INTELLIGENT BASED PACKET SCHEDULING SCHEME USING INTERNET PROTOCOL/MULTI-PROTOCOL LABEL SWITCHING (IP/MPLS) TECHNOLOGY FOR 5G.

O.Z. MUSTAPHA

PhD

2019

INTELLIGENT BASED PACKET SCHEDULING SCHEME USING INTERNET PROTOCOL/MULTI-PROTOCOL LABEL SWITCHING (IP/MPLS) TECHNOLOGY FOR 5G.

DESIGN AND INVESTIGATION OF BANDWIDTH MANAGEMENT TECHNIQUE FOR SERVICE-AWARE TRAFFIC ENGINEERING USING INTERNET PROTOCOL/ MULTI-PROTOCOL LABEL SWITCHING (IP/MPLS) FOR 5G.

Oba Zubair MUSTAPHA

Submitted for the Degree of

Doctor of Philosophy

Faculty of Engineering and Informatics

University of Bradford

2019

ABSTRACT

Oba Zubair Mustapha

Intelligent based Packet Scheduling Scheme using Internet Protocol/Multi-Protocol Label Switching (IP/MPLS) Technology for 5G.

Design and Investigation of Bandwidth Management Technique for Service-Aware Traffic Engineering using Internet Protocol/ Multi-Protocol Label Switching (IP/MPLS) for 5G

Keywords: Bandwidth Management, Packet Scheduling, Internet Protocol (IP), Multi-Protocol Label Switching (MPLS), Fuzzy Algorithm, Neuro-Fuzzy Algorithm, Quality of Service, Packet Processing Algorithm (PPA), Traffic Engineering (TE), Weighted Fair Queuing (WFQ)

Multi-Protocol Label Switching (MPLS) makes use of traffic engineering (TE) techniques and a variety of protocols to establish pre-determined highly efficient routes in Wide Area Network (WAN). Unlike IP networks in which routing decision has to be made through header analysis on a hop-by-hop basis, MPLS makes use of a short bit sequence that indicates the forwarding equivalence class (FEC) of a packet and utilises a predefined routing table to handle packets of a specific FEC type. Thus header analysis of packets is not required, resulting in lower latency. In addition, packets of similar characteristics can be routed in a consistent manner. For example, packets carrying real-time information can be routed to low latency paths across the networks. Thus the key success to MPLS is to efficiently control and distribute the bandwidth available between applications across the networks.

A lot of research effort on bandwidth management in MPLS networks has already been devoted in the past. However, with the imminent roll out of 5G, MPLS is seen as a key technology for mobile backhaul. To cope with the 5G demands of rich, context aware and multimedia-based user applications, more efficient bandwidth management solutions need to be derived.

This thesis focuses on the design of bandwidth management algorithms, more specifically QoS scheduling, in MPLS network for 5G mobile backhaul. The aim is to ensure the reliability and the speed of packet transfer across the network. As 5G is expected to greatly improve the user experience with innovative and high quality services, users' perceived quality of service (QoS) needs to be taken into account when deriving such bandwidth management solutions. QoS expectation from users are often subjective and vague. Thus this thesis proposes the use of fuzzy logic based solution to provide service-aware and user-centric bandwidth management in order to satisfy requirements imposed by the network and users.

Unfortunately, the disadvantage of fuzzy logic is scalability since dependable fuzzy rules and membership functions increase when the complexity of being modelled increases. To resolve this issue, this thesis proposes the use of neuro-fuzzy to solicit interpretable IF-THEN rules.The algorithms are implemented and tested through NS2 and Matlab simulations. The performance of the algorithms are evaluated and compared with other conventional algorithms in terms of average throughput, delay, reliability, cost, packet loss ratio, and utilization rate.

Simulation results show that the neuro-fuzzy based algorithm perform better than fuzzy and other conventional packet scheduling algorithms using IP and IP over MPLS technologies.

ACKNOWLEDGMENTS

The writing of a thesis for the degree of Doctor of Philosophy is a very lonely experience, yet not possible without personal and practical support from various people. I would like to express my thanks to the people who have dedicated and made themselves available for assistance during my programme.

First and foremost, I would like to express my profound gratitude to the Almighty God for sparing my life to undertake this task. I also wish to express my sincere gratitude to my PhD supervisors, Prof Yim Fun Hu and Prof Raed A. Alhameed for their encouragement, technical advice, moral, and sacrificial support during the duration of this PhD research work, who complemented each other wonderfully well.

I really appreciate the efforts of Prof Ray E. Sheriff, Dr Muhammad Ali, and Dr Felicia for their technical advice during my PhD programme. A well-deserved thanks to all my colleagues in the research group.

A special thank you to Tertiary Education Trust Fund (TETFUND), for sponsoring the first 3 sessions of my PhD.

Above all, I would like to thank all the members of my family for their continuous support, constant love, encouragement, and prayers.

DEDICATION

To omnipresence, omnipotent, omniscience and Almighty Allah

And

To my Parents.

TABLE OF CONTENTS

CONTENTS **PAGE**

1 INTRODUCTION ... 1

1.1 RESEARCH MOTIVATION ... 1
1.2 INTERNET TRAFFIC ANALYSIS ... 2
1.2.1 Data Traffic Forecast in Networking ... 5
1.2.2 Bandwidth Utilization Analysis ... 6
1.3 PROBLEM STATEMENT ... 9
1.4 AIM AND OBJECTIVES ... 10
1.5 CONTRIBUTED WORK AND ACHIEVEMENTS ... 11
1.5.1 Published Work ... 12
1.6 STRUCTURE OF THE THESIS ... 14

2 LITERATURE REVIEW ... 15

2.1 COMMUNICATION SERVICES ... 15
2.1.1 Connection-oriented Packet Switching (Virtual Circuit) ... 16
2.1.2 Connectionless Packet Switching (Datagram) ... 17
2.2 WIRED AND WIRELESS NETWORKING ... 20
2.2.1 Wired Networks ... 20
2.2.2 Wireless Networks ... 22
2.2.3 Comparison between Wired and Wireless Networking ... 26
2.2.4 Type of Wired and Wireless Technologies ... 28
2.3 INTERNET ... 31
2.3.1 Internet Protocol Stack ... 32
2.3.2 Internet Protocol version 4 (IPv4) ... 33
2.3.3 Internet Protocol version 6 (IPv6) ... 38
2.3.4 User Datagram Protocol ... 40
2.3.5 Transmission Control Protocol ... 40
2.3.6 Source Rate Control ... 44
2.4 CHALLENGES OF TODAY'S INTERNET PROTOCOL ... 45
2.4.1 Internet of Things ... 46
2.4.2 Security and Information protection ... 47
2.4.3 Complexity of Infrastructure ... 47

2.4.4 Time and Cost 47
2.4.5 Hardware issues 48
2.5 CHALLENGES AND PROSPECTS OF 5G SYSTEM 48
2.5.1 5G Concepts and Requirements 48
2.6 SUMMARY 53

3 MULTI-PROTOCOL LABEL SWITCHING (MPLS) NETWORK........... 54

3.1 INTRODUCTION 54
3.2 FIXED OR STATIC ROUTING PATH SELECTION 55
3.2.1 MPLS Functions and Components 56
3.2.2 Mathematical Model of MPLS operation 63
3.2.3 MPLS Label Distribution 64
3.2.4 Resource Reservation Protocol (RSVP) 66
3.3 DATA FORWARDING ARCHITECTURES 67
3.3.1 IP and MPLS 67
3.3.2 MPLS and VPLS 69
3.4 MPLS TRAFFIC ENGINEERING (MPLS-TE) 72
3.4.1 Differentiated MPLS Traffic Engineering (DSTE) 74
3.4.2 MPLS-TE Model 76
3.5 HIERARCHICAL MODEL 79
3.5.1 Core Network Model 80
3.5.2 Technical Challenges of IP/MPLS 84
3.6 SUMMARY 86

4 BANDWIDTH MANAGEMENT IN IP/MPLS BASED NETWORK ENVIRONMENT 87

4.1 INTRODUCTION 87
4.2 BANDWIDTH MANAGEMENT TECHNIQUES 87
4.2.1 Bandwidth Management in IP Environment 91
4.2.2 Bandwidth Management in MPLS Environment 92
4.3 PROPOSED BANDWIDTH MANAGEMENT USING MPLS TECHNOLOGY 92
4.3.1 Node and Network Design 93
4.3.2 Design of Dynamic and Static Label Switch Paths 96
4.4 COMMUNICATION SERVICE COMPONENTS 98
4.4.1 Quality of Service Functions 99
4.4.2 Class of Service 106

4.4.3 Service Level Specifications 107
4.5 QOS SERVICE MODELS 108
4.5.1 Traffic Models 108
4.5.2 Integrated Services 110
4.5.3 Differentiated Services 111
4.5.4 Best Effort 115
4.6 CONCLUDING REMARKS 116

5 PROPOSED INTELLIGENT BASED PACKET SCHEDULING FRAMEWORK 118

5.1 INTRODUCTION 118
5.2 PROPERTIES OF PACKET SCHEDULING 119
5.3 PACKET SCHEDULING ALGORITHMS 121
5.3.1 WRR Algorithm 124
5.3.2 SFQ Algorithm 126
5.3.3 WFQ Algorithm 126
5.3.4 PPA Algorithm 128
5.4 FUZZY LOGIC 129
5.4.1 Fuzzy Logic Operation 129
5.4.2 Structure of Fuzzy Logic Controller 131
5.4.3 Defuzzification Methods 135
5.5 PROPOSED FUZZY AND NEURO-FUZZY BASED PACKET SCHEDULING SCHEMES 140
5.5.1 Service-Aware LSP Selection Example 142
5.5.2 Neuro-Fuzzy Packet Scheduling 152
5.6 CONCLUDING REMARKS 162

6 SIMULATION MODEL AND RESULTS 163

6.1 INTRODUCTION 163
6.2 PERFORMANCE METRICS 164
6.3 SIMULATION PLATFORMS 166
6.3.1 OPNET Simulation Platform 166
6.3.2 Network Simulator NS2 Platform 170
6.3.3 MATLAB Platform 171
6.4 SIMULATION PARAMETERS 172
6.5 SIMULATION SCENARIOS 177
6.5.1 Scenario 1: Proposed Bandwidth Management using Multimedia

Services in the MPLS network 177

6.5.2 Scenario 2: Conventional FIFO and WFQ Algorithms in IP/MPLS ... 179

6.5.3 Scenario 3: Proposed Fuzzy based Packet Scheduling Algorithms (2 MB). 179

6.5.4 Scenario 4: Proposed Fuzzy based Packet Scheduling Algorithms (5 MB). 180

6.5.5 Scenario 5: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (2 MB). 180

6.5.6 Scenario 6: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB). 181

6.6 RESULTS EVALUATION 181

6.6.1 Results for Scenario 1: Bandwidth Management using Multimedia Services in the MPLS network. 181

6.6.2 Results for Scenario 2: Conventional FIFO and WFQ Algorithms in IP/MPLS. 189

6.6.3 Results for Scenario 3: Proposed Fuzzy based Packet Scheduling Algorithms (2 MB). 191

6.6.4 Results for Scenario 4: Proposed Fuzzy based Packet Scheduling Algorithms (5 MB). 196

6.6.5 Results for Scenario 5: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (2 MB). 201

6.6.6 Results for Scenario 6: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB). 205

6.7 CONCLUDING REMARKS 211

7 CONCLUSION AND FUTURE WORK 213

7.1 CONCLUSION 213

7.1.1 Summarised key Contributions and Simulation Results Summary ... 214

7.2 FUTURE WORK 215

7.2.1 Improving the Fuzzy based Packet Scheduling Scheme 215

7.2.2 Improving Neuro-Fuzzy based Packet Scheduling Scheme. 215

7.2.3 SD-WAN Technologies 215

7.2.4 Traffic Efficiency in the 5G 216

LIST OF FIGURES

Figure 1-1. Global Internet Video Traffic by type [7]. 3

Figure 1-2. Global IP traffic forecast in Exabyte's per month from Cisco VNI Global traffic [8]. 4

Figure 1-3. Global IP Data Traffic Forecast [9]. 5

Figure 1-4. HTTP Bandwidth Savings in WAN [10]. 6

Figure 1-5. Large Data Centre WAN Link in MPLS Interface [11]. 7

Figure 1-6. The trend of Bandwidth Utilisation [12]. 8

Figure 2-1. Connection-oriented and Connectionless Transmission Services [17]. 16

Figure 2-2. Global IP Traffic, wired and wireless [18] 18

Figure 2-3. Global MPLS IP VPN Services market by service (USD Million), 2012 – 2020 [19]. 19

Figure 2-4. Wired Network Topologies [22] 21

Figure 2-5. Wireless Networks [24]. 23

Figure 2-6. Different wired and wireless technologies [27]. 28

Figure 2-7. Internet TCP/IP based Communications Stack [20]. 32

Figure 2-8. IPv4 Header [33] 34

Figure 2-9. IPv6 Header [34]. 39

Figure 2-10. Technologies of 5G Mobile and Wireless Networks [44]. 51

Figure 3-1. Hierarchical fixed routing path selection methods [52]. 56

Figure 3-2. MPLS Shim Header 57

Figure 3-3. Architecture of IP/MPLS Networks. 58

Figure 3-4. Flow charts of Packet Processing Algorithm in MPLS. 62

Figure 3-5. MPLS VPN Architecture [61]. 71

Figure 3-6. Maximum Allocation Model (MAM) [70]. 74

Figure 3-7. Maximum Allocation with reservation (MAR) [70]. 75

Figure 3-8. Russian Dolls Model (RDM) [70]. 75

Figure 3-9. Average delays of MAM, RDM, and MAR [70]. 76

Figure 3-10. MPLS Traffic Engineering Model. 76

Figure 3-11. Three-Layered Hierarchical Model [75]. 80

Figure 3-12. Operation Network Model [76]. 82

Figure 3-13. CAPEX and OPEX analysis of MPLS technology [78]. 85

Figure 4-1. Proposed node model diagram. 94

Figure 4-2. Peripheral nodes and hub node of an ingress LER 95

Figure 4-3. Source node, peripheral nodes, and processor of a site 96

Figure 4-4. Architecture of MPLS Network mechanism. 97

Figure 4-5. MPLS modules and flow charts of Dynamic and Static LSP models 98

Figure 4-6. Static and Dynamic MPLS LSPs 99

Figure 4-7. Communication Service Components. 100

Figure 4-8. QoS functional block diagram. 100

Figure 5-1. Conventional Packet Scheduling and Proposed Intelligent-based Packet Scheduling. 119

Figure 5-2. First-In-First-Out (FIFO) packet scheduling. 123

Figure 5-3. Priority Queueing (PQ) packet scheduling. 124

Figure 5-4. Round Robin (RR) packet scheduling. 124

Figure 5-5. Packet processing algorithm. 130

Figure 5-6. Fuzzy logic flow diagram. 132

Figure 5-7. Input parameters and Membership functions. 135

Figure 5-8. Bandwidth versus Utilization Rate (UR). 138

Figure 5-9. Bandwidth versus Reliability 138

Figure 5-10. Bandwidth versus Cost 139

Figure 5-11. Bandwidth versus Delay 139

Figure 5-12. Throughput versus Utilization Rate (UR) 139

Figure 5-13. Throughput versus Reliability 140

Figure 5-14. Throughput versus Cost 140

Figure 5-15. Reliability versus Cost 140

Figure 5-16. Proposed Fuzzy based Packet Scheduling framework. 142

Figure 5-17.Implementation of Fuzzy based Packet Scheduling using Matlab. 143

Figure 5-18. Membership functions for (a) Bandwidth, Bw (b) Delay, D (c) Cost, C (d) Reliability, R, (e) Throughput, Thr, (f) Packet loss rate, Plr and (g) Utilisation rate, Ur 146

Figure 5-19. Membership functions for Service, $Serv$ 153

Figure 5-20. Neuro-Fuzzy System Architecture.. 156

Figure 5-21. Implementation of Neuro-Fuzzy based Packet Scheduling in Matlab.. 159

Figure 5-22. Comparison of Training Random variables and FIS output (Neuro-Fuzzy). 160

Figure 5-23. Comparison of Neuro-Fuzzy and Fuzzy Inference system with 400 random input samples 161

Figure 5-24. Probability Distribution of Fuzzy Inference system 161

Figure 5-25. Probability Distribution of Neuro-Fuzzy Inference system 162

Figure 5-26. Training Error of Neuro-Fuzzy Inference system. 162

Figure 6-1. Network model, Node model and Process model of Mobile device 169

Figure 6-2. Processors and Queues modules. 170

Figure 6-3. Procedure of NS2 Implementation 172

Figure 6-4. Architecture of the MPLS network... 174

Figure 6-5. Label Mapping, Label Distribution, and Packet Forwarding. 174

Figure 6-6. Comparison of Throughput sent and received for voice and video traffics.. 183

Figure 6-7. Comparison of Throughput sent and received for voice in both IP and MPLS networks... 184

Figure 6-8. Comparison of both Packet delay variation and End-to-End delay for voice and video traffics... 184

Figure 6-9. Comparison of End-to-End delay for voice in IP and MPLS networks... 185

Figure 6-10. Service utilization versus some sources. 185

Figure 6-11. Percentage of Bandwidth utilization using MPLS LSP 187

Figure 6-12. Circle view of bandwidth distribution. 187

Figure 6-13. Average Throughput for Voice in MPLS................................. 188

Figure 6-14. Average Throughput for Video Conferencing in MPLS. 188

Figure 6-15. Average End-to-End delay for Voice...................................... 189

Figure 6-16. Average End-to-End delay for Video Conferencing................ 189

Figure 6-17. Average Throughput for VoIP in IP/MPLS networks using FIFO and WFQ Algorithms. 190

Figure 6-18. The average end-to-end delay for VoIP. 191

Figure 6-19. Packet Loss Ratio for VoIP... 192

Figure 6-20. Average Throughput of IP/MPLS with baseline FIFO and WFQ for 2 MB. 192

Figure 6-21. Average Throughput of IP/MPLS (Fuzzy) with FIFO and WFQ for 2 MB. 193

Figure 6-22. Average End-to-End delay in IP/MPLS with baseline FIFO and WFQ for 2 MB. 194

Figure 6-23. Average End-to-End delay in IP/MPLS (Fuzzy) with FIFO and WFQ for 2 MB. 194

Figure 6-24. Reliability of the system. 195

Figure 6-25. Effects of the Fuzzy on service utilization. 195

Figure 6-26. Effects of the Fuzzy on average waiting time. 196

Figure 6-27. Link cost for all the schemes. 197

Figure 6-28. Average Throughput of IP/MPLS with baseline FIFO and WFQ for 5 MB. 198

Figure 6-29. Average Throughput of IP/MPLS (Fuzzy) with FIFO and WFQ for 5 MB. 198

Figure 6-30. Average End-to-End delay in IP/MPLS with baseline FIFO and WFQ for 5 MB. 199

Figure 6-31. Average End-to-End delay in IP/MPLS (Fuzzy) with FIFO and WFQ for 5 MB. 199

Figure 6-32. Reliability of the system (5 MB). 200

Figure 6-33. Effects of the Fuzzy on service utilization (5 MB). 201

Figure 6-34. Effects of the Fuzzy on average waiting time (5 MB). 201

Figure 6-35. Link cost for all the schemes (5 MB). 202

Figure 6-36. Average Throughput of IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 2 MB. 203

Figure 6-37. Average End-to-End delay in IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 2 MB. 204

Figure 6-38. Reliability for packet delivery of the schemes (2 MB). 204

Figure 6-39. Service utilization of the schemes (2 MB). 205

Figure 6-40. Link cost of the schemes (2 MB). 206

Figure 6-41. Packet loss ratio of the schemes (2 MB). 206

Figure 6-42. Average Throughput of IP/MPLS (Neuro-Fuzzy) with (FTP) for 5 MB. 207

Figure 6-43. Average End-to-End delay in IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 5 MB. 208

Figure 6-44. Reliability for packet delivery of the schemes (5 MB). 209

Figure 6-45. Service utilization of the schemes (5 MB). 209

Figure 6-46. Link cost of the schemes (5 MB). 210

Figure 6-47. Packet loss ratio of the schemes (5 MB). 211

List of Tables

Table 1. 5G requirements [47] 50

Table 2. Comparison of MPLS with Frame Relay and ATM [53] 81

Table 3. Survey of Routing Algorithms 90

Table 4. Network QoS requirement for multimedia services [105]........................... 106

Table 5. Application specific QoS requirements [106].. 109

Table 6. PHB and DSCP [107]... 113

Table 7. Mapping 6-bit DSCP field into 3bits experimental (EXP) field in MPLS [108].

