov. 1995.

[89] Y. Koh and K. Kim, "Loss Probability Behavior of Pareto/M/1/K Queue", *IEEE Communications Letters*, Vol.7, No.1, Jan. 2003.

[90] S. Subramanian and T. LeNgoc, "Traffic Modeling in a Multi-Media Environment", *Canadian Conference on Electrical & Computer Engineering*, Vol. 2, pp. 838-841, Sept. 1995.

[91] I. Norros, "A Storage Model with Self-Similar Input", *Queueing Systems*, Vol. 16, pp. 387-396, 1994.

[92] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks", *IEEE JSAC*, Vol. 13, No. 6, pp. 953-962, August 1995.

[93] H. Ahn and J-K. Kim, "Shifting-Level Process as a LRD Video Traffic Model and Related Queueing Results", *Computer Communications.*, Vol. 23, pp. 371-378, 2000.

[94] M. Grasse, "A New Approach to Long-Range Dependence in Variable Bit Rate Video Traffic", *Telecommunications Systems*, Vol. 12, pp. 79-100, 1999.

[95] E. Hafner, Z. Nendal, and M. Tschanz, "A digital loop communication system", *IEEE Transactions on Communications*, vol. 22, no. 6, pp. 877–881, June 1974.

[96] K. Tanno, "Buffer Insertion/Self-Token (BIST) Protocol for Multimedia LANs", *IEEE International Conference on Communication ICC'95*, pp. 350-355, November 1995.

[97] M. J. Francisco, F. Yuan, C. Huang and H. Peng, "A comparison of two buffer insertion ring architectures with fairness algorithms", *IEEE International Conference on Communication ICC'03*, Vol. 2, pp. 1375-1380, May 2003.

[98] I. Cidon, L. Georgiadis, R. Guerin and Y. Shavitt, "Improved Fairness Algorithm for Rings with Spatial Reuse", *IEEE/ACM Transaction on Networking*, Vol. 5, No. 2, pp. 190-204, April 1997.

[99] B. Vandalore, "General Weighted Fairness and its support in explicit rate switch algorithms", *Computer Communications*, Vol. 23, pp. 149-161, Jan. 2000.

[100] F. Alharbi and N. Ansari, "SSA: Simple Scheduling Algroithm for Resilent Packet Ring Networks", *IEE Proceedings on Communications*, Vol. 153, No. 2, pp. 183-188, April 2006.

[101] M. Chen, S. Zhou and F. Ao, "A Fairness Algorithm Based on Flow for RPR", *IEEE 2006 International Conference on Communications, Circuits and Systems Proceedings*, Vol. 3, pp. 2013-2015, June 2006.

[102] M. Chen, X. Wei, S. Zhou and F. Ao, "On Queue Length of RPR", *IEEE 2006 International Conference on Communications, Circuits and Systems Proceedings*, Vol. 3, pp. 1772-1776, June 2006.

[103] P. Yue, Z. Liu and J. Liu, "High Performance Fair Bandwidth Allocation Algorithm for Resilient Packet Ring", *Proceedings of the 17th International Conference on Advanced Information Networking and Applications* (AINA'03), IEEE Computer Society, September 2003.

[104] X. Zhou, G. Shi, H. Fang and L. Zeng, "Fairness Algorithm Analysis in Resilient Packet Ring", *International Conference on Communication Technology*, Vol. 1, pp. 622-624, April 2003.

[105] D. Bertsekas and R. Gallager, *"Data Networks"*, Prentice Hall, 1992.