....114

Table 8. Differences between the three QoS models... 117

Table 9. Degree of Membership Function for LSP1. 146

Table 10. Degree of Membership Function for LSP2. 146

Table 11. Result of IF-ELSE rule for LSP1 149

Table 12. Result of IF-ELSE rule for LSP2 149

Table 13. Fuzzy Rule Base for Fuzzy Controller. 150

Table 14. Simulation Parameters for Multimedia Services 170

Table 15. Relationship between R-values and MOS... 171

Table 16. Simulation platform specifications 173

Table 17. Simulation platform specifications for Voice... 176

Table 18. Simulation platform specifications for File Transfer Protocol (FTP) 177

Table 19. Neuro-Fuzzy Inference System Information.. 177

Table 20. Performance Analysis of FIFO using FTP in MPLS................................ 211

Table 21. Performance Analysis of WFQ using FTP in MPLS................................ 211

LIST OF ACRONYMS

4G	Fourth Gneration
5G	Fifth Generation
ACK	Acknowledge
AIAD	Additive Increase Additive Decrease
AIMD	Additive Increase Multiplicative Decrease
AN	Access Node
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
API	Application Programming Interface
AR	Augment Reality
ARPANET	Advance Research Project Agency Network
AS	Autonomous System
ATM	Asynchronous Transfer Mode
ATMoMPLS	Asynchronous Transfer Mode over Multi-Protocol Label Switching
BB	Backbone node
BGP	Border Gateway Protocol
CAPEX	Capital Expenditure
CE	Customer Edge router
CDMA	Code Division Multiple Access
CoS	Class of Service
CR-LDP	Constraint Routing Label Distribution Protocol
D2D	Device to Device
DBA	Dynamic Bandwidth Allocation
DiffServ	Differentiated Service
DSCP	Differentiated Services Code Point
DSTE	Differentiated Service Traffic Engineering
ECMP	Equal Cost Multi-Path
ECN	Explicit Congestion Notification
EIGRP	Enhance Internet Gateway Routing Protocol
EoMPLS	Ethernet over Multi-Protocol Label Switching

FDDI	Fibre Distributed Data Interface
FEC	Forwarding Equivalence Class
FCC	Federal Communications Commission
FIFO	First In First Out
FFD	First Fit Decreasing
FFT	Fast Fourier Transform
FR	Frame Relay
FRoMPLS	Frame Relay over Multi-Protocol Label Switching
FTP	File Transfer Protocol
GGSN	Gateway GPRS Serving Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HD	High Definition
HLR	Home Location Register
HTTP	Hypertext Transfer Protocol
IHL	Internet Header Length
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IGP	Internet Gateway Protocol
IMT	International Mobile Telecommunications
INSPs	Internet Network Service Providers
IntServ	Integrated Service
IoT	Internet of Things
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISPs	Internet Service Providers
IS-IS	Intermediate System-Intermediate System
IT	Information Technology
ITU	International Telecommunication Union

LAN	Local Area Network
LDP	Label Distribution Protocol
LER	Label Edge Router
LISP	Locator/Identifier Separation Protocol
LoS	Line of Sight
LSP	Label Switch Path
LSR	Label Switched Router
M2M	Machine to Machine
MATLAB	Matrix Laboratory
MAC	Medium Access Control
MAM	Maximum Allocation Model
MAR	Maximum Allocation with Reservation
MIAD	Multiplicative Increase Additive Decrease
MIMD	Multiplicative Increase Multiplicative Decrease
mmW	Millimetre Waves
MNs	Moving Networks
MPLS	Multi-protocol Label Switching
MSS	Maximum Segment Size
MTBF	Mean Time Between Failure
MTTR	Mean Time to Restore
NAT	Network Address Translation
NGN	Next Generation Networks
NOC	Network Operating Centre
NRT	Non- Real-Time service
NS2	Network Simulator 2
O&M	Operation and Management
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
OPNET	Optimised Network Evaluation Tool
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First

PE	Provider Edge router
PHP	Penultimate Hop Popping
POP	Post Office Protocol
PQ	Priority Queueing
QoE	Quality of Experience
QoS	Quality of Service
RDM	Russian Dolls Model
RFC	Request for Comments
RTT	Round-Trip Time
RF	Radio Frequency
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RSVP	Resource Reservation Protocol
RSVP-TE	Resource Reservation Protocol Traffic Engineering
RT	Real Time Service
SDN	Software Defined Networks
SFQ	Stochastic Fair Queueing
SGSN	Serving GPRS Support Nod
SLA	Service Level Agreements
SMTP	Simple Mail Transfer Protocol
SON	Self-Organizing Networks
SSL	Secure Sockets Layer
T1/E1	Digital Link of 1.544Mbps/2.048Mbps
TDM	Time Division Multiplexing
TCO	Total Cost of Ownership
TCP	Transmission Control Protocol
TE	Traffic Engineering
ToS	Type of Service
TTL	Time To Live

UDP	User Datagram Protocol
UL/DL	Uplink/Downlink
UMTS	Universal Mobile Telecommunications System
URCs	Ultra-Reliable Communications
VLLs	Virtual Leased Lines
VM	Virtual Machine
VoD	Video on Demand
VoIP	Voice over Internet Protocol
VPLS	Virtual Private LAN Services
VPN	Virtual Private Network
VR	Virtual Reality
WAN	Wide Area Network
WEP	Wired Equivalent Privacy
WFQ	Weighted Fair Queueing
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Networks
WRR	Weighted Round Robin
WWW	World Wide Web

1 Introduction

1.1 Research Motivation

The internet continues to grow at a phenomenal rate. This is reflected in the tremendous popularity of the World Wide Web (WWW), the opportunities that businesses in reaching customers from virtual storefronts, and the emergence of new ways of doing business. Expanding the business and public awareness will continue to increase demand for access to resources on the internet.

In an accelerating world, extensive use of networks has led to a significant increase in the number of Internet users, which causes huge resource pressure on networks and internet services. Moreover as stated in [1], it affects the performance negatively, reduces the available quality of service, and perhaps threaten its performance seriously. With all these challenges and with the increasing popularity of real-time and multimedia applications, such as voice and video, traditional IP-based networks relying on hop-by-hop header inspection for packet forwarding will not be able to satisfy the stringent network performance requirements such as short delay, high bandwidth, etc. imposed by these applications and will not be able to guarantee the quality of services (QoS) [2]. Alternative approaches such as Multi-Protocol Label Switching (MPLS) and so on are now used. MPLS is a fast-growing technology, which plays a vital role in providing quality of service (QoS). It uses a variety of protocols and careful traffic engineering to establish Label Switched Paths (LSP) across the network. MPLS uses the information contained in the labels, which are attached to Internet Protocol (IP) packets to improve the fast forwarding of these packets [3]. It provides scalability as well as congestion control in order to overcome limitations such as high packet loss and excessive delays in the network [4, 5].

Packet Switching is a method of transferring the data to a network in the form of packets.The need to transfer the file fast and efficient manner over the network and minimize the transmission latency, the data is broken into small pieces of variable length, called Packet. At the destination, all these small-parts (packets) should be reassembled, belonging to the same file. A packet composes of payload and various control information [6]. It uses Store and Forward technique while switching the packets; while forwarding the packet each hop first store that packet than forward. This technique is very beneficial because packets may get discarded at any hop due to some reason. More than one path is possible between a pair of source and destinations. Each packet header contains source and destination addresses to identify the path for the packet to travel through the network. Thus packets belonging to the same file may or may not travel through the same path. If there is congestion at some path, packets can choose different path possible over an existing network. Packet-switched networks were designed to overcome the weaknesses of circuit-switched networks since circuit-switched networks is connection-oriented which is not bandwidth effective for data communications especially small messages [6].

1.2 Internet Traffic Analysis

Overall IP traffic will increase by nearly three times by 2021 to reach 35 GB per person per month, up from 13 GB last year. Smartphones will account for 33% of total IP traffic in 2021, up from 13% in 2016 as shown in Figure 1-1. PCs accounted for 46% of total IP traffic last year; this will decrease to 25%

by 2021. The video will represent 82% of all IP traffic in 2021, including consumer and business IP traffic, amounting to a million minutes of video transmitted through networks every second. Internet video, IP video-on-demand (VoD), videos exchanged through file sharing, video-streamed gaming, and video conferencing are factored into this estimate. Last year, video accounted for 73% of global IP traffic.

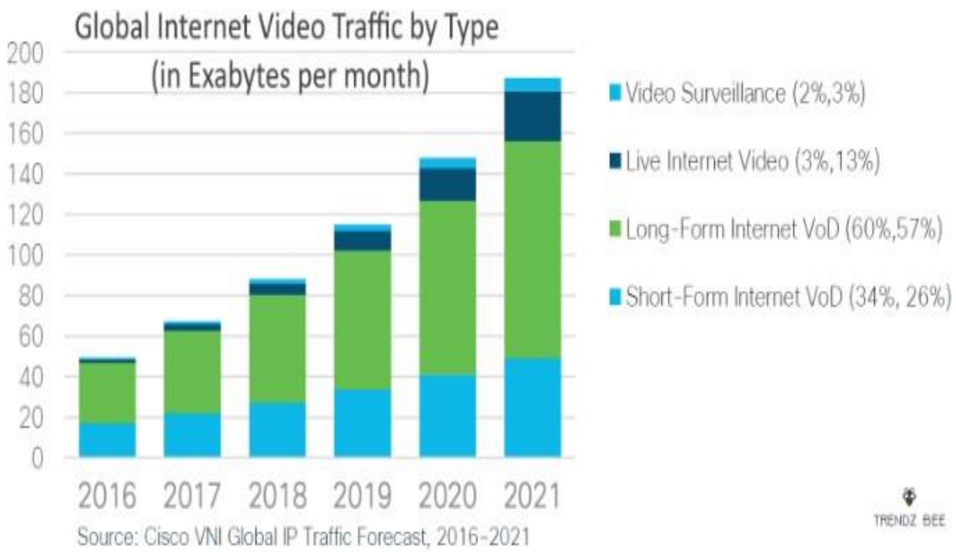


Figure 1-1. Global Internet Video Traffic by type [7].

Specific video formats are poised for growth. Live video is expected to grow 15-fold, from 3% of internet video traffic to 13% by 2021. Virtual reality (VR) and augmented reality (AR) will increase by 20 times between 2016 and 2021 globally, at a compound annual growth rate of 82% [7]. Stemming largely from downloads of VR content files and apps, and an uptick VR streaming would push Cisco's prediction even higher. Internet video delivered to TV is taking hold. This category grew by 50% in 2016 and will grow 3.6-fold by 2021. By 2021, internet video on TV will be 26% of fixed consumer internet video traffic.

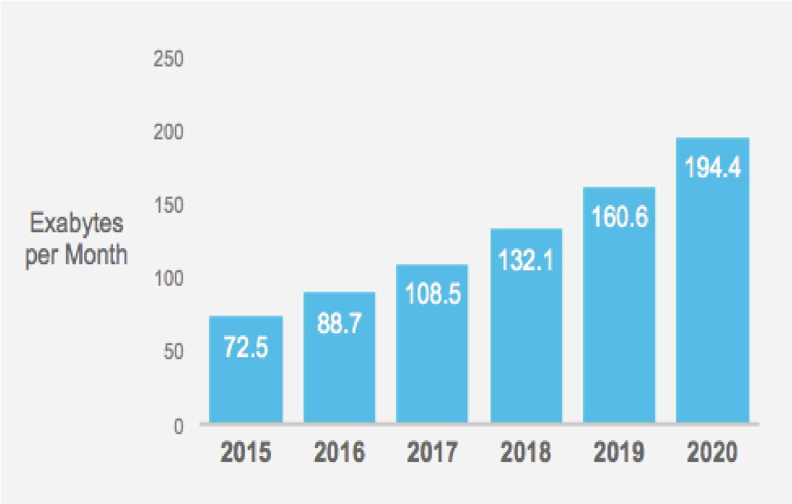


Figure 1-2. Global IP traffic forecast in Exabyte's per month from Cisco VNI Global traffic [8].

Internet video will account for 79% of global internet traffic by 2020, up from 63% in 2015, according to new research from Cisco. The Cisco Visual Networking Index claims that internet video will increase four-fold between 2015 and 2020 and that HD and Ultra HD Internet video will make up 82% of this internet video traffic by 2020 – up from 53% as evidenced in Figure 1-2.

"The world will reach three trillion Internet video minutes per month by 2020, which is five million years of video per month, or about one million video minutes every second," according to the forecast as evidenced [8] in Figure 1-2.

Breaking down video traffic by usage, Cisco claims that consumer internet video traffic will be 82% of consumer internet traffic by 2020, up from 68% in 2015, while business internet video traffic will be 66% of business internet traffic by 2020, up from 44% in 2015 [8].

1.2.1 Data Traffic Forecast in Networking

According to the Global IP data traffic forecast, the statistic gives information on the global IP data traffic from 2016 to 2021 is shown in Figure 1-3. In 2017, global IP data traffic is expected to amount to 121,694 petabytes per month. In 2021, IP data traffic worldwide is expected to reach 278,108 petabytes per month. This shows that there is an approximate of 27% increment from 2016 to 2017 while the percentage increase is about 24% from 2017 to 2018. Therefore, it is observed that the difference between IP data traffic from 2016 to 2021 would be 156414 petabytes per month with about 56% [9].

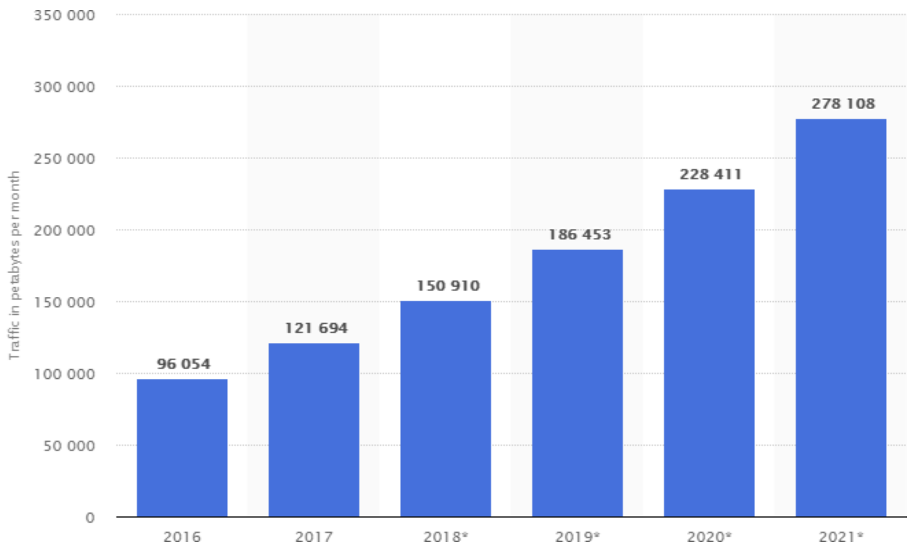


Figure 1-3. Global IP Data Traffic Forecast [9].

It is important to understand the internet traffic to find solutions to any bandwidth consumption issues. In [10], Many studies have been carried out for characterizing the internet traffic of the core network as well as distribution and access networks.

Bandwidth Savings chart in Figure 1-4 displays how much traffic, by percentage, is served by the Cache Engine (CE) that did not have to be fetched from the source. When this information is combined with overall incoming traffic into the router from the WAN, it indicates how effective the cache is in boosting the WAN performance in terms of request-response latency. The combination of the incoming (WAN) traffic flow to the router, plus the WAN data offload incoming traffic provides a truer measure of the traffic flow the router's clients (in aggregate) experience.

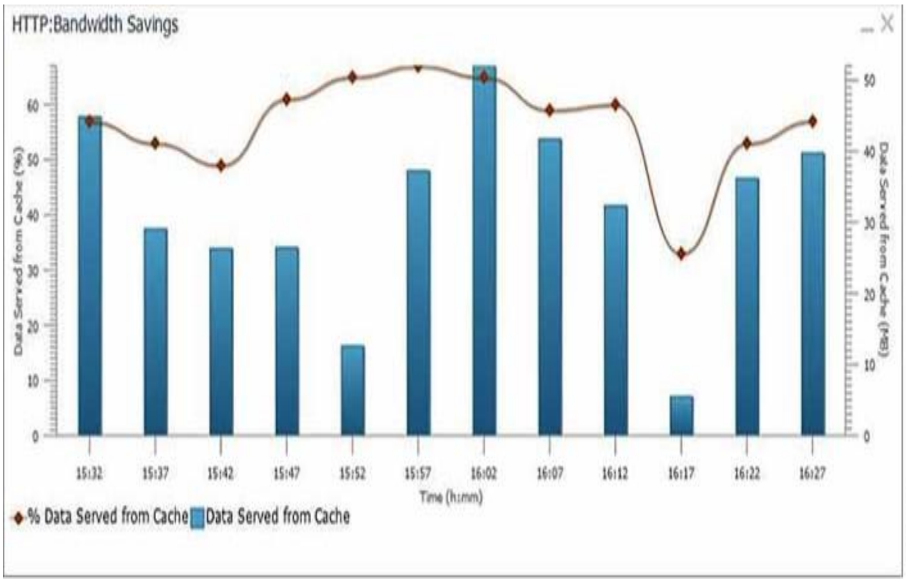


Figure 1-4. HTTP Bandwidth Savings in WAN [10].

1.2.2 Bandwidth Utilization Analysis

The report for the WAN link as shown in Figure 1-5 is much more informative and provides a wide range of statistical values including peak and average values, and average total bandwidth utilization. These offer useful information about typical usage levels or patterns across the business day and despite the obviously excessive, recurring out-bound peaks. The values indicated for the average-out value is only 255 Mbps [11].

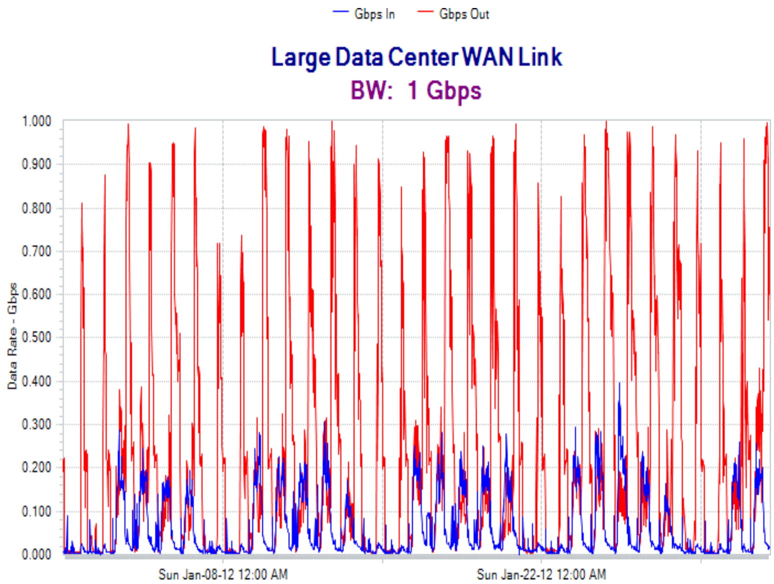


Figure 1-5. Large Data Centre WAN Link in MPLS Interface [11].

It can be noticed that the traffic after 22nd of January 2012 had 0.395 Gbps in as the maximum value with 39.54 % bandwidth utilization. For example, the business areas may be heavily loaded during daytimes but only lightly loaded at night times. Additionally, the traffic profile during weekends/holidays, even during peak hours, is much lower than that of a normal weekday. Therefore, weekdays and weekends appear to show distinct trends as shown in Figure 1-5.

1.2.2.1 Bandwidth Utilization

The bandwidth of a device on a network is the maximum amount of data that the device can take in from the network or give out into the network at a given instant. Bandwidth is efficiently allocated and prioritized according to mission-critical needs (latency, jitter, and packet loss minimum thresholds). The Network Interface Card (NIC) in the device determines this maximum limit. The amount of data transferred across a device's network card, at a given instant, is the actual bandwidth utilized by the device at that instant. The chart in Figure 1-6 shows the inbound and outbound traffic for the interface measured over time [12]. The minimum bandwidth utilization is 0% at the early time while maximum bandwidth utilization is 7% at peak time during the day.

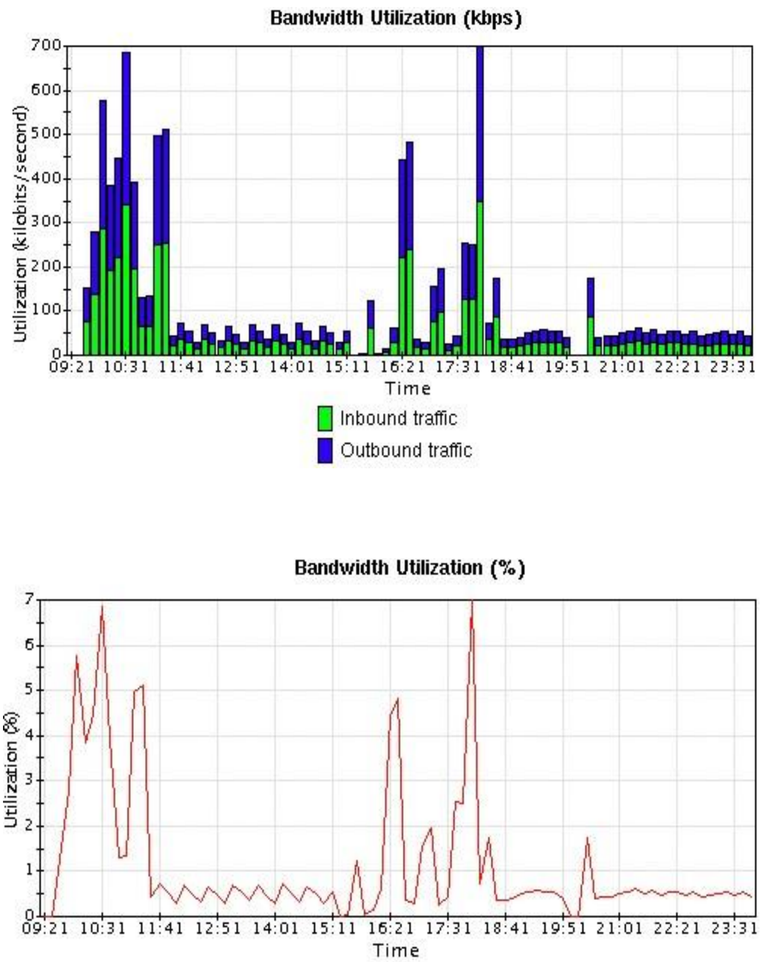


Figure 1-6. The trend of Bandwidth Utilisation [12].

However, consider the different time zones around the globe, the traffic intensity in an IP network can be quite diverse over time and space. Such diversity in traffic intensity may result in underutilization of network resources during a particular period of time resulting in both system and network utilisation inefficiencies. However, traffic dynamics can provide significant opportunities for bandwidth savings. For example, if the traffic variation can be traced and the resource allocation strategy for individual or the whole network is adopted accordingly, a significant amount of bandwidth could be saved. From the above discussion, it can be concluded that the variation of traffic intensity in IP networks shows significant underutilization of system capacity given the network being designed based on the peak-traffic scenarios.

1.3 Problem Statement

5G promises to deliver ultra-fast speeds and responsiveness to connect everything without interruption, with peak speed 20 times faster than 4G. However, with the imminent roll out of 5G, MPLS is seen as a key technology for mobile backhaul to ensure the reliability of the communication minimizing the delays and enhancing the speed of packet transfer through carefully designed traffic engineering (TE) solutions to handle the congestion and to improve throughput.

A lot of research effort on bandwidth management in MPLS networks has already been devoted in the past. To cope with the 5G demands of rich, context aware and multimedia-based user applications, more efficient bandwidth management solutions need to be derived. As 5G is expected to greatly improve the user experience with innovative and high quality services, users' perceived quality of service (QoS) needs to be taken into account when deriving such bandwidth management solutions. However, QoS expectation from users are often subjective and vague. It is imperative that users' expected QOS can be interpreted as accurately as possible when designing such bandwidth management solutions. This thesis addresses the problem of how to control and distribute bandwidth available to forward packets from different applications across the network and at the same time satisfying both network and users QoS requirements.

1.4 Aim and objectives

Bandwidth management can be described as the various techniques, technologies, tools and policies employed by an organization to enable the most efficient use of its bandwidth resources [13]. In [14], bandwidth management is a process of allocating bandwidth resources to critical applications on a network. The critical function of bandwidth management is to control the flow of packets on a network link to avoid traffic exceeding the capacity of the network, which would lead to congestion.

In order to address the increasing demand for data capacity expected to increase a thousand fold by 2020 [15], this thesis aims to design bandwidth management algorithm and more specifically bandwidth allocation for packet scheduling in MPLS-based mobile backhaul networks for 5G with a view to minimize packets loss, delay variation, and at the same time utilizing the available bandwidth resource in an efficient manner while maintaining QoS. The objectives of this project are mentioned as follows:

1. To design suitable link resource allocation mechanism to improve link utilisation, minimise packet loss and delay for packet scheduling in 5G MPLS-based backhaul network, taking into account both applications and users QoS requirements and network bandwidth constraints.

2. To define suitable scheduling techniques for real-time and non real-time applications.
3. To design the project in NS2, then to simulate the adopted intelligent based Packet Scheduling algorithm in MATLAB;
4. To review and conclude the performance outcomes that were achieved in the NS2 simulation that is the end-to-end delay and accuracy performance of the intelligent based packet scheduling algorithm that was employed in MATLAB.

The ultimate goal is ensure reliable, scalable and efficient packet transfer over the network and at the same to guarantee QoS support.

1.5 Contributed Work and Achievements

The main contributions of this thesis include:

1. A bandwidth management algorithms combining WFQ and fuzzy logic for real-time and non real-time packet scheduling in MPLS networks, taking into the application and users QoS requirements and the network resource constraints.
2. An intelligent neuro-fuzzy bandwidth management algorithm to enhance the scalability of the fuzzy logic based approach by solicit interpretation of the fuzzy rules.
3. The consideration of static and dynamic LSPs to enhance link utilisation adaptive to current state of network resource.
4. Performance comparison of the fuzzy and neuro-fuzzy bandwidth allocation for packet scheduling algorithms derived in this PhD work with

conventional scheduling techniques such as First-In First-Out (FIFO) and Weighted Fair Queuing (WFQ).

1.5.1 Published Work

The following materials have been published taking material from the thesis, which forms the core contribution of this thesis.

Accepted Journal

1. **Oba Zubair Mustapha**, Yim Fun Hu, Raed A. Abd-Alhameed, M. Ali, and R.E.Sheriff " Adoption of Bandwidth Management Technique using Dynamic LSP Tunneling and LDP in MPLS for sustainable Mobile Wireless Networks, International Journal of Computing and Digital Systems, 2019**.**

Published Conference Papers

- **Oba Zubair Mustapha**, Muhammad Ali, Yim Fun Hu, and Raed A. Abd-Alhameed," Fuzzy-based Packet Scheduling Scheme using Non-Real time Traffic in IP/MPLS," in *IEEE International Conference on Future Internet of Things and Cloud (FiCloud 2019),* Istanbul, Turkey, 2019.
- **Oba Zubair Mustapha**, Yim Fun Hu, Ray E. Sheriff, Raed A. Abd-Alhameed, M. Ali and T. Sigwele, " Adoption of Bandwidth Management Technique using Dynamic LSP Tunnelling and LDP in MPLS for sustainable Mobile Wireless Networks," in *IEEE International Innovative Computing Technology (INTECH 2018),* Luton, UK, 2018.
- **Oba Zubair Mustapha**, Yim Fun Hu, and Raed A. Abd-Alhameed," Approach to Label Distribution Protocol Signalling using Multimedia Services for Bandwidth Allocation," in *IEEE International Conference on Modelling & Simulation (UKSim 2018),* Cambridge, UK, 2018.

- **Oba Zubair Mustapha**, Ray E. Sheriff, and Felicia L.C. Ong, " Bandwidth Management using MPLS Model for Future Mobile Wireless Networks," in *International Conference on Wireless and Satellite Systems (WiSATS 2017)*, Oxford, UK, 2017.

Unpublished Conference Papers

- **Oba Zubair Mustapha,** Ray E. Sheriff, and Felicia L.C. Ong, "Resource Reservation Protocol Tunnelling Extension in MPLS for sustainable Mobile Wireless Networks," in *Annual Innovative Engineering Research Conference, Faculty of Engineering and Informatics*, University of Bradford, UK, 2017.

Chapter

- **Oba Zubair Mustapha**, Ray E. Sheriff, and Felicia L.C. Ong, " Bandwidth Management using MPLS Model for Future Mobile Wireless Networks," in *Wireless and Satellite Systems*. vol. 231, P. Pillai, K. Sithamparanathan, G. Giambene, M. Ángel Vázquez and P. D. Mitchell (Eds.), ed: Springer International Publishing, pp. 62-72, 2018.

Poster Presentation

- **Oba Zubair Mustapha,** Ray E. Sheriff, and Felicia L.C. Ong, " Modelling and Simulation of MPLS using Multimedia Services for Future Mobile Wireless Networks," *Poster Presentation at Faculty of Engineering and Informatics Workshop*, University of Bradford, UK, 2016.

1.6 **Structure of the Thesis**

This thesis is divided into seven chapters. Chapter 1 presents a brief introduction of the research topic, states the overarching aim and objectives of the work and also highlights the problem statement. Chapter 2 begins by presenting the current state of art including the concepts of wired and wireless networking, and then introduces the challenges of today's Internet Protocol and 5G system with their prospects. A literature review of MPLS and Traffic Engineering is discussed in Chapter 3. Chapter 4 presents a more detailed review of the proposed bandwidth management using MPLS technology. The mathematical model for the operation and MPLS mechanism are present here. In chapter 5, the existing packet scheduling algorithms and proposed intelligent based packet scheduling scheme in IP/MPLS are presented. Chapter 6 then discusses the simulation platforms used and the various simulation scenarios. This chapter also presents the results obtained for these different scenarios to critically evaluate the performance of the proposed framework. Finally, Chapter 7 presents the thesis conclusion which summarises the contributions made by this research and also discusses some recommendations for future work and development.

2 Literature Review

In this chapter, an overview of the technology used to develop the network of an Intelligent Network Service Providers will be discussed. It will present various challenges faced by today's packet forwarding services ranging from layer 2.5 to layer 3 in an Internet Protocol network. The details of network architectures such as IP and Multi-Protocol Label Switching data forwarding architecture will be strictly described including the motivation and their technologies.

2.1 Communication Services

Communication between various network components requires a common understanding between sender and receiver regarding the transmission procedures, the types of signals carrying the useful information, and the format of the data transmitted. There are four major aspects involved in communications services: communication interfaces, communication protocols, layered communication stacks, and information models.

Communication interfaces are connection points between network components designed according to standard or proprietary specifications. An interface is defined by mechanical, electrical, and functional characteristics that allow communication to take place between adjacent network devices. Communication protocols are formal descriptions of data unit formats and transmission rules for message exchange between network entities. These are organized in multiple layered communication stacks where each layer provides services to the layer above. The information models are collections of abstracted managed object definitions and attributes for devices having common characteristics. Currently, there are many communications architectures and protocols in use.

Internet Protocol (IP) is a connectionless datagram-based network layer protocol that performs addressing, routing, and control functions for transmitting and receiving packets. As packets are received by the router, IP addressing information, such as the destination address, is used to determine the best "next hop" the packet should take reroute to its final destination [16]. Figure 2-1 shows the diagram used to represent connection-oriented and connectionless Transmission Services.

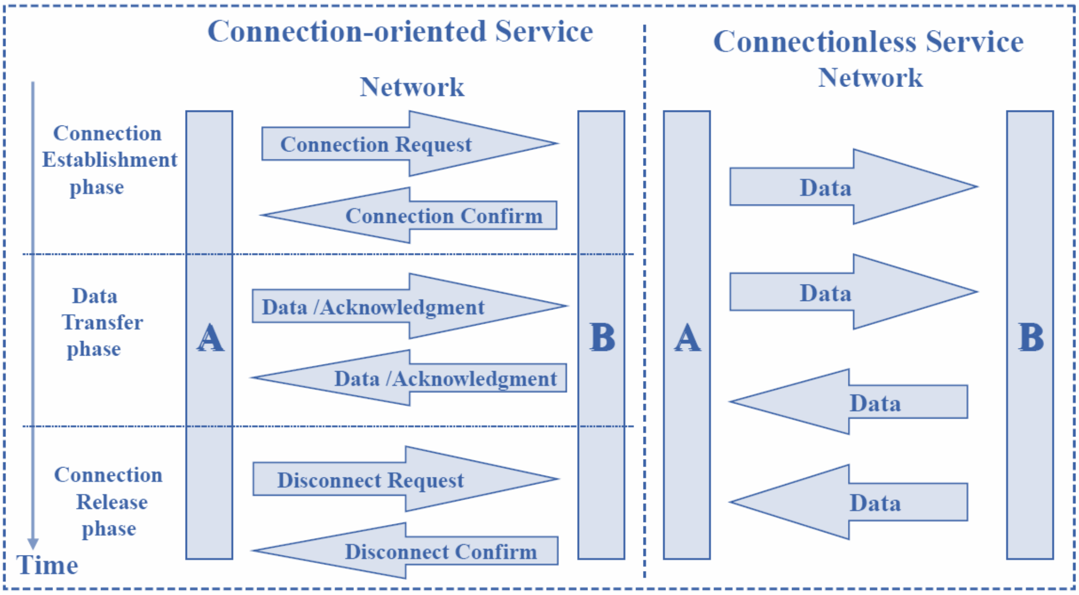


Figure 2-1. Connection-oriented and Connectionless Transmission Services [17].

2.1.1 Connection-oriented Packet Switching (Virtual Circuit)

Before starting the transmission, it establishes a logical path or virtual connection using signaling protocol, between sender and receiver and all packets belongs to this flow will follow this predefined route. Virtual Circuit ID is provided by switches/routers to uniquely identify this virtual connection. Data is divided into small units and all these small units are appended with the help of sequence number. Therefore, connection-oriented service involved three phases, namely: connection establishment phase, data transfer phase, and connection tear-down.

2.1.2 Connectionless Packet Switching (Datagram)

Unlike Connection-oriented packet switching, in which each packet contains all necessary addressing information such as source address, a destination address, and port numbers, etc. In Datagram Packet Switching, each packet is treated independently. Packets belonging to one flow may take different routes because routing decisions are made dynamically, so the packets arrived at the destination might be out of order. It has no connection setup and teardown phase, like Virtual Circuits. However, data transfer forms the only phase exhibited by this type of network.

In, the rapid growth of mobile data traffic has been widely recognized and reported. The trend toward mobility carries over into the realm of fixed networks as well, in that an increasing portion of traffic will originate from portable or mobile devices. It shows the growth in Wi-Fi and mobile traffic in relation to traffic from wired devices. By 2021, wired networks will account for 37 percent of IP traffic, and Wi-Fi and mobile networks will account for 63 percent of IP traffic. In 2016, wired networks accounted for the majority of IP traffic, at 51 percent; Wi-Fi accounted for 41 percent; and mobile or cellular networks accounted for 7.5 percent of total global IP traffic [18].

Video is the underlying reason for accelerated busy hour traffic growth. Unlike other forms of traffic, which are spread evenly throughout the day (such as web browsing and file sharing), video tends to have “prime time.” Because of video consumption patterns, the Internet now has a much busier busy hour. Because video has a higher peak-to-average ratio than data or file sharing, and because the video is gaining traffic share, peak Internet traffic will grow faster than average traffic. The growing gap between peak and average traffic is amplified further by the changing composition of Internet video. Real-time video such as live video, ambient video, and video calling has a peak-to-average ratio that is higher than on-demand video. The global IP Traffic of wired and wireless is shown in Figure 2-2.

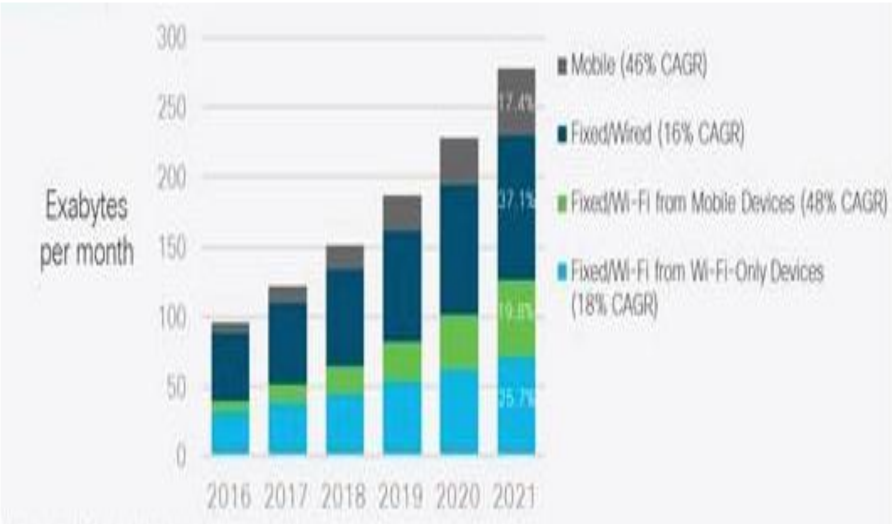


Figure 2-2. Global IP Traffic, wired and wireless [18].

The IP/MPLS Virtual Private Network (VPN) services market is poised for high growth as several enterprises are expected to incorporate these networks to harness various benefits offered by them such as performance maximization and cost minimization. Even though the VPN market has considerably matured over the last five years as most of enterprises have implemented some of the other form of VPN, managed network-based IP/MPLS VPN services that exhibit appreciable growth potential.

The declining ATM/frame relay market is expected to buoy the IP/MPLS VPN Services market growth throughout the forecast period. Enterprises migrating from a packet have implemented layer 2 and layer 3 IP/MPLS VPN Services, with the latter generating a major proportion of the revenue as shown in Figure 2-3. IP/MPLS VPN Services helps enterprises in prioritization of applications such as Voice over Internet Protocol (VoIP) by Class of Service (CoS), creation of disaster recovery infrastructures and reducing complexity to simplify network management. Ability to provide scalable bandwidth and convergence of voice, data, and video from multiple platforms on a single platform is expected to encourage the incorporation of IP/MPLS VPN Services. The technology considerably helps meet increasing traffic requirements where additional capacity is required.

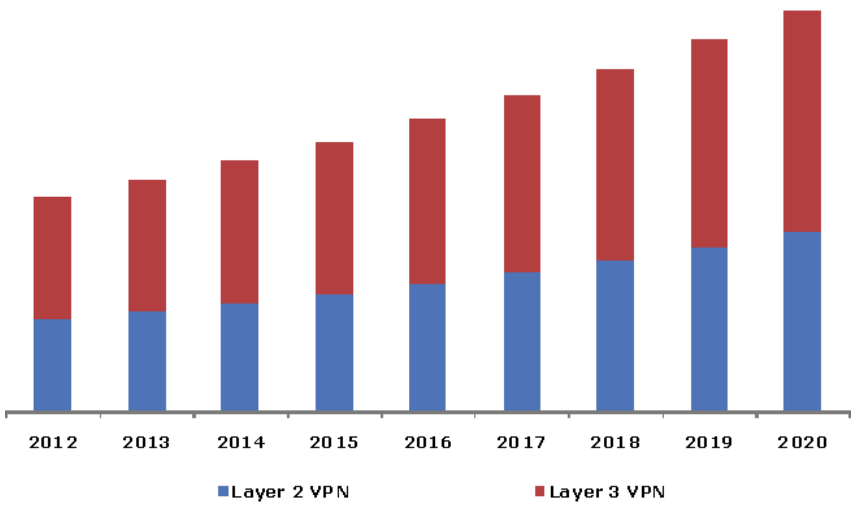


Figure 2-3. Global MPLS IP VPN Services market by service (USD Million), 2012 – 2020 [19].

2.2 Wired and Wireless Networking

2.2.1 Wired Networks

Wired networks, also called Ethernet networks, are the most common type of local area network (LAN) technology. A wired network is simply a collection of two or more computers, printers, and other devices linked by Ethernet cables. Ethernet is the fastest wired network protocol, with connection speeds of 10 megabits per second (Mbps) to 100 Mbps or higher. Wired networks can also be used as part of other wired and wireless networks. To connect a computer to a network with an Ethernet cable, the computer must have an Ethernet adapter (sometimes called a network interface card, or NIC). Ethernet adapters can be internal (installed in a computer) or external (housed in a separate case). Some computers include a built-in Ethernet adapter port, which eliminates the need for a separate adapter (Microsoft) [20, 21]. Four basic network topologies are most commonly used today as shown in Figure 2-4.

The star network, a general more simplistic type of topology, has one central hub that connects to three or more computers and the ability to network printers. This type can be used for small businesses and even home networks. The star network is very useful for applications where some processing must be centralized and some must be performed locally. The major disadvantage is the star network is its vulnerability. All data must pass through one central host computer and if that host fails, the entire network will fail [20].

On the other hand, the bus network has no central computer and all computers are linked to a single circuit. This type broadcasts signals in all directions and it uses special software to identify which computer gets what signal. One disadvantage with this type of network is that only one signal can be sent at one time, if two signals are sent at the same time they will collide and the signal will fail to reach its destination [20]. One advantage is that there is no central computer so if one computer goes down others will not be affected and will be able to send messages to one another [20, 21].

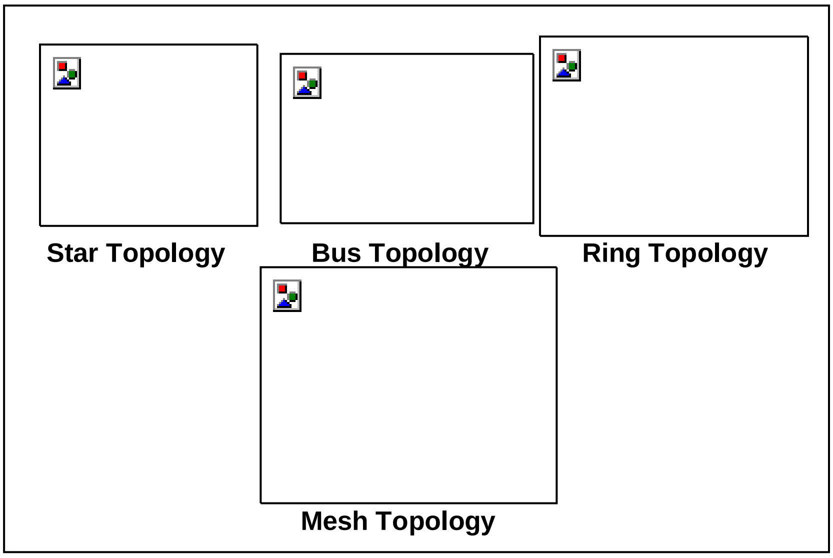


Figure 2-4. Wired Network Topologies [22]

The third type of network is the ring network. Similar to the bus network, the ring network does not rely on a central host computer either. Each computer in the network can communicate directly with any other computer, and each process its own applications independently. A ring network forms a closed loop, data is sent in one direction only and if a computer in the network fails, the data can still be transmitted.

Finally, mesh network is a local network topology in which the infrastructure nodes (i.e. bridges, switches, and other infrastructure devices) connect directly, dynamically and non-hierarchically to as many other nodes as possible and cooperate with one another to efficiently route data from/to clients. This lack of dependency on one node allows for every node to participate in the relay of information. Mesh networks dynamically self-organize and self-configure, which can reduce installation overhead. The ability to self-configure enables dynamic distribution of workloads, particularly if a few nodes should fail. This, in turn, contributes to fault-tolerance and reduced maintenance costs. It may be contrasted with conventional star/tree local network topologies in which the bridges/switches are directly linked to only a small subset of other bridges/switches, and the links between these infrastructure neighbours are hierarchical.

Typically, the range of a wired network is within a 2,000-foot-radius. The disadvantage of this is that data transmission over this distance may be slow or non-existent. The benefit of a wired network is that bandwidth is very high and that interference is very limited through direct connections. Wired networks are more secure and can be used in many situations; corporate LANs, school networks and hospitals. The biggest drawback to this type of network is that it must be rewired every time it is moved [21].

2.2.2 Wireless Networks

Wireless communication has become a ubiquitous part of modern life, from global cellular communication systems to local and even personal area networks. Even to the most casual observer, it is apparent that a veritable revolution in telecommunications has taken place within recent years. The use of wireless communications has expanded dramatically globally, as more and more users are using data applications [23]. The past years have experienced phenomenal growth in the wireless industry, both in terms of mobile technology and its subscribers. There has been a significant shift from fixed to mobile cellular telephony, especially since the turn of the century. The cellular concept was a breakthrough in solving the problem of spectral congestion and user capacity. It is a system-level idea which calls for replacing a single, high power transmitter (covering a large area) with many low power transmitters covering small geographic areas called cells that are represented as a hexagon. Each cell is served by a BS providing coverage to only a small portion of the service area.

A wireless network, which uses high-frequency radio waves rather than wires to communicate between nodes, is another option for home or business networking. Individuals and organizations can use this option to expand their existing wired network or to move completely for wireless. Wireless allows devices to be shared without networking cable, which increases mobility but decreases range. The wireless network is illustrated in Figure 2-5 and there are two main types of wireless networking; peer-to-peer or ad-hoc and infrastructure [23].

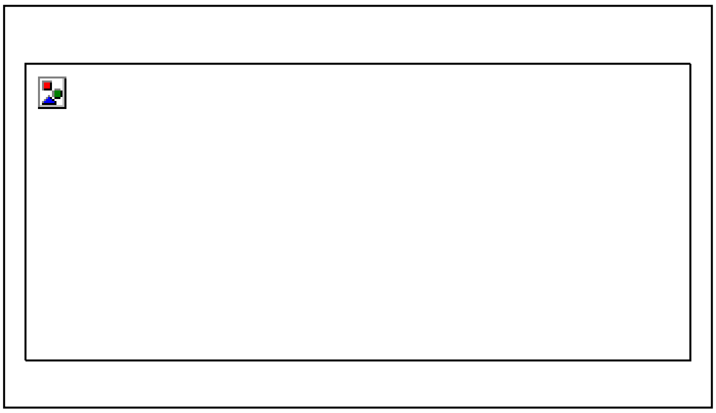


Figure 2-5. Wireless Networks [24].

An ad-hoc or peer-to-peer wireless network consists of several computers each equipped with a wireless networking interface card. Each computer can communicate directly with all the other wireless-enabled computers. They can share files and printers this way, but may not be able to access wired LAN resources, unless one of the computers acts as a bridge to the wired LAN using special software.

An infrastructure wireless network consists of an access point or a base station. In this type of network, the access point acts as a hub, providing connectivity for wireless computers. It can connect or bridge the wireless LAN to a wired LAN, allowing wireless computer access to LAN resources, such as file servers or existing Internet Connectivity.

There are four basic types of transmissions standards for wireless networking. The Institute of Electrical and Electronic Engineers (IEEE) produce these types. These standards define all aspects of radio frequency wireless networking. They have established four transmission standards; 802.11, 802.11a, 802.11b, 802.11g.

The basic differences between these four types are connection speed and radio frequency. 802.11a and 802.11b are the slowest at 1 or 2 Mbps and 5.5 and 11Mbps respectively. They both operate at 2.4 GHz radio frequency. 802.11a operates at 5 GHz frequency and can transmit up to 54 Mbps and the 802.11g operates at 2.4 GHz frequency and can transmit up to 54 Mbps. Actual transmission speeds vary depending on such factors as the number and size of the physical barriers within the network and any interference in the radio transmissions.

Wireless networks are reliable, but when interfered with it can reduce the range and the quality of the signal. Interference can be caused by other devices operating on the same radio frequency and it is very hard to control the addition of new devices on the same frequency. Usually, if your wireless range is compromised considerably, more than likely, interference is to blame [21].

A major cause of interference with any radio signals are the materials in your surroundings, especially metallic substances, which tend to reflect radio signals. The potential sources of metal around a home are numerous--things like metal studs, nails, building insulation with a foil backing and even lead paint can all possibly reduce the quality of the wireless radio signal. Materials with a high density, like concrete, tend to be harder for radio signals to penetrate, absorbing more of the energy. Other devices utilizing the same frequency can also result in interference with your wireless. For example, the 2.4GHz frequency used by 802.11b-based wireless products to communicate with each other. Wireless devices do not have this frequency all to themselves. In a business environment, other devices that use the 2.4GHz band include microwave ovens and certain cordless phones [21].

On the other hand, many wireless networks can increase the range of the signal by using many different types of hardware devices. A wireless extender can be used to relay the radio frequency from one point to another without losing signal strength. Even though this device extends the range of a wireless signal it has some drawbacks. One drawback is that it extends the signal, but the transmission speed will be slowed.

There are many benefits to a wireless network. The most important one is the option to expand your current wired network to other areas of your organization where it would otherwise not be cost effective or practical to do so. An organization can also install a wireless network without physically disrupting the current workplace or wired network. Wireless networks are far easier to move than a wired network and adding users to an existing wireless network is easy. Organizations opt for a wireless network in conference rooms, lobbies, and offices were adding to the existing wired network may be too expensive to do so.

2.2.3 Comparison between Wired and Wireless Networking

The major difference between these two types of networks is that one uses network cables and the other uses radio frequencies. A wired network allows for a faster and more secure connection and can only be used for distances shorter than 2,000 feet. While a wireless network is a lot less secure and transmission speeds can suffer from outside interference. Although wireless networking is a lot more mobile than wired networking and the range of the network is usually 150-300 indoors and up to 1000 feet outdoors depending on the terrain [25].

The cost of wired networking has become rather inexpensive. Ethernet cables, hubs, and switches are very inexpensive. Some connection sharing software packages, like internet connection sharing (ICS), are free; some cost a nominal fee. Broadband routers cost more, but these are optional components of a wired network, and their higher cost is offset by the benefit of easier installation and built-in security features.

Wired LANs offer superior performance. A traditional Ethernet connection offers only 10 Mbps bandwidth, but 100 Mbps Fast Ethernet technology costs a little more and is readily available. Fast Ethernet should be sufficient for file sharing, gaming, and high-speed Internet access for many years into the future [26]. Wired LANs utilizing hubs can suffer performance slowdown if computers heavily utilize the network simultaneously. Use Ethernet switches instead of hubs to avoid this problem; a switch costs little more than a hub.

Wireless networks using 802.11b support a maximum bandwidth of 11 Mbps, roughly the same as that of old, traditional Ethernet. 802.11a and 802.11g LANs support 54 Mbps, that is approximately one-half the bandwidth of Fast Ethernet. Furthermore, wireless networking performance is distance sensitive, meaning that maximum performance will degrade on computers farther away from the access point or other communication endpoints. As more wireless devices utilize the 802.11 LAN more heavily, performance degrades even further.

The greater mobility of wireless LANs helps offset the performance disadvantage. Mobile computers do not need to be tied to an Ethernet cable and can roam freely within the wireless network range. However, many computers are larger desktop models, and even mobile computers must sometimes be tied to an electrical cord and outlet for power. This undermines the mobility advantage of wireless networks in many organizations and homes.

In theory, wireless LANs are less secure than wired LANs, because wireless communication signals travel through the air and can easily be intercepted. The weaknesses of wireless security are more theoretical than practical [26].

Wireless networks protect their data through the Wired Equivalent Privacy (WEP) encryption standard that makes wireless communications reasonably as safe as wired ones.

No computer network is completely secure. Important security considerations for organizations tend to not be related to whether the network is wired or wireless but rather ensuring that the firewall is properly configured, employees are aware of the dangers of spoof emails, they are a way of spyware and how to avoid and that anyone outside the organization does not have unauthorized access to the network.

2.2.4 Type of Wired and Wireless Technologies

There are various wired and wireless technologies, which can be used for networking. These technologies are described below as shown in Figure 2-6.

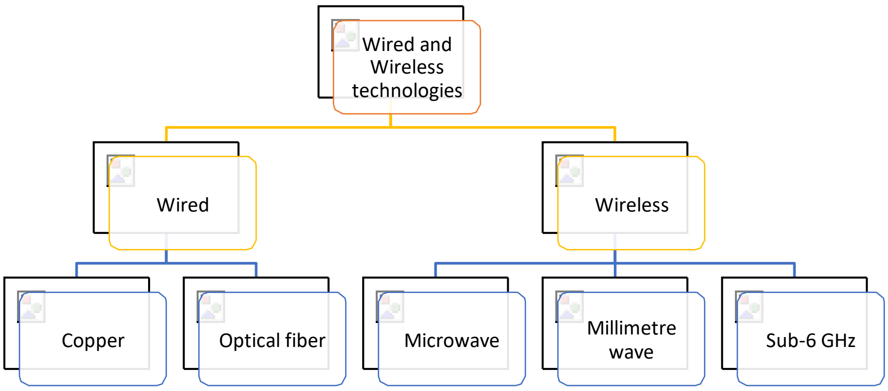


Figure 2-6. Different wired and wireless technologies [27].

1. *Wired Technology*

   i. Copper

Considering copper-based solutions, leased T1/E1 copper lines are widely utilized as they can provide suitable support for voice traffic, with deterministic QoS, low latency, and jitter. However, copper lines do not scale easily to provide sufficient bandwidth at distances exceeding a few hundred meters to support emerging broadband technologies. Even with 8-pair bonding and vectoring technology, the bandwidth is limited to around 140Mbps [28].

ii. Optical fibre

Optical fibre can provide a 100Gbps+ throughput connectivity for tens of kilometres that can be accomplished using gigabit passive and active optical network technologies [29]. An active optical system uses electrically powered switching equipment, such as a router or a switch aggregator, to manage signal distribution and direct signals to specific customers. A passive optical network uses optical splitters to separate and collect optical signals as they move through the network. Optical fibres are usually deployed in urban and sub-urban areas where very high traffic-carrying capacity is required. Although a fibre-based fronthaul offers long-term support with respect to increasing capacity requirements, this comes at a relatively high CAPEX and costly deployment.

2. Wireless Technology

Various wireless Technologies solutions exist in terms of the type of propagation, the spectrum used and the network topology. In general, the advantage of wireless backhaul is the freedom from cabling, which is expensive to deploy due to the high costs. Wireless solutions need only equipment’s at the small cell and the point of presence offering reduced costs and speed of deployment. The main categories are the following;

i. Microwave

Microwave radio can be viewed as an alternative of fronthaul connectivity particularly in areas where a wired connection is not accessible. Microwave transmission works mainly in the licensed spectrum (6GHz to 38GHz) and requires Line of Sight (LoS) [30, 31]. Microwave radio can provide a capacity of some hundred Mbps.

ii. Millimetre Wave (mmW)

The mmW-radio refers to any RF technology operation in the 30-300GHz range, but it is generally used to discuss 60-80GHz, otherwise called E-band. In this context, several GHz-wide bandwidths are available and can provide multiple Gbps even with low order modulation schemes [32]. On top of high-data rates, mmW radio band can offer superb invulnerability to interference, high security, and the frequency reuse. However, mmW radio requires clear LoS propagation and its range is restricted by the oxygen absorption which strongly attenuates signals greater than 60GHz over distances. Therefore, high gain directional antennas are used to compensate for the large free space propagation losses.

iii. Sub-6 GHz

This classification of fronthaul can be seen as a non-LoS category and incorporates carrier frequencies below 6 GHz (3.5GHz licensed and 2.4/ 5.8GHz unlicensed) [27]. Sub-6 GHz fronthaul can be easy to plan and deploy in urban regions, subsequently reducing the cost and duration of small cell network roll out. The 3.5GHz band has emerged as a promising candidate for dedicated use in small cells. On the contrary, the unlicensed spectrum gives a large amount of freely available bandwidth and it is already heavily used by Wi-Fi hotspots, Bluetooth, and other equipment.

2.3 Internet

The predecessor of today's Internet was the ARPANET (Advance Research Project Agency Network). In autumn 1969, the first computer was connected to a node at the University of California and by the end of the year; the ARPANET consisted of four connected computers with different operating systems. The ARPANET grew continuously and soon the TCP/IP suite was adopted as the official protocol suite. TCP/IP was used by other networks to link to ARPANET since 1977, which led to rapid growth.

In 1989, the World Wide Web (WWW) was invented. Mosaic, the first graphical Web browser was released in 1993 and the first search engine – 'Yahoo' (Yet Another Hierarchical Officious Oracle) – went online in 1994. The WWW led to the Internet becoming 'attractive' for ordinary people, which led to an even higher growth rate. Today, more than 1 billion people are connected to the Internet and this number is still rapidly growing.

Internet refers to the global information system that:

1. is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons;
2. is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and

3. It provides, uses or makes accessible, either publicly or privately, high-level services layered on the communications and related infrastructure described herein.

2.3.1 Internet Protocol Stack

The five-layer reference model of the Internet as related to the OSI protocol stack is described below; it is shown in Figure 2-7. It is a hybrid of the OSI reference model and the TCP/IP reference model of Clark. Each layer consists of a set of protocols for communication with another entity on the same layer and of communication services that are offered to the next higher layer. The layers can be distinguished as follows:

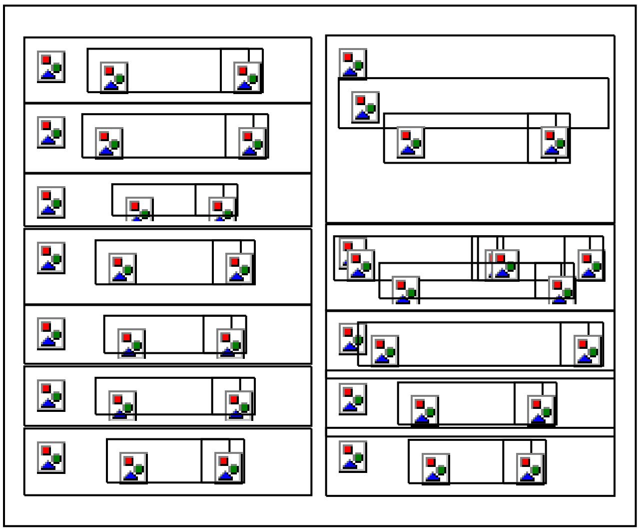


Figure 2-7. Internet TCP/IP based Communications Stack [20].

- **Physical layer:** This defines the mechanical, electrical and timing interfaces of the network.
- **Data link layer:** The main task of this layer is to transform the raw layer 1 transmission facility into a line free of undetected transmission errors

between two directly connected systems, typically by using the concept of data frames.

- **Network layer:** It is concerned with forwarding and routing of packets from sender to receiver end systems. The basic network layer protocol of the Internet is IP (Internet Protocol); it offers a connection-less datagram forwarding service.
- **Transport layer:** This layer uses the network layer to provide sender to receiver application communication. The most important transport layer protocols of the Internet are the connection-oriented virtual error-free TCP (Transmission Control Protocol) and the connection-less UDP (User Datagram Protocol).
- **Application layer:** It contains high-level protocols such as, for example, HTTP. It handles issues like network transparency, resource allocation and problem partitioning for an application. The application layer is not the application itself; it is a service layer that provides high-level services.

2.3.2 Internet Protocol version 4 (IPv4)

IP is a connection-less unreliable layer 3 protocol. ‘Unreliable’ in this context means that there are no guarantees that an IP datagram will be successfully delivered. IP is the least common denominator on top of the different layer 2 technologies that are used as infrastructure in the different autonomous systems that form the Internet.

The IPv4 header is depicted in Figure 2-8, It has a size of 20 bytes if no options are used. It consists of 14 fields, of which 13 are required. The 14th field is optional and aptly named: options [33]. IPv4 options should be avoided because IP packets with options are often processed on the slow path of an IP router and this can lead to additional delay and performance problems. The source and destination address fields contain the sender's and receiver's IP address. Information about the sender/receiver port is transport layer information and is therefore not contained in the IP header but in the TCP/UDP header instead [33].

In practice, the time to live field is used as a hop counter used to limit packet lifetimes. It is decreased at each intermediate router, and the packet is discarded if it reaches zero.

32 bits

<table>
<tr><td>Version</td><td colspan="2">Header
Lenght</td><td>Type of Service</td><td colspan="2">Total length</td></tr>
<tr><td colspan="4">Identification</td><td>Flags</td><td>Fragment offset</td></tr>
<tr><td colspan="2">Time to live</td><td colspan="2">Protocol</td><td colspan="2">Header checksum</td></tr>
<tr><td colspan="6">Source address</td></tr>
<tr><td colspan="6">Destination address</td></tr>
<tr><td colspan="6">Options</td></tr>
</table>

Figure 2-8. IPv4 Header [33]

- **Version:** The first header field in an IP packet is the four-bit version field. For IPv4, this is always equal to 4.

- **Internet Header Length (IHL):** The Internet Header Length (IHL) field has 4 bits, which is the number of 32-bit words. Since an IPv4 header may contain a variable number of options, this field specifies the size of the header (this also coincides with the offset to the data). The minimum value for this field is 5, [22] which indicates a length of 5 × 32 bits = 160 bits = 20 bytes. As a 4-bit field, the maximum value is 15 words (15 × 32 bits, or 480 bits = 60 bytes).
- **Differentiated Services Code Point (DSCP):** Originally defined as the type of service (ToS), this field specifies differentiated services (DiffServ) per RFC 2474 (updated by RFC 3168 and RFC 3260). New technologies are emerging that require real-time data streaming and therefore make use of the DSCP field. An example is Voice over IP (VoIP), which is used for interactive voice services.
- **Explicit Congestion Notification (ECN):** This field is defined in RFC 3168 and allows end-to-end notification of network congestion without dropping packets. ECN is an optional feature that is only used when both endpoints support it and are willing to use it. It is only effective when supported by the underlying network.
- **Total Length:** This 16-bit field defines the entire packet size in bytes, including header and data. The minimum size is 20 bytes (header without data) and the maximum is 65,535 bytes. All hosts are required to be able to reassemble datagrams of size up to 576 bytes, but most modern hosts handle much larger packets. Sometimes links impose further restrictions on the packet size, in which case datagrams must

be fragmented. Fragmentation in IPv4 is handled in either the host or in routers.

- **Identification:** This field is an identification field and is primarily used for uniquely identifying the group of fragments of a single IP datagram. Some experimental work has suggested using the ID field for other purposes, such as for adding packet-tracing information to help trace datagrams with spoofed source addresses, but RFC 6864 now prohibits any such use.
- **Flags:** A three-bit field follows and is used to control or identify fragments. They are (in order, from most significant to least significant):

  bit 0: Reserved; must be zero; bit 1: Don't Fragment (DF); bit 2: More Fragments (MF)

  If the DF flag is set, and fragmentation is required to route the packet, then the packet is dropped. This can be used when sending packets to a host that does not have resources to handle fragmentation. It can also be used for path MTU discovery, either automatically by the host IP software, or manually using diagnostic tools such as ping or traceroute. For unfragmented packets, the MF flag is cleared. For fragmented packets, all fragments except the last have the MF flag set. The last fragment has a non-zero Fragment Offset field, differentiating it from an unfragmented packet.

- **Fragment Offset:** The fragment offset field is measured in units of eight-byte blocks. It is 13 bits long and specifies the offset of a fragment relative to the beginning of the original unfragmented IP datagram. The first fragment has an offset of zero. This allows a maximum offset of $(2^{13} – 1) \times 8 = 65{,}528$ bytes, which would exceed the maximum IP packet length of 65,535 bytes with the header length included (65,528 + 20 = 65,548 bytes).

- **Time To Live (TTL):** An eight-bit time to live field helps prevent datagrams from persisting (e.g. going in circles) on an internet. This field limits a datagram's lifetime. It is specified in seconds, but time intervals less than 1 second are rounded up to 1. In practice, the field has become a hop count—when the datagram arrives at a router, the router decrements the TTL field by one. When the TTL field hits zero, the router discards the packet and typically sends an ICMP Time Exceeded message to the sender. The program traceroute uses these ICMP Time Exceeded messages to print the routers used by packets to go from the source to the destination.
- **Protocol:** This field defines the protocol used in the data portion of the IP datagram. The Internet Assigned Numbers Authority maintains a list of IP protocol numbers as directed by RFC 790.
- **Header Checksum:** The 16-bit IPv4 header checksum field is used for error-checking of the header. When a packet arrives at a router, the router calculates the checksum of the header and compares it to the checksum field. If the values do not match, the router discards the packet. Errors in the data field must be handled by the encapsulated protocol. Both UDP and TCP have checksum fields. When a packet arrives at a router, the router decreases the TTL field. Consequently, the router must calculate a new checksum.

- **Source address:** This field is the IPv4 address of the sender of the packet. Note that this address may be changed in transit by a network address translation device.
- **Destination address:** This field is the IPv4 address of the receiver of the packet. As with the source address, this may be changed in transit by a network address translation device.
- **Options:** The options field is not often used. Note that the value in the IHL field must include enough extra 32-bit words to hold all the options

2.3.3 Internet Protocol version 6 (IPv6)

IPv6 is a numerical label that is used to identify a network interface of a computer or a network node participating in an IPv6 computer network. It is the successor to the first addressing infrastructure of the Internet, Internet Protocol version 4 (IPv4). In contrast to IPv4, which defined an IP address as a 32-bit value, IPv6 addresses have a size of 128 bits. Therefore, IPv6 has a vastly enlarged address space compared to IPv4 [34]. The core set of IPv6 protocols form an IETF Draft Standard since 1998. Many other RFCs describe further details of IPv6 architectures and the transition from IPv4 to IPv6.

The main motivation behind the development of IPv6 was the predicted shortage of IPv4 addresses soon. IPv6 is therefore designed to 'never' run out of addresses. IPv6 addresses are 16-byte addresses, which increase the address space drastically compared to the 4-byte IPv4 addresses.

32 bits

Version	**Traffic class**	**Flow label**
Payload length	**Next Header**	**Hop limit**
Time to live	**Protocol**	**Header checksum**
Source address (16 bytes)		
Destination address (16 bytes)		

Figure 2-9. IPv6 Header [34].

Besides this, IPv6 contains other improvements over IPv4. The most important ones are as follows:

- Simplification of the IP header. By comparing the IPv4 header in Figure 2-8 with the IPv6 header in Figure 2-9, it will immediately recognizes the streamlined header layout. Many IPv4 fields are dropped or made optional. This allows routers to process packets faster and can thus improve throughput.
- IPv6 header options are encoded differently compared to those of IPv4. This results in less stringent limits on the length of options and greater flexibility in introducing new options in the future.
- The IPv6 header contains a 20-bit flow label that can be used for marking packets belonging to certain flows to give them preferential treatment. The 8-bit traffic class field can be used as Diffserv byte similar to the type of service field in IPv4.

- Authentication, data integrity and data confidentiality are other important features of IPv6.

2.3.4 User Datagram Protocol

User Datagram Protocol (UDP) is the Internet's connection-less transport layer protocol. It is a simple message-oriented connectionless protocol that is documented in [35, 36]. Connectionless protocols do not set up a dedicated end-to-end connection. Communication is achieved by transmitting information in one direction from source to destination without verifying the readiness or state of the receiver. Although UDP provides integrity verification (via checksum) of the header and payload, it provides no guarantees to the upper layer protocol for message delivery and the UDP layer retains no state of UDP messages once sent. For this reason, UDP sometimes is referred to as Unreliable Datagram Protocol. If transmission reliability is desired, it must be implemented in the user's application.

UDP does not support flow control or the reliable or even-ordered delivery of datagrams. This is mostly used for multimedia application protocols such as VoIP or video streaming.

2.3.5 Transmission Control Protocol

Transmission Control Protocol (TCP) was designed to transmit a byte stream reliably using the unreliable IP datagram service. TCP is the most commonly used transport protocol today. It is best suited for application protocols such as SMTP, FTP or the HTTP that need a reliable connection-oriented service. It is less suited for real-time streaming applications that do not need the retransmission of lost packets and that prefer to have more influence on the transmission rate [35, 36].

The following are the main features of TCP:

- TCP is a connection-oriented protocol. A TCP connection is a byte stream, not a message stream.
- TCP connections are full duplex (traffic can go in both directions) and point to point (no multicast or broadcast).
- TCP takes care of the reliable in-sequence delivery of the TCP segments. Lost packets are retransmitted and out-of-sequence segments are reordered at the end system.
- TCP's window-based flow control mechanisms allow a slow receiver to slow down a fast sending sender.
- TCP has a congestion control mechanism that tries to detect congestion in the network and adapt the window size accordingly.

TCP interprets packet loss as an indication for congestion in the network and reacts by decreasing its window size and, therefore, the number of packets it can have in the network at one point in time. A TCP sender keeps track of two windows for sending data. The advertised window of the receiver (for flow control) and the congestion window (for congestion control) [2, 37].

TCP is a self-clocked algorithm. This means that the window size is adapted in intervals proportional to the round-trip time (RTT). TCP starts in a phase called slow start. The initial value of the congestion window (cwnd) is one maximum segment size (MSS). Each time an ACK is received while in a slow start, the congestion window is increased by one segment size. Therefore, if the sender receives its full window's worth of ACKs per RTT, cwnd is doubled per RTT. Another variable, the slow-start threshold (ssthresh), is used at the sender to keep track of when to end the slow-start phase and enter the congestion avoidance phase.

2.3.5.1 TCP Rate Approximation Model

This maximal window size (W_{max}) is given by the buffer limit set aside for the connection on the receiver side. It is considered in the first part of the main min term in equation (1). The assumptions made for the derivation of the formula are that the effect of the fast recovery algorithm can be neglected and that the time spent in the slow-start phase is negligible. The latter holds true only for long-lived TCP connections. Among other assumptions, it is assumed that the round-trip time RTT is not affected by the window size [2]. This assumption is acceptable for most connections except when the connections go through an extreme bottleneck such as a low-bandwidth modem line with a large buffer. The rate r_{av} is specified in packets/second.

To obtain the rate in bytes/second, r_{av} should be multiplied with the average packet size, which is typically around 1500 bytes.

W_{max} = Maximum receiver window size (pkts); b = Number of packets acknowledged by a received ACK (typically b =2); p = Loss probability; T_0 = Retransmission timeout [s] (initially T_0 =3s, RTT = Round-Trip Time (s); r_{av} = TCP rate (pkts/s)

$$r_{av} = min\left(\frac{W_{max}}{RTT}, \frac{1}{RTT\sqrt{\frac{2bp}{3}} + T_0\, min\left(1, 3\sqrt{\frac{3bp}{8}}\right) p\left(1 + 32p^2\right)}\right) \tag{1}$$

2.3.5.2 TCP Latency Model

Many HTTP transfers of web pages, however, are short-lived TCP connections that spend little or no time in the congestion avoidance phase of TCP. For these connections, the slow-start phase has to be taken into account. The total duration D of the short-lived TCP transfer consists of the connection setup time L and the time T for the data transfer itself. The connection setup consists of a three-way handshake; Equation (2) shows the expected duration of the handshake, taking into account that the last ACK of the handshake already carries data (and is, therefore, part of T).

It is assumed that in total (d) segments are to be transmitted (d is approximately the number of bytes to be transmitted divided by the MSS which again is 1460 bytes in most cases). During slow start, there are two possibilities: If the maximum congestion window W_{max} is very large, the window size will be W at the end of the transfer. Duration T is then given by the second part of the equation (4). Otherwise, the maximum congestion window Wmax is reached during the transfer and T is expressed by the first part of the equation (4).

D = Total duration of the TCP transfer (s); r_{av} = Average TCP rate (pkts/s); L = Duration of the TCP connection establishment (s); T = Duration of the data transfer itself (s); d = Number of data segments to be transferred (pkts); RTT = Round-trip time (s); W = Unconstrained window size at end of transfer (pkts)

γ = Slow-start growth rate (γ =1.5 if one ACK per two data segments is sent) w_1 Initial congestion window size [pkts] (typically w_1 =2); W_{max} = Maximum receiver window size (pkts)

$$D = L + T \tag{2}$$

where $L = RTT$

$$W = \frac{d(\gamma-1)}{\gamma} + \frac{w_1}{\gamma} \tag{3}$$

$$T = \begin{cases} RTT.\left(log_\gamma\left(\frac{W_{max}}{w_1}\right) + 1 + \frac{1}{W_{max}}\left(d - \frac{\gamma W_{max} - w_1}{\gamma - 1}\right)\right) & when\ W > W_{max} \\ RTT.\ \left(log_\gamma\left(\frac{d(\gamma-1)}{w_1}\right) + 1\right) otherwise \end{cases} \tag{4}$$

$$r_{av} = \frac{d}{D}$$

2.3.6 Source Rate Control

The goal of rate control rule is to efficiently use the available capacity, thereby reducing queue growth and loss. Another goal is that taken fairness into consideration, which is, in fact, a somewhat more complex issue. If the rate of a sender at time t is denoted by s (t), y (t) represents the binary feedback with values 0 meaning ‘no congestion’ and 1 meaning ‘congestion’ and we restrict our observations to linear controls, the rate update function can be expressed as:

$$s(t+1) = \begin{cases} a_i + b_i s(t) & if\ y(t) = 0 \\ a_d + b_d s(t) & if\ y(t) = 1 \end{cases} \tag{5}$$

where a_i , b_i , $a_d$and b_d are constants.

The linear control rules, which has both additive and multiplicative components, are as follows: MIMD, AIAD, AIMD, and MIAD.

- Multiplicative Increase Multiplicative Decrease (MIMD)

$$a_i = 0;\ a_d = 0;\ b_i > 1;\ 0 < b_d < 1 \tag{6}$$

- Additive Increase Additive Decrease (AIAD)

$$a_i > 0;\ a_d < 0;\ b_i = 1;\ b_d = 1 \tag{7}$$

- Additive Increase Multiplicative Decrease (AIMD)

$$a_i > 0;\ a_d = 0;\ b_i = 1;\ 0 < b_d < 1 \tag{8}$$

- Multiplicative Increase Additive Decrease (MIAD)

$$a_i = 0;\ a_d < 0;\ b_i > 1;\ b_d = 1 \tag{9}$$

In the case of AIAD, the system ends up in an overloaded state, which means that it now sends the feedback ‘there is congestion’ to the sources. The same is true for MIMD, but here, a multiplication by a constant factor corresponds with moving along an equi-fairness line. By moving upwards along an equi-fairness line and downwards at an angle of 45^0, MIAD converges towards a very unfair rate allocation. Finally, AIMD approaches perfect fairness and efficiency, but because of the binary nature of the feedback.

2.4 Challenges of today’s Internet Protocol

Most importantly, the IPv4 address space ran out in 2011, forcing providers to change to the newer IPv6, which solve other problems in addition to those caused by the exploding address space, including the need for better control over the quality of service and for larger packets [38]. The author further explored the impact of MPLS technology as well as the flattening of the

Internet's autonomous system (AS) topology. Currently, the IP of today are facing the following challenges:

1. Internet of Things
2. Security and Information protection
3. Complexity
4. Time and cost
5. Legacy-Hardware issues

2.4.1 Internet of Things

Today, it is mainly humans that are connected to the Internet. Even though it requires being driven by its vast expansion, this growth is clearly coming to an end in the first world, given that most humans are already highly connected. This might lead us to ask whether the growth and therefore the demand on the scale of Internet protocols, such as on routing protocols, is coming to an end.

The Internet of Things, which envisions that everything that has power will connect to the Internet, will drive growth for years to come. It will demand not only semantically rich, high-level protocols and agents to make it usable for humans, but also basic underlying infrastructure protocols that can support the expected growth of mobile, connected devices, each of which might communicate infrequently but must be reachable at all times [38]. The authors in [39] present Locator/Identifier Separation Protocol (LISP), which aims to achieve better scalability in the routing domain, in particular for mobile devices. It does this by separating the IP addresses' location from their identification function. Changing such a fundamental concept on the Internet is a challenging task. The article highlights this challenge and shows how LISP and other new protocols can overcome it by being incrementally deployable today.

2.4.2 Security and Information protection

The most prominent concern is security; the consensus, therefore, leans toward IPv6, as that protocol covers security without requiring Network Address Translation (NAT). Most security issues relating to the individual hardware, not the protocol. Although most people are for IPv6, they also realize IPv4 has gone through the discovery and repair of vulnerabilities that IPv6 has yet to do. This process takes time. Ultimately, both protocols have their own security issues and will continue to present challenges.

2.4.3 Complexity of Infrastructure

To move a smoothly running operation from IPv4 to IPv6, a company must audit, review, upgrade, reconfigure and test its entire technology infrastructure. Everything from routers and servers to smartphones, PCs, laptops and other network-connected devices would require an upgrade to make the shift. Even the way a business runs a process and establishes policies will have to change. Moreover, planning will have to be detailed to ensure everyone, across departments, is working toward the same goal.

2.4.4 Time and Cost

Businesses that operate large networks on IPv4 could take years to complete this entire process. In addition, given the nature and complexity of the adoption process, it involves several costs. Apart from the considerable expenses of upgrading hardware and software is a direct cost to the company in time to make the transfer, both for planning and execution.

In order for the operators to remain profitable and competitive, they become more and more cautious about the Total Cost of Ownership (TCO) of their network. The TCO, including the CAPEX and the OPEX. The CAPEX is mainly related to network infrastructure building, whereas OPEX is mainly associated with network operation and management.

2.4.5 Hardware issues

When the transition takes place, network-connected devices may need a dual-stack configuration: that is, both an IPv4 address and an IPv6 address. Older devices that lack support for an IPv6 address could create network-communication issues and eventually become a liability.

2.5 Challenges and Prospects of 5G System

5G technology is a term used by many researchers to describe the forthcoming 5th generation of mobile networks or future generation of wireless systems [40-45]. It may include a dense, highly integrated network comprised of small cells supporting data rates in the order of 10 Gbps, with a roundtrip latency of 1 ms or less [46]. Some of the advantages of 5G over the previous technologies are fast broadband speeds, improved coverage, ability excellent QoS, and increasingly reliable connectivity.

It was reported from Japan that there will be a 5G system in use during Olympic Games in Tokyo 2020. However, there are no 5G standards available for now [15].

2.5.1 5G Concepts and Requirements

The key performance requirements from end-users' perspective are as follows: traffic volume density, throughput of end-user, latency, reliability, and sustainability. To have support for the increasing data traffic and maximise profits, the operators, therefore, need to perform a feasibility study of operating their networks. The key principle is to deliver higher capacity at a reduced cost [47, 48]. The data rate can be measured in several different ways, and there will be a 5G goal target for each such metric [47].

a) Aggregate data rate or area capacity refers to the total amount of data the network can serve, characterized in bits/s per unit area.

b) Edge rate or 5% rate is the worst data rate that a user can reasonably expect to receive when in the range of the network, and so is an important metric and has a concrete engineering meaning. Goals for the 5G edge rate range from 100 Mbps (easily enough to support high-definition streaming) to as much as 1 Gbps. This requires about a 100× advance since current 4G systems have a typical 5% rate of about 1 Mbps, although the precise number varies quite widely depending on the load, the cell size, and other factors.

c) Peak rate is the best-case data rate that a user can hope to achieve under any conceivable network configuration. The peak rate is likely to be in the range of (1 to 10Gbps).

Table 1. 5G requirements [47]

Requirements	Desired value	Application example
Data rate	1 to 10 Gbps	Virtual reality office
Data volume	9 Gbytes/h in busy period 500 Gbytes/mo/subscriber	Dense urban information society
Latency	Less than 5 ms	Traffic efficiency and safety
Battery life	One decade	Massive deployment of sensors and actuators
Connected devices	300,000 devices per AP	Massive deployment of sensors and actuators
Reliability	99.999%	Teleprotection in smart grid network Traffic efficiency and safety

Table 1 depicts the basic challenging requirements for the 5G mobile and wireless communication systems. More recent attention has focussed on the provision of 5G future integration of access technologies into one seamless experience such as device-to-device (D2D) communication, massive machine communication (M2M), moving networks (MNs), ultra-dense networks (UDNs), and ultra-reliable communication (URC) [49]. This is shown in Figure 2-10.

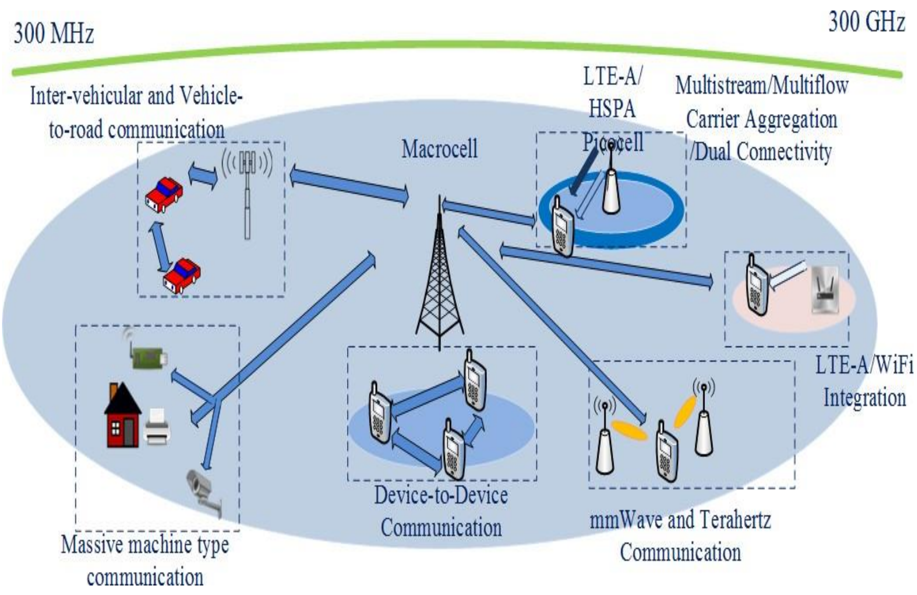


Figure 2-10. Technologies of 5G Mobile and Wireless Networks [44].

2.5.1.1 Device-to-Device (D2D) Communication

This is a means of having direct communication between devices, without the involvement traffic passing through the core network infrastructure. This will enhance the minimisation of resulting interference due to the environment whereby increasing coverage, offload backhaul, and increase capacity per area. D2D as stated in [49] is future technology for mobile systems, which is further divided into two types: co-channel frequency D2D and dedicated frequency D2D. The former works in the same frequency as that between UE and cell while later owns a specific frequency, which is different from that between UE and cell.

2.5.1.2 Machine-to-Machine (M2M) Communication

M2M is refers to method of communication for billions of different network devices, mostly essential for future mobile and wireless communication systems. This possesses a wide range of traffic characteristics and requirements such as data rate, latency, and cost.

2.5.1.3 Moving Networks (MNs)

A Moving Networks provide end-users on the move with high level of quality of experience (QoE). It is an extension of coverage for large populations, which are substantial part of communication of devices on the move or group of moving network nodes communicating with other nodes in it environment.

2.5.1.4 Ultra-Dense Networks (UTNs)

This focuses on the accelerating demands of traffic due to the densification of infrastructure. This will enable increase in capacity and energy efficiency of radio links. Alternatively, there will an increase of approximate magnitudes, in which the growth rate of access nodes is about 78% within period of 2012-2016 as confirmed in [49].

The 5G system will consists of various system densely deployed on top of the core network. Based on these features, the 5G systems will have the following disadvantages;

- High-bandwidth consumption in the WAN due to deployment of VPNs in IP/MPLS networks.
- Additional cost of NOC infrastructures.
- Traffic congestion due to billions of devices connected to the 5G network.
- Underutilisation of NOC infrastructures especially during low traffic periods.

2.6 **Summary**

In this chapter, both connection-oriented and connectionless-oriented packet switched networks are described. Then, full details of wired and wireless networking are provided with their types and comparison between them. Further discussion on the basic network and transport layer protocols of the Internet: Internet Protocol Stack, IPv4, IPv6, UDP, and TCP. TCP is the most commonly used transport protocol on the Internet and uses complex congestion and flow control mechanism that was discussed in this chapter. Methods for estimating the throughput of a TCP connection for different network conditions were discussed as well. In addition to this, we presented the challenges of today's internet protocol and 5G system with their prospects. Chapter 3 will be focusing on an elaborate description of Multi-Protocol Label Switching (MPLS) network.

3 Multi-Protocol Label Switching (MPLS) Network

3.1 Introduction

The details of MPLS network in terms of its components and functions will be discussed in this chapter. Then, this will be followed by the discussion of MPLS traffic engineering and technologies. Furthermore, the mathematical model for MPLS operation and bandwidth constraint models will be presented.

MPLS is a high-performance packet forwarding technology that integrates the performance and traffic management capabilities of the data link layer (Layer 2) switching with the scalability, flexibility, and performance of network-layer (Layer 3) routing. It enables enterprises and service providers to provide differentiated services without sacrificing the existing infrastructure [50, 51]. With and without explicit traffic engineering, it is growing in popularity for provisioning and managing core networks. Practically every modern router can do plain IP packet forwarding and MPLS.

It is a mechanism in which packets are forwarded based on labels and technology for the IP services delivery. The labels usually carried in the Layer 2 header (Shim), which is located between layer 2 header and IP. Therefore, MPLS is a layer 2.5 (between MAC and IP layers) [50]. It can also correspond to other parameters such as QoS or source address and is designed to support forwarding of other protocols as well. The technology of MPLS is employed in any environment regardless of the Layer 1 media and Layer 2 protocol. It uses a 32-bit label field that is inserted between Layer 2 and Layer 3 headers [3].

Most often, the network protocols used by the conventional network are based on the shortest path algorithms, indicating only one path between source and destination. In MPLS, multiple paths can be chosen simultaneously for forwarding the packets, which improve the overall performance. MPLS provides explicit routing, which is helpful in traffic engineering and to increase traffic-oriented performance. MPLS manages the bandwidth and meets the service requirements for the next generation when it is used as a traffic-engineering tool.

3.2 Fixed or Static Routing Path Selection

Fixed Routing (FXR) is an important routing topology employed in all type of networks, including IP-, ATM- and TDM-based networks. The routing pattern, in this case, is where the route and route selection sequence are adequately measured and maintained over a long period of time [52] . The main idea behind fixed routing is to carry traffic which is economically feasible over direct links between pairs of nodes low in the hierarchy. Therefore, the routing connection requests in a hierarchical network involves an originating ladder, a terminating ladder, and links interconnecting the two ladders. For instance, BB1 and BB2 could be backbone nodes in a backbone area and AN1 and AN2 could be access nodes in separate access areas which are difference from backbone area. A preferable routing sequence for AN1 to AN2 is illustrated in Figure 2-2.

A connection request involving no via nodes: It indicates no intermediary connection between the nodes (path AN1 – AN2). There is another connection that requests one via node: This has one interconnecting node between the originating and terminating nodes (path AN1 – BB2 – AN2, AN1 – BB1 – AN2).

Also, the connection request can involve two via nodes: Here there is the possibility of having two interconnecting nodes (path AN1 – BB1 – BB2 – AN2). The above procedure gives only the first-choice inter-ladder link from AN1 to AN2. To follow the route AN2 – AN1 request, it requires different route entirely with a reverse diagram, which the AN2 – BB2 the originating ladder and AN1 – BB1 the terminating ladder. In Figure 3-1, the preferred path from AN2 – AN1 is AN2 – AN1, AN2 – BB1 – AN1, AN2 – BB2 – AN1, and AN2 – BB2 – BB1 – AN1, in that order [52].

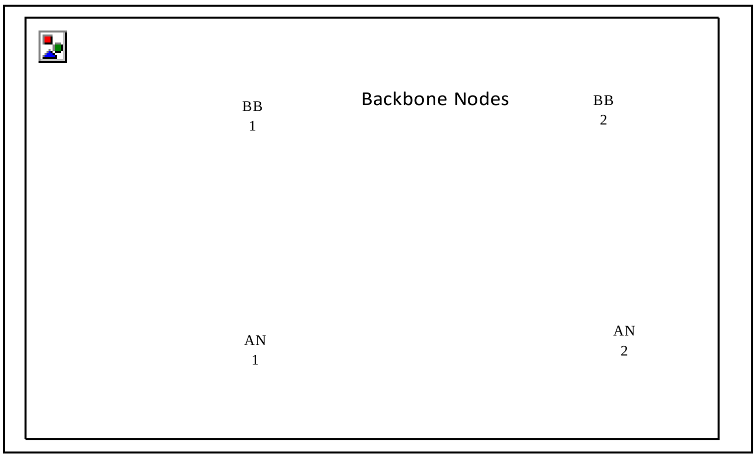


Figure 3-1. Hierarchical fixed routing path selection methods [52].

3.2.1 MPLS Functions and Components

A label (20 bits) is a short, fixed length, an identifier that is used to acknowledge Forwarding Equivalence Class. A packet may be assigned to FEC based on its network layer destination address; however, the label does not directly encode any information from the network layer header.

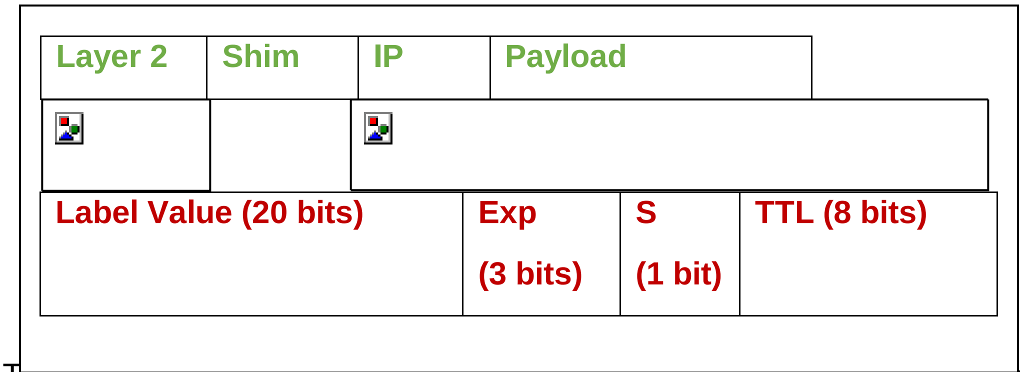


Figure 3-2. MPLS Shim Header

A labeled packet is a packet into which a label has been encoded. It may reside in an encapsulation header that exists specifically for this purpose, called an MPLS ‘shim’ header or a stack entry, as it is often referred, and as shown in Figure 3-2. This is the 32 bits of the label field of MPLS [53].

Experimental bits (3 bits) perform the function of carrying the IP precedence value. They are used for Quality of Service (QoS) to set the priority that the labeled packet should have. Then the bottom of stack bit (1 bit) indicates whether the label to the next header is another label or a Layer 3 header. It tells the MPLS Router if it is the last leg of the destination and there are no more labels for the onward process. This usually means the router is an egress router. Moreover, TTL (Time to leave) (8 bits) field is mainly for preventing the indefinite looping of packets. This identifies how many hops the packet can make before it is discarded.

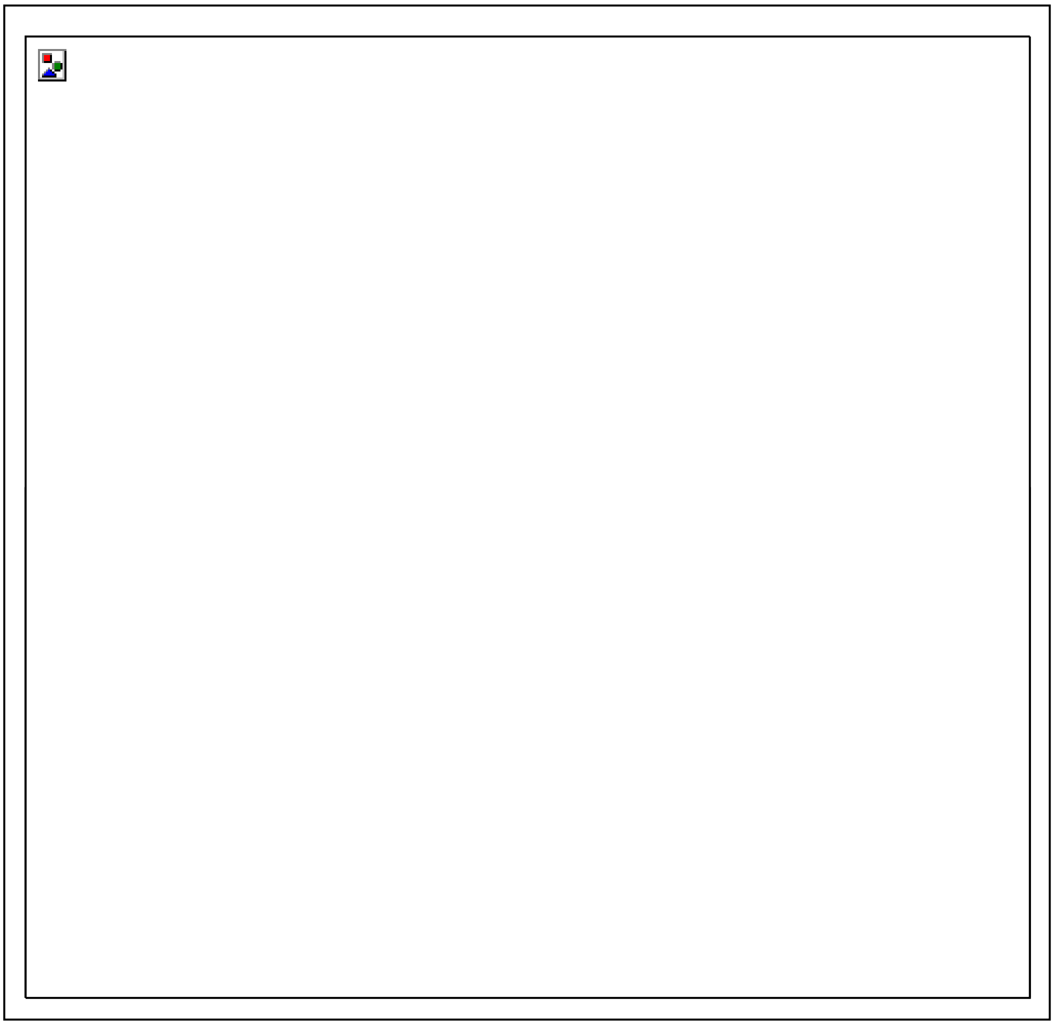


Figure 3-3. Architecture of IP/MPLS Networks.

The architecture of IP/MPLS networks and the detailed process of routing information transfer from the control plane to the data plane in the LSR are shown in Figure 3-3. The router functionality is divided into two major parts: the control plane and data plane.

1. Control Plane: In this plane, the layer 3 routing information and labels are exchanged. It contains complex mechanisms for exchanging routing information such as Open Shortest Path First (OSPF), Enhance Internet Gateway Routing Protocol (EIGRP), Intermediate System to -Intermediate

System (IS-IS), and Border Gateway Protocol (BGP). Other mechanisms such as LDP and RSVP are also used to exchange labels.

2. Data plane: This forwards packets based on labels and has a simple forward engine. It can directly influence the forwarding behaviour of packets. Data plane QoS mechanisms can be categorized in terms of the primitive behavioural characteristics such as classification, marking, policing, prioritization, queuing and scheduling.

3. Label Switch Router

A Label Switch Router (LSR) is a device that is capable of forwarding packets at layer 3 and forwarding frames that encapsulate the packet at layer 2. The label swapping mechanism is implemented at layer 2. When an LSR receives a packet, it uses the label included in the packet header as an index to determine the next hop on the label-switched path (LSP) and a corresponding label for the packet from a lookup table. The old label is then removed from the header and replaced with the new label before the packet is routed forward.

4. Label Edge Router

A Label Edge Router (LER) is both a router and a layer 2 switch that is capable of forwarding MPLS frames to and from an MPLS domain. It performs the IP to MPLS FEC binding including the aggregation of incoming flows. It also communicates with interior MPLS LSRs to exchange label bindings. Often referred to as an ingress or egress LSR, because it is situated at the edge of an MPLS domain. When forwarding an IP datagram into the MPLS domain, an LER uses routing information to determine the appropriate label to be affixed, labels the packet accordingly, and then forwards the labeled packet into the MPLS domain. Likewise, upon receiving a labeled packet which is destined to exit the MPLS domain, the LER strips off the label and forwards the resulting IP packet using normal IP forwarding rules.

5. Label Switch Path

A Label Switch Path (LSP) is an ingress-to-egress switched path built by MPLS nodes to forward the MPLS encapsulated packets of a FEC using the label swapping forwarding mechanism.

6. Forwarding Equivalence Classes

A Forwarding Equivalence Class (FEC) is a set of packets that are treated identically by a router, i.e., forwarded out the same interface with the same next hop and label, and assigned the same class of service. When a packet enters the MPLS domain at the ingress node, it is mapped into an FEC. The mapping can be done according to several factors, i.e., the address prefix, source/destination address pair, or ingress interface.

The label-forwarding in MPLS begins at the ingress edge router called Label Edge Router (LER router) in which the label is assigned and imposed by the IP routing process. This is followed by the swapping of labels on the contents of the label forwarding table in the core using Label Switch Router (LSR). At the egress edge router, the label is disposed and a routing lookup is used to forward the packet. Therefore, LSR forms the basis for labeled packets forwarding (label swapping), while Edge LSR labels IP packets and forwards them into the MPLS domain, or removes labels and forwards IP packets out of the MPLS domain.

MPLS forwarding is based on a fixed-length label. The label exchange happens in the control plane in much the same way that routing information is exchanged prior to the forwarding of IP packets [53]. Label exchanges are done through a label distribution protocol such as LDP, RSVP, or BGP. In the forwarding plane, the ingress LER imposes one or more labels on the data packet. The core devices switch the packet to the egress LER based on the top label only.

In MPLS, packets are assigned to FEC at the ingress Label Edge Router (LER) located at the edge of the MPLS domain. The FEC to which they are assigned can be dependent on several attributes including the address prefix in the packet's header. However, the assignment of a packet to FEC is done just once, as the packet enters the MPLS domain [16, 54]. Specifically, MPLS routers establish a label-switched path (LSP), a pre-determined path to route traffic in an MPLS network, based on the criteria in the FEC. It is only after an LSP has been established that MPLS forwarding can occur. LSPs are unidirectional which means that return traffic is sent over a different LSP. The detail of the Packet Processing Algorithm in MPLS is illustrated in Figure 3-4. Packets are forwarded alongside a label switch path (LSP) where each LSR (label switch router) makes forwarding decision entirely based on the content of the label, in this way eliminating the need of the IP address so that transitional router does not have to execute routing lookup which is very time-consuming process [3]. In this process, the LSR acquires the exit of the label of each hop and for the next hop, it should be put on a new label. The packet forwarding decision is also resolved by next hop by doing an interpretation of the label on the packet these established paths [4].

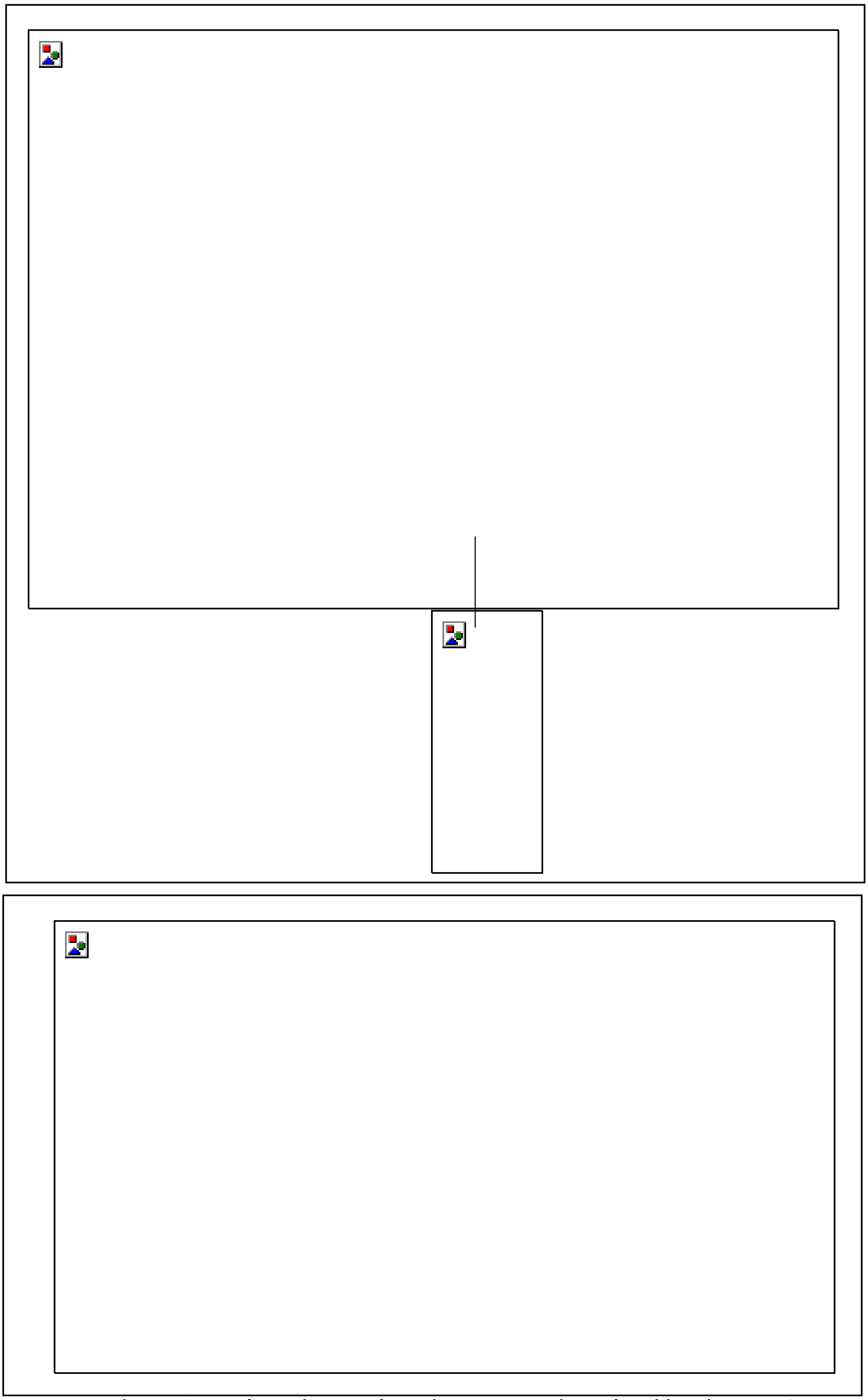


Figure 3-4. Flow charts of Packet Processing Algorithm in MPLS.

3.2.2 Mathematical Model of MPLS operation

Let G = (N, E) be a graph depicting the physical topology of the network. Then, N is the set of nodes in the network and E is the set of links; Let H = (U, F, d) be the induced MPLS graph, where U is a subset of N representing the set of LSRs in the network, F is the set of LSPs, and d is the set of demands [55]. All the set of routers, in accordance with MPLS network formation, can be categorized into two subsets:

In MPLS network, finding a solution to routing issues in terms of flow models is necessary in order to calculate one or a multitude of paths (Label Switching Path, LSP) between a pair of edge “sender-receiver” nodes and define the sequence of the set intensity of traffic distribution between them [56].

$N^{+} = \{U^{+}_{r}, r = \overline{1, m_{LER}}\}$ - A subset of edge routers.

$N^{-} = \{U^{-}_{j}, j = \overline{1, m_{LSR}}\}$ - A subset of intermediate routers.

U_r^{+} –: $r - LER$ at which k-traffic arrives into MPLS Network.

U_e^{+} –: $e - LER$ at which k-traffic leaves MPLS Network.

K_r^s -: Multitude of s is in a class of services (CoS), arriving into $r - LER$.

$I^{k^s_r}$ –: Intensity of $k^s_r -$ traffic with servicing class, arriving into $r - LER$.

$p_{ij}^{k_r^s}$ –: Routing variable, which characterized the intensity of $k^s_{\bar{r}}$ traffic in (i,j) $\in E$ link for every $r - LER$ and $k^s_r \in K^s_r$.

φ_{ij} –: Intensity of the available link bandwidth from i to j

p_{ij} -: traffic from i to j

$$\begin{cases} \sum_{J:(i,j)\in E} p_{ij}^{k_r^s} - \sum_{J:(i,j)\in E} p_{ji}^{k_r^s} = I^{k_r^s}, if\ i = U_r^+; \\ \sum_{J:(i,j)\in E} p_{ij}^{k_r^s} - \sum_{J:(i,j)\in E} p_{ji}^{k_r^s} = 0, if\ i \neq U_r^+, U_e^+; \\ \sum_{J:(i,j)\in E} p_{ij}^{k_r^s} - \sum_{J:(i,j)\in E} p_{ji}^{k_r^s} = -I^{k_r}, if\ i = U_e^+; \end{cases} \quad (10)$$

The equations in (10) imply the number of LERs and LSRs in the network system. Furthermore, it shows the process of packet forwarding in MPLS from the ingress LER (entry) through LSRs to the egress LER (exit). This is to prevent packet loss on the routers in the MPLS network [55, 57]. The whole set of k - traffics, arriving from users (access networks), depending on which edge router this traffic comes from and according to which class it will be serviced.

$$\sum_{s=1}^{S} \sum_{k_r^s \in K_r^s} p_{ij}^{k_r^s} \leq \varphi_{ij} - \sum_{i=1}^{S} \sum_{\substack{g \in U^+ \\ g \neq r}} \sum_{k_r^s \in K_r^s} p_{ij}^{k_g^s} \left(r \in U^+, (i,j) \in E\right) \quad (11)$$

The meaning of equation (11) inequality is that the traffic, routed from r - LER, cannot exceed by its intensity of the available bandwidth of the link, which is left after traffics service, routed from other edge routers.

3.2.3 MPLS Label Distribution

Modern networks change on a dynamic basis. To accommodate the adjusted needs of these networks, many network engineers have chosen to use the second method of programming MPLS switches: dynamic signaling and label distribution. Dynamic label distribution and signaling can use one of several protocols.

Labels are distributed between LERs and LSRs using the Label Distribution Protocol (LDP). In the LDP, there are LSR discovery mechanisms, which implies that the protocol will initially discover the neighbouring label switch routers through the LSR mechanisms. Basically, LDP runs over transmission control protocol to provide reliable delivery of messages. TCP possesses its own congestion control and flows control mechanisms, which may not be appropriate for the distribution of the labels in the LSRs. However, it is agreeable to the MPLS standards that LDP will run over a TCP and make use of TCP's capabilities for providing the reliable delivery of information.

Within the IETF, two label distribution protocols that also allow the set-up of explicit paths for traffic engineering namely: RSVP-TE and Constraint-based Routing support for LDP (CR-LDP). The LDP, BGP, and IS-IS protocols establish the Label Switch Path (LSP but do little in the service of traffic engineering because routed traffic can potentially be redirected onto a high-priority LSP, thereby causing congestion.

To overcome this problem, signaling protocols were established to create traffic tunnels (explicit routing) and allow for better traffic engineering. These protocols are Constraint Route Label Distribution Protocol (CR-LDP) and Resource Reservation Setup Protocol (RSVP-TE).

In RSVP-TE, RSVP messages are exchanged directly via raw IP datagrams and the protocol uses soft state. The label distribution method is downstream on demand: If an ingress LSR determines that a new LSP should be set up to a certain egress LSR, a PATH message is sent containing a specified explicit route. That route can be different from the standard hop-by-hop route. The message also contains the traffic parameters for the new route. Each router along the path receives the message builds up the state. While in the case of CR-LDP, CR-LDP uses TCP connections for reliable message exchange and is a hard-state protocol. It does not need to refresh the set-up of an LSP. The label distribution method is downstream on demand as in RSVP-TE. A label request message is sent by the ingress LSR towards the egress. It contains an explicit route. Contrary to RSVP-TE, intermediate routers reserve resources immediately when the label request reaches them. The egress-router responds to the label request with a label-mapping message that contains the label and information about the final resource reservation. It is routed back to the ingress nodes.

3.2.4 Resource Reservation Protocol (RSVP)

RSVP is a state-establishment protocol that will enable the internet to support real-time and multimedia applications, such as teleconferencing and video conferencing applications. These applications require reservations to be on the internet routers, and RSVP is the protocol to set up for these reservations [58-60]. It allocates bandwidth along the LSP for tunnels to establish [53]. The key features of RSVP include flexibility, robustness, scalability through the receiver control reservations, sharing of reservations and use of IP multicast for data distribution.

RSVP can be defined as a means by which applications communicate their requirement to the network in an efficient and robust manner [58]. In other words, it enables real-time applications to obtain the different (QoS) for their data flows. It does not provide any service; it simply communicates any end-system requirement to the network. RSVP is not a routing protocol but works in conjunction with the routing protocol. It is usually referred to as a signaling protocol used in MPLS Traffic Engineering

RSVP’s three levels of service include best-effort, rate-sensitive, and delay-sensitive service [59, 60].

Best-effort service is used for applications that require reliable delivery rather than timely delivery.

Rate-sensitive service is used for any traffic that is sensitive to variation in the amount of bandwidth available. Such applications include H.323 video conferencing, which was designed to run at a nearly constant rate.

Delay-sensitive traffic requires timely but not reliable delivery of data.

3.3 Data Forwarding Architectures

The data forwarding architecture describes the actual technical packet forwarding technology. Internet Network Service Providers (INSPs) can use plain IP packet forwarding where every hop in the path of the packet through the network is an IP router that looks up IP header information in its routing table to decide on how to forward the packet. An alternative data forwarding architecture is label switching packets using MPLS technology.

3.3.1 IP and MPLS

In traditional IP routing, each router analyses the header of the arriving packet and independently chooses the next hop based on the distributed routing algorithm that uses a routing protocol and the information of the IP header.

Contrary to that, in a network using MPLS as data forwarding architecture, the FEC assignment is done at the MPLS ingress node just once when the packet enters the MPLS domain. An MPLS domain is a contiguous set of nodes using MPLS as the forwarding architecture. The FEC is encoded into a 4-byte label that is attached to the packet between the layer-2 and layer-3 headers. Subsequent hops no longer should examine the IP header of the packet, the label is used as an index into a forwarding table that specifies the next hop as outgoing interface and a new label that replaces the old one (label swapping).

MPLS offers some advantages over the conventional IP forwarding architecture:

- MPLS forwarding could be done by switches that do not have to can analyse the IP headers. MPLS forwarding is a simpler operation than IP routing and less expensive to implement for operations at state-of-the-art line speeds.
- The ingress router assigning the label can use any information to assign a label to a packet. Apart from analysing the destination address of the IP header, the transport layer ports could be evaluated, or the DSCP of a packet in a Diffserv domain.
- Additionally, the process of determining the label can become more and more sophisticated without any impact at all on the core routers.
- As information about the ingress router does not travel with an IP packet, traditional IP routing does not allow differentiation between packets from two different ingress routers in the core. With MPLS, this can easily be done if each ingress routers assign a different label.
- For traffic engineering or policy reasons, it may be desirable to force packets to follow a path different to the standard shortest path as it is determined by the routing protocol algorithm. With MPLS traffic

engineering, the path set-up can be controlled centrally and any path through the network can be used.

These advantages make it obvious that a network with MPLS-based forwarding architecture is well-suited for traffic engineering.

3.3.2 MPLS and VPLS

It is generally observed that MPLS and VPLS (Virtual Private LAN Service) are common in terms of point-to-multipoint connectivity and transport of IP packets. MPLS can carry various different protocols, it is known by its enabled applications such as L3VPNs, L2VPLS, EoMPLS, AToMPLS, FRoMPLS [61, 62]. While VPLS is a protocol for building a virtual multipoint Ethernet network on top of a MPLS network or IP network). VPLS performs the same way of MPLS, but it has the capability to do more services than MPLS. However, MPLS is still a vital option for many businesses. With VPLS customer level of achievement increases dramatically from a technical perspective. VPLS is also easier to support, which in turns reduces costs and improves the customer experience for a business. However, hybrid deployment of both MPLS and VPLS technologies and services may be better suitable to meet business-critical needs.

VPLS provides a more complete multipoint communication solution, integrating the advantages provided by Ethernet and MPLS. By emulating traditional LAN functions, VPLS enables users on different LANs to communicate with each other over MPLS networks as if they were on the same LAN. MPLS VPNs can be classified into MPLS L2VPNs and MPLS L3VPNs:

Traditional MPLS L2VPNs, such as the Virtual Leased Lines (VLLs) or Virtual Private Wire Services (VPWSs), can provide P2P services but not P2MP services over a public network.

Virtual Private LAN services are growing in popularity for delivering Ethernet services. They combine MPLS and Ethernet allowing both customers and carriers to benefit. For over two decades, Ethernet switching has dominated the local area network while IP routing has dominated the carrier network. IP backbones have been used to provide Internet access and more recently to provide IP VPN access [63].

MPLS L3VPNs can provide P2MP services on the precondition that Provider Edge routers (PEs) keep routes destined for end users. This implementation requires a high routing performance of PEs.

PE1, PE2, and PE3 belong to the same VPLS domain. AC links connected to the VPLS domain are mapped to PWs through a VSI to generate the forwarder of the VSI. When Customer Edge router (CE1) receives a Layer 2 packet from a user at Site1, it forwards the Layer 2 packet to PE1 through the AC link as shown in Figure 3-5.

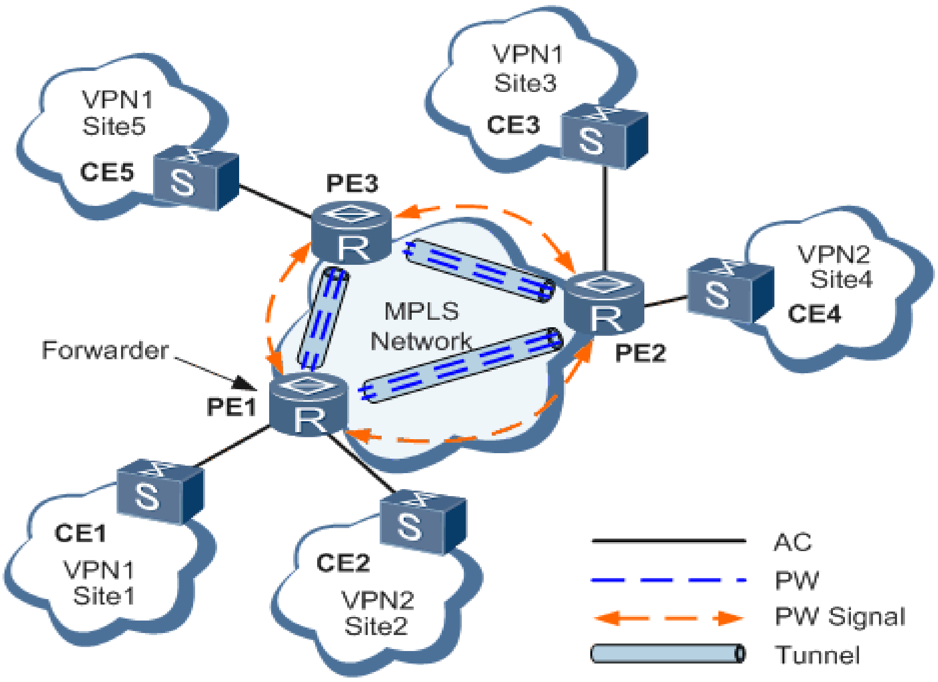


Figure 3-5. MPLS VPN Architecture [61].

The entire VPLS network is like a switch. PWs are established over MPLS tunnels between VPN sites to transparently transmit Layer 2 packets between sites. When forwarding packets, PEs learn the source MAC addresses of these packets and create MAC entries, mapping MAC addresses to attachment circuits (ACs) and PWs [61].

There is need to evolve from seamless MPLS to meet the more comprehensive capabilities of 5G. An evolved seamless MPLS needs to enable agile and seamless connectivity for distributed Network Functions in the cloud edge and core datacentres.

Network virtualization enables operating several logical networks over one physical network. To support network virtualization, it is necessary to tag the packets so that routers can determine to which VPN instance a packet belongs. Labels in MPLS provide this functionality. LISP supports such tagging of packets and mappings with the Instance ID in the LISP specific header.

A unified IP/MPLS wide area network is still critical but it needs to support End-to-End (E2E) topology awareness and service-aware traffic engineering for deterministic SLA guarantees. A number of key protocols that should be supported [64]: (1) Ethernet VPNs can be used for network virtualization in datacentres as well as providing layer 2 or layer 3 virtual private network connectivity between virtual networks (2) Segmented Routing is a scalable approach, which can used to leverage regular Internal Gateway Routing Protocols (IGP) such as OSPF, IS-IS and BGP to distribute topology information.

3.4 MPLS Traffic Engineering (MPLS-TE)

In the data communications world, traffic engineering provides an integrated approach to engineering traffic at layer-3 in the OSI model. The integrated approach means that routers are configured to divert from destination-based forwarding to move the traffic load from congested parts of the network to non-congested parts. Traditionally, this diversion is done using overlay networks where routers use carefully engineered ATM or Frame Relay to distribute the traffic load on layer 2. The routing tables are created by an interior routing protocol, IGP, which finds the least-cost route according to its metric to each destination in the network [65].

Traffic engineering (TE) is the process of routing data traffic in order to balance the traffic load on various links, routers, and switches in the network [66, 67]. In other words, it is a technique that makes better use of the existing bandwidth in a network by moving traffic from over-utilized links to less-utilized links. It is most effective in networks where some links are heavily utilized and have little or no bandwidth available while others carry little or no traffic.

In many networks, this method works well. However, in some networks the destination based forwarding results in the over-utilization of some links while others are under-utilized. This imbalance will be the case when there are several possible routes to reach a certain destination and the IGP selects one of them as the best and uses only that. In the extreme case, the best path may have to carry so large a volume of traffic that packets are dropped while the non-best path is almost idle. One solution to the problem would be to adjust the link bandwidths to more appropriate values. Reduce the underutilised link and increase the over-utilized one. However, this adjustment is not always possible. The alternate path is a backup path. In the case of a primary link failure, the backup must be able to forward at least the major part of the traffic volume normally forwarded by the primary. Therefore, it may not be possible to reduce the bandwidth. Without a cost saving, the budget may not allow an increase to the primary link bandwidth. To provide better network performance within budget, network administrators move a portion of the traffic volume from the over-utilised link to the under-utilized link. During normal operations, this move results in less packet drops and quicker throughput. In the case of a failure to any of the links, all traffic is forwarded over the remaining link, which then of course becomes over-utilized.

The essence for the requirement of traffic engineering can be categorised into three as stated in [53, 54]:

Firstly, congestion in the network occurs during the period of election news, online trading and special sports events. Secondly, it has to do with the better utilization of available bandwidth in which packets route on the non-shortest path and fast reroute around the failed links/nodes. Finally, capacity planning in terms of improving the aggregate availability of the network by traffic engineering.

3.4.1 Differentiated MPLS Traffic Engineering (DSTE)

The Differentiated Service MPLS Traffic Engineering (DSTE) is an aspect, which combines the capabilities of QoS and DSTE capabilities of MPLS to allocate bandwidth and control QoS for various virtual networks (also known as the class of service in DSTE) [52]. Traffic engineering is an application of MPLS, which gives support to enhance the efficiency and reliability of the network operations and also optimizes the utilization of network resources [68]. The allocation of bandwidth to each class type (CT) and the provision of bandwidth protection and QoS can be implemented using admission control. There are three “Bandwidth Constraint Models (BCM)”, which have been experimental to control bandwidth allocation/protection within the DSTE framework. The illustrations of three BCM are shown in Figure 3-6, Figure 3-7, and Figure 3-8 respectively.

Maximum Allocation Model (MAM) [69]: This model provides the maximum allocation of bandwidth to each class type on each link, and another class type protects it from use.

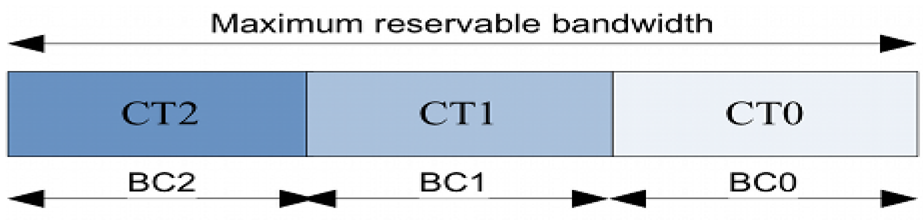


Figure 3-6. Maximum Allocation Model (MAM) [70].

Maximum Allocation with Reservation (MAR) [71] allocates bandwidth to each class type (CT) on each link and uses dynamic bandwidth reservation to protect bandwidth under congestion, but allows bandwidth sharing in absence of congestion.

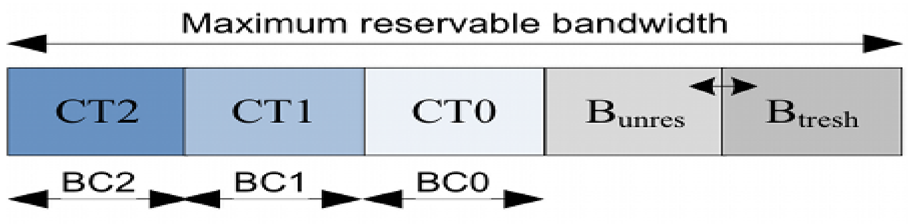


Figure 3-7. Maximum Allocation with reservation (MAR) [70].

Russian Dolls Model (RDM) [72] allocates bandwidth progressively to each CT but requires LSP pre-emption to work well.

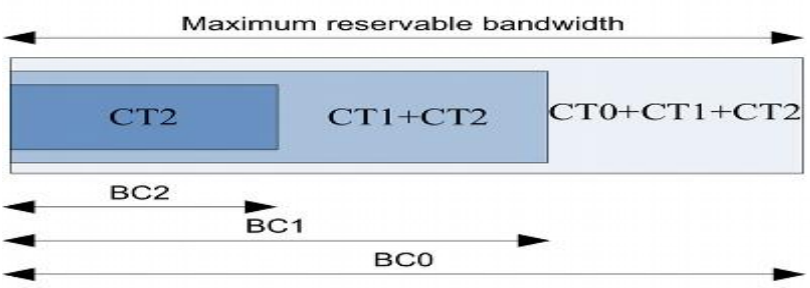


Figure 3-8. Russian Dolls Model (RDM) [70].

It is illustrated that with the implementation of the above-mentioned bandwidth constraint models, RDM can yield poor results since the pre-emption is not enabled. In the case of analysis and simulation results of MAR and MAM bandwidth constraint models, the MAR bandwidth constraints model perform better than the MAM bandwidth constraints model [33] as shown in Figure 3-9.

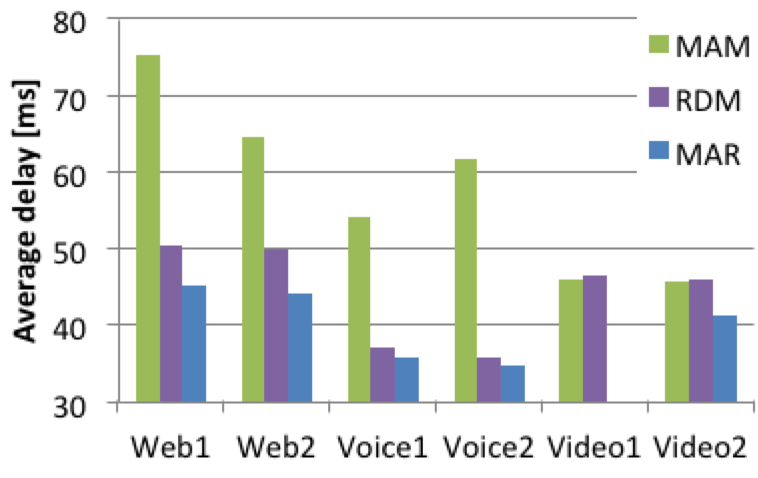


Figure 3-9. Average delays of MAM, RDM, and MAR [70].

3.4.2 MPLS-TE Model

MPLS is one of the tools that can be used to implement traffic engineering. It is of the type that gives preferential treatment to certain types of traffic needs to have TE-configured differently from a network that does not [73]. The components of MPLS-TE Model is shown in Figure 3-10.

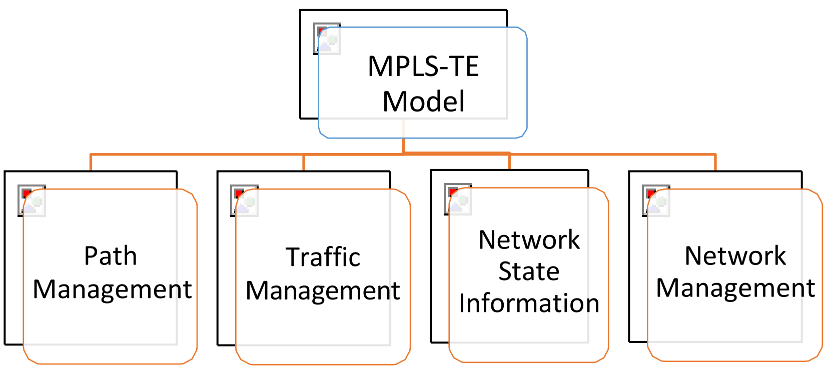


Figure 3-10. MPLS Traffic Engineering Model.

Path management, which includes label distribution, path placement, path maintenance, and path revocation. These are used to establish, maintain, and tear down LSPs in the MPLS context. The label distribution, path placement, and path revocation functions are implemented through a signaling protocol, such as the Resource Reservation Protocol (RSVP) extensions [3] or constraint routed label distribution protocol (CR-LDP) [4].

Path selection, which is used to select an appropriate route through the MPLS network for explicit routing. Traffic management is essential for both the service provider and Internet service provider (ISP). It is implemented by introducing the concept of constraint-based routing, which is used to compute paths that satisfy certain specifications subject to certain constraints, including constraints imposed by the operational environment [1].

State information dissemination, which is used to distribute relevant information concerning the state of the network, including topology and resource availability information. In the MPLS context, this is accomplished by extending conventional IP link state routing protocols to carry additional information in their link state advertisements [5–6].

Network Management, The five functional areas of network management for MPLS are known by the acronym FCAPS[74]:

- Fault: Network devices generate data indicating problems or matters of interest to a network manager.
- Configuration: Modifies the network in some fashion, such as creating a label-switched path (LSP). Often called provisioning in the telecoms world.
- Accounting (or billing): Enables an operator to determine usage of network resources. End users may be billed or the data may be used for accounting analysis.

- Performance: Determines whether the network is operating within required limits. This factor is increasingly critical as Service-Level Agreements (SLAs) are used by SPs to differentiate their services. SLAs are being used within enterprise networks in the form of contracts between Information Technology (IT) and the various departments. Performance analysis may also be used by network planners to decide whether infrastructure upgrades are required.
- Security. This area is increasingly critical with the growing number and level of sophistication of network attacks. The focus here is ensuring that network resources are protected from unauthorized access.

The motivation of MPLS for Traffic Engineering can be attributed to the following factors:

- Explicit label switched paths which are not constrained by the destination based forwarding paradigm can be easily created through manual administrative action or through automated action by the underlying protocols,
- LSPs can potentially be efficiently maintained,
- Traffic trunks can be instantiated and mapped onto LSPs,
- A set of attributes can be associated with traffic trunks which modulate their behavioural characteristics,
- A set of attributes can be associated with resources which constrain the placement of LSPs and traffic trunks across them,

- MPLS allows for both traffic aggregation and disaggregation whereas classical destination only based IP forwarding permits the only aggregation,
- It is relatively easy to integrate a "constraint-based routing" framework with MPLS,
- A good implementation of MPLS can offer significantly lower overhead than competing alternatives for Traffic Engineering.

3.5 Hierarchical Model

The first and foremost that occupies the top position in the hierarchy network is shown in Figure 3-11. It is the core of the network, or many call it the backbone of the network. The core layer provides interconnectivity between distribution layer devices it usually consists of high-speed devices, like high-end routers and switches with redundant links. In this layer, the data is forwarded as soon as possible by using the fastest network method and protocol (high speed). For example fast Ethernet 100Mbps, Gigabit Ethernet, FDDI or ATM. Data traffic is used in Sweden because delivery is certain and fast [75].

Secondly, below the core layer is the distribution layer, also called the workgroup layer that connects the access layer and the core layer. The main function of the distribution layer is to provide routing, filtering, and to determine the best way to handle service requests in the network. After the distribution layer determines the best path then the request is forwarded to the core layer. The core layer quickly forwards the request to the correct service. For example, how a file request is forwarded to a server and if necessary, forwards the request to the core layer. This layer usually consists of routers and multilayer switches.

The third is the access layer, also called the desktop layer. Access layers control access of users with workgroups to internetwork resources. The access layer design is needed to provide network access facilities. This layer usually incorporates Layer 2 switches and access points that provide connectivity between workstations and servers. At this layer, access control and policy can be managed, separate collision domains can be created, and implementation of port security. The Access Layer function includes Shared bandwidth; Switched bandwidth; MAC layer filtering; Micro-segmentation [75].

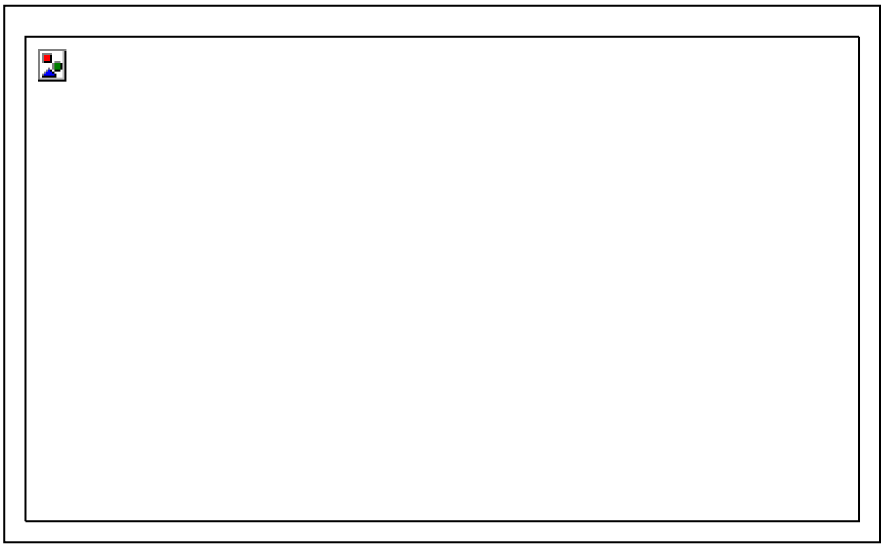


Figure 3-11. Three-Layered Hierarchical Model [75].

3.5.1 Core Network Model

In Figure 3-11 the internet has driven operation network model, which is shown below. It is comprised of TDM, IP/MPLS, and ATM. MPLS works in conjunction with the Internet Protocol (IP) and its routing protocols, such as the Interior

Gateway Protocol (IGP). MPLS LSPs provide dynamic, transparent virtual networks with support for traffic engineering, the ability to transport layer-3 (IP) VPNs with overlapping address spaces, and support for layer-2. While the underlying protocols and technologies are different, both MPLS and Asynchronous Transfer Mode (ATM) provide a connection-oriented service for transporting data across computer networks. In both technologies, connections are signaled between endpoints, the connection state is maintained at each node in the path, and encapsulation techniques are used to carry data across the connection.

Table 2. Comparison of MPLS with Frame Relay and ATM [53].

Protocols	Layer	Network Bandwidth Utilisation	Support for Traffic Tunnelling
MPLS	2.5	Full utilization	Yes (Unidirectional Tunnels)
Frame Relay	2	Underutilisation	No support for Tunnels
ATM	1	Underutilisation	Yes (Bidirectional Tunnels)

MPLS can work with variable length packets while ATM transports fixed-length (53 bytes) cells. Packets must be segmented, transported, and re-assembled over an ATM network using an adaptation layer, which adds significant complexity and overhead to the data stream. MPLS, on the other hand, simply adds a label to the head of each packet and transmits it on the network as shown in Figure 3-12. Both ATM and MPLS support tunneling of connections inside connections. MPLS uses label stacking to accomplish this while ATM uses virtual paths. MPLS can stack multiple labels to form tunnels within tunnels.

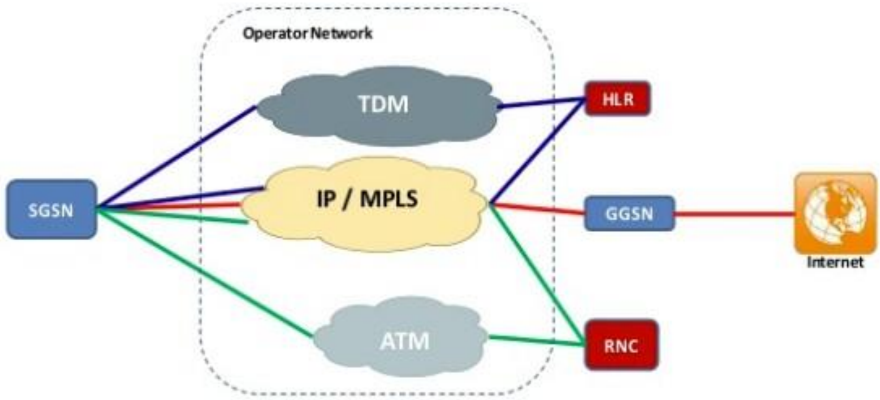


Figure 3-12. Operation Network Model [76].

The operation network model of the combined technologies can be seen in Figure 3-12 and the description of each component is as follows:

Serving GPRS Support Node (SGSN) is a main component of the GPRS network, which handles all packet switched data within the network, e.g. the mobility management and authentication of the users.

Gateway GPRS Support Node (GGSN) is a main component of the GPRS network. The GGSN is responsible for the interworking between the GPRS network and external packet switched networks, like the Internet.

Radio Network Controller (or RNC) is a governing element in the UMTS radio access network (UTRAN) and is responsible for controlling the Node Bs that are connected to it.

Home Location Register (HLR) is the main database of permanent subscriber information for a mobile network. The HLR is an integral component of CDMA (code division multiple access), TDMA (time division multiple access), and GSM (Global System for Mobile communications) networks.

Advantages and Disadvantages of MPLS

It should be observed that MPLS is not a routing protocol, but is a fast-forwarding technique which is designed to work with the existing IP routing such as OSPF or BGP [77]. Here are a few of the reasons why it might make sense for an organization to keep or adopt an MPLS network solution:

- It is a reliable technology that has served enterprises for years with direct routes from one edge to another.
- It possesses a quality connection that offers a consistent user experience with no packet loss, fixed latency, and low jitter.
- It is the safe bet for organizations that are averse to change and do not require higher-level functionality.

On the other side, here are a few of the drawback associated with MPLS:

- It can be optimized for point-to-point connectivity and not point to the cloud, meaning there is no way to directly access every cloud or software as a service (SaaS) application with MPLS. Only 2% of cloud services provide this access and with a significant premium
- It requires WAN optimization to streamline the delivery, which adds extra cost on top of a solution that can already be pricey.
- It takes a long time to deploy, especially when office locations are spread across different states or countries. It can take up to 6-8 months to get each new site up and running
- The cost is high because of the limited competition in the marketplace.

The standardization of MPLS has the goal of achieving a combination of the label switched forwarding and network layer routing [79]. It is anticipated to proffer solution to the pending issues of IP routing techniques as follows:

- Scalability issues and overheads that were associated with IP-over-ATM overlay networks.
- Enable label forwarding at expedite speeds of terabit by reducing the complexity of operation in the core network

Other benefits of MPLS are the moderate performance, better bandwidth utilization, reduced network congestion, and a better end-user experience.

3.5.2 Technical Challenges of IP/MPLS

Service providers face the challenge of controlling costs as they manage this transition. Though control of capital expenses (CAPEX) is a concern as traffic continues to grow rapidly, control of operation expenses (OPEX) is an even greater worry. Day-to-day operations and service provisioning activities are major OPEX components, and service providers seek technology that minimizes operational complexity because it is a root cause of high OPEX [78].

MPLS-TP is a set of MPLS protocols that are being defined in IETF. It is a simplified version of MPLS for transport networks with some of the MPLS functions turned off, such as Penultimate Hop Popping (PHP), Label-Switched Paths (LSPs) merge, and Equal Cost Multi-Path (ECMP). MPLS-TP extends IP/MPLS beyond the core network into the metro network and provides a reliable packet-switching transport between these networks [78].

MPLS-TP simplifies MPLS by eliminating elements of MPLS that are not necessary for a transport-oriented network. TCO analysis of MPLS-TP and IP/MPLS are shown in Figure 3-13.

The five-year TCO, CAPEX, and OPEX of the MPLS-TP solution are 55% lower than that of the IP/MPLS solution. The lower CAPEX and OPEX of MPLS-TP are attributed to the reduced complexity of the MPLS-TP protocol. This is compared to IP/MPLS and the centralized control plane approach of the MPLS-TP implementation versus the distributed control plane implementation associated with IP/MPLS routers. The simplified MPLS-TP protocol reduces the work of implementing, operating and maintaining the mobile backhaul network. The centralized control plane approach takes cost out of each network element and achieves higher economies of scale and scope via the centralized design.

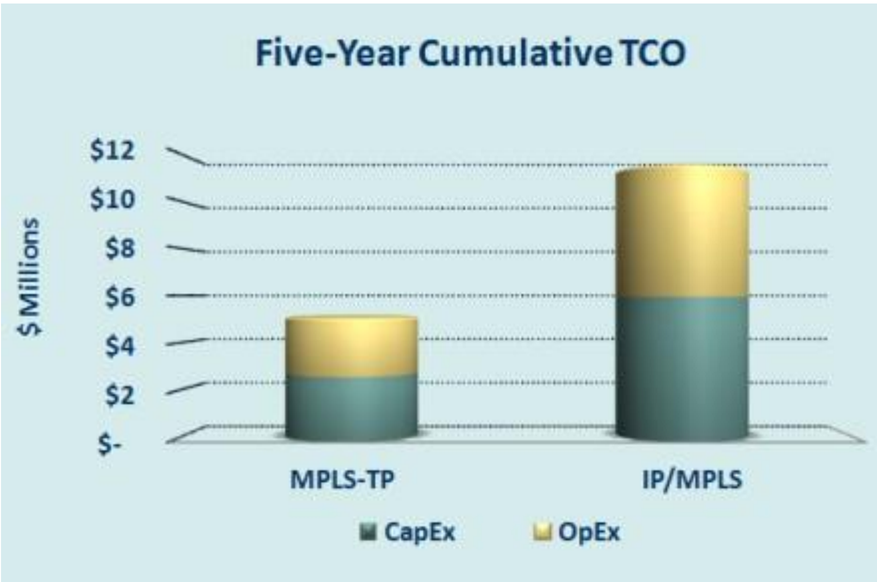


Figure 3-13. CAPEX and OPEX analysis of MPLS technology [78].

3.6 **Summary**

This chapter discussed MPLS functions and components in conjunction with the mathematical model of its operation. This is followed by a description of protocols used in MPLS such as LDP and RSVP. Subsequently, we provided an overview of MPLS technology by surveying the MPLS in combination with other technologies such as IP and VPLS in Section 3.3. It has been noted that the advancement of MPLS technologies is to optimize resource utilization of the network. The differentiated services traffic engineering of MPLS systems were discussed in Section 3.4 including bandwidth constraint models (BCM) and MPLS-TE models. Furthermore, the hierarchical model consisting of the core network, distribution layer and access layer based on cloud computing technology was discussed in Section 3.5 where the core network operation and technical challenges were presented. In the next chapter, details of bandwidth management using MPLS technology compared with IP network environment will be discussed.

4 Bandwidth Management in IP/MPLS Based Network Environment

4.1 Introduction

In this chapter, an overview of the need for bandwidth management using MPLS technology will be discussed. The design of process, node and MPLS network models will be presented. The goal of this chapter is to present the current techniques of traffic engineering used within MPLS to improve network performance.

Over the last two decades, there has been a surge in the development of new technologies, which leads to a great deal of accelerating demand from all users having access to real-time information such as data, voice, and video services. Therefore, the operators and service providers require seamless integration of network protocols with an improved QoS. Subsequently, managing network bandwidth efficiently is a top-of-mind concern for many network engineers today. With increasing bandwidth demands, network professionals are continually looking to optimize network resources, ensure adequate bandwidth, and deliver high performance.

4.2 Bandwidth Management Techniques

As mentioned in [79], when a bandwidth varies with the specific user over a period, it is called dynamic bandwidth management. On the other hand, static bandwidth management remains the same over time. In the former, the allocation of the available network resources and the adaptation of the control mechanism to support the traffic can be done by using the monitoring system to determine the user's traffic behaviour [80]. It will make a positive impact on the user due to the bandwidth constant monitoring and its accessibility. While in the latter, there is no such monitoring or control and lacks variation in the allotted network resources [79].

Bandwidth management is a dynamic approach that provides adaptability, feasibility, and efficiency for real-world network operations. Thus many schemes were contributed and designed either with a simulation study, new algorithm, or enhanced model which approximates optimized solution under widely diverse traffic load intensities [14, 81]. The research in [82] has identified that critical bandwidth used would require a mechanism for bandwidth reservation and not only for prioritization. Ideas on specific mechanism should enable bandwidth re-use such that the entire available bandwidth could be used for QoS demanding traffic and waste of bandwidth prevented [82].

Subsequently, the resource allocation controls and priority mechanisms, and the information required to support them discussed in [83]. Based on the QoS resource management method, the connection/bandwidth allocation admission control for each link in the path is performed as the status of the link required [52, 83]. Dynamic bandwidth allocation can be described as a form of bandwidth management or statistical multiplexing. This is a method by which traffic bandwidth in a shared channel can be allocated on demand and establish fairness to different users of that bandwidth.

The author provides an intriguing idea in [84] for the service providers to manage their network efficiently by improving the QoS to the customer. Further issues were also mentioned as to allocate limited bandwidth with fairness to the users and the application of network management to monitor and control the traffic of multiple applications, although, there are still a lot of controversial issues yet to be resolved such as increasing network capacity and metered pricing. Authors in [85] discuss bandwidth management in the next generation of packet networks. However, there are issues surrounding the bandwidth management for next-generation voice and multimedia over packet networks. End-to-End QoS requirements for PSTN-grade voice and multimedia service and how it might be best supported over a packet network infrastructure were investigated.

The authors in [14] presented a survey of bandwidth management in an IP based network. In addition, [86] develop ideas on a survey of Dynamic Bandwidth Allocation (DBA) algorithms. Moreover, the performance analysis of this algorithm is highlighted. Yoanes B. et al [87] pointed out that bandwidth management as a solution to data transfer for video streaming and multimedia sharing. Rural area network, virtual class and user were the main targeted areas for the bandwidth management improvement [87]. But, the specific technique of bandwidth management had not been adequately implemented. Bandwidth management is a broad aspect of different mechanisms.

A Scalable and Robust Solution for Bandwidth Allocation is proposed, according to Sridhar M. et al [88]. Admission control mechanism and recursive monitoring algorithm were used to propose an approach using scalable and robust architecture for bandwidth allocation and reservation. Lakshmanan and Qing [89] highlighted performance evaluation of dynamic adaptive resource allocation in wireless and mobile networks. There is a consideration for the existing system model by applying system level parameters to meet the expectation of this scheme.

Table 3 highlights a broad overview of the existing routing algorithms pertaining to MPLS Networks. The Authors in [90] proposed MDMF algorithms, which illustrated an improved performance than other algorithms in terms of maximizing the number of flows and minimizing the delay using label switched path mechanism.

Table 3. Survey of Routing Algorithms

Authors	Year of Publication	Algorithms	Performance Evaluation
Awduche et al [91]	2001	Minimum Hop Routing (MHA)	Create bottleneckfor future LSPs leading to network underutilization
Kodialam and Lakshman; Kar et al [92]	2000	Minimum Interference Routing (MIRA)	Cannot guarantee hop count and delay, choose the longer path leading to an increase in delay, and reject request where there are enough resources.
Kodialam and Lakshman [93]	2000	Widest Shortest Path (WSA)	Create bottleneck for future LSPs leads to network underutilization
Kodialam and Lakshman [93]	2000	Shortest Widest Path (SWP)	Create bottleneckfor future LSPs leading to network underutilization
Alidadi et al [66]	2009	Bandwidth Guarantee Routing (BGLC)	No limitation to maximum flow
Kotti et al [94]	2007	Bandwidth-Constraint Routing (BCRA)	No consideration for the ingress-egress pairs and network topology.

Gopalan et al [95]	2004	Link Critically Based Routing (LCBR	Need for previous information about traffic between ingress and egress.
Yang et al [96] **Kulkarmi et al [97]** **Van Mieghem and Kuipers [90]**	2003 2012 2004	Self-Adaptive Multi-Constraint Routing (SAMCRA)	Queue delay is affected by the bandwidth of LSPs routed in particular link
Mehdi Naderi Soorki and Habib Rostami [98]	2014	Minimum Delay Maximum Flow (MDMF)	No consideration for the bandwidth reservation for future LSPs leading the network to over utilization.
Rabah Guedrez et al [99]	2016	Segment routing-Label Encoding	Fast rerouting is not considered.
Tamara Radivilova et al [100]	2017	Minimimum cost of routing	Require optimization of network resources.
Wim Henderickx [101]	2018	Segment Routing	No Consideration for application –aware load balancer and limitation of maximum segment identifiers (SID).
Masood Mohsin and Mohamed Mostafa Fouad [102]	2018	dolphin-echolocation	Need for optimizing the bandwidth utilization.

4.2.1 Bandwidth Management in IP Environment

Bandwidth is an extremely valuable and scarce resource in any networks. Efficient bandwidth management plays a vital role in determining network performance in an organization [14, 81]. A few adaptive bandwidth reservation and algorithms have been proposed in the network to control the bandwidth consumption in the network. This is to ensure QoS guarantees for higher-priority traffic services either in IP Based network, wireless and other Next

Generations Network (NGN) communications. Under the dynamic network condition, changes, and control decisions in the proposed algorithms are made adaptively to strike a reliable network performance.

In many markets around the world, IPv6 is a requirement, and over time, this trend increased as 3G networks emerged in countries that have previously remained focussed on IPv4. From a bandwidth management perspective, the ideal solution would probably be to support MPLS in an IPv6 environment; however, today many operators are running native IPv6 networks. The IP core networks support Differentiated Services (Diffserv), to avoid the need for per-flow reservation in the core, and flows that require QoS are typically assigned to the Diffserv EF (Diffserv Expedited Forwarding) class by classifiers at the edge of the network. The IPv6 network implements priority queuing within its network elements, and EF traffic is assigned high priority queues thus guaranteeing a low loss, low latency, and low jitter.

4.2.2 Bandwidth Management in MPLS Environment

MPLS traffic engineering provides the most rigorous QoS capabilities available today in packet networks, approaching ATMs traffic management capabilities. From a bandwidth management perspective, the use of MPLS allows the complexities of core networks to be abstracted to a series of tunnels. Bandwidth management in an MPLS network is performed by using MPLS TE tunnels across the network. These form the pool of resources from which the bandwidth manager allocates bandwidth when authorizing sessions.

4.3 Proposed Bandwidth Management using MPLS technology

MPLS ensures the reliability of the communication minimizing the delays and enhancing the speed of packet transfer. It is valuable in its capability of providing Traffic Engineering (TE) for minimizing the congestion by efficient throughput. In this section, the verification of the MPLS model will be the focus of the performance evaluation. An elaborate description of MPLS and its principle of operation are already provided in section 3.2.1. It will eventually address the challenges of packet loss, high latency, high operational cost, more bandwidth utilization, and poor QoS.

The critical function of bandwidth management is to control the flow of packets on the network links to avoid traffic exceeding the capacity of the network, which would lead to congestion. This implies that more capacity requires more bandwidth. Therefore, there is an expectation that the demand for data capacity will increase a thousand-fold by 2020 [5].

Packets flow would suffer long queuing delays at congested nodes and possibly packet loss if buffers overflow [6]. To solve the problem mentioned, managing the available bandwidth would be of benefit to both the users and operators. These, in turn, play a vital role in minimizing the cost of operation rather than demanding more bandwidth, which will be very expensive. While at the same time used to monitor the effectiveness and efficiency in the performance of the network. The approach stated in [7, 26] is the purpose of supplying bandwidth on a network to reserve capacity for users. Nevertheless, the demand is low as compared with the operational capacity of the network.

4.3.1 Node and Network Design

The design of the MPLS node and network models have been implemented as shown in Figure 4-1. This is following the configuration of the individual node starting with the application and profile definitions nodes, where the parameters are set out for voice and video. Therefore, seek to obtain statistics collection of the performance behaviour of the model, then running of simulations to view the obtainable results.

The first level of planning design is to create the process model of the peripheral node with define variables, macros, and transitions. A finite state machine (FSM) implements the behaviour of a process model using the states and transitions to determine the actions to be performed [23]. The peripheral nodes proceed to the central hub by point-to-point links, which can be unidirectional, or bidirectional. The primary role of the MPLS node model is to simulate packets forwarding from one site node to another site node through packet switching technology. Looking at the nodes shown below in Figure 4-1 indicate the linking of mesh topology using LSRs as the core network. This makes a further connection to the sites through LERs both at the ingress and egress points. A source module is another essential aspect of the node model, which generate the packets. The processor assign destination addresses to the packets, sends them to a node of the point-to-point transmitter, and retrieve packets arriving from the point-to-point receiver.

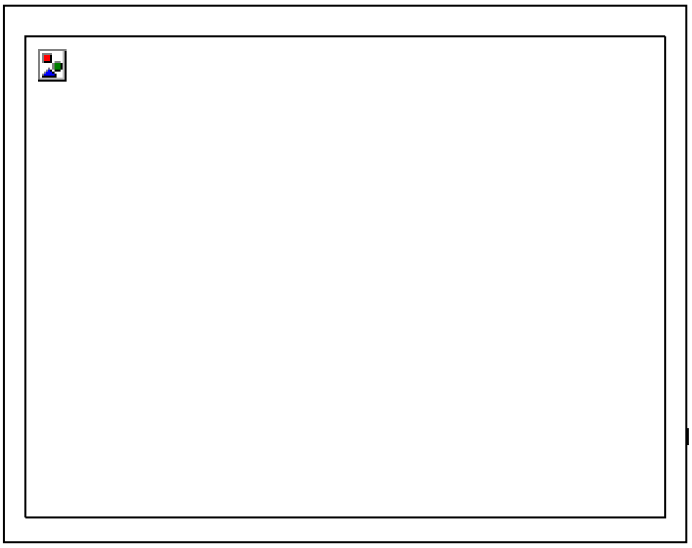


Figure 4-1. Proposed node model diagram.

The central hub node has the role in relaying packets from the receiver to a definite destination (transmitter end). There must be a direct association between many point-to-point receivers and point-to-point transmitters as shown in Figure 4-2. A source module is another essential aspect of node model, which generate the packets. As it can be seen in Figure 4-3, the processor assign destination addresses to the packets and sends them to a node of the point-to-point transmitter. Also, retrieve packets arriving from the point-to-point receiver. Furthermore, the architecture of the MPLS network mechanism is shown in Figure 4-4.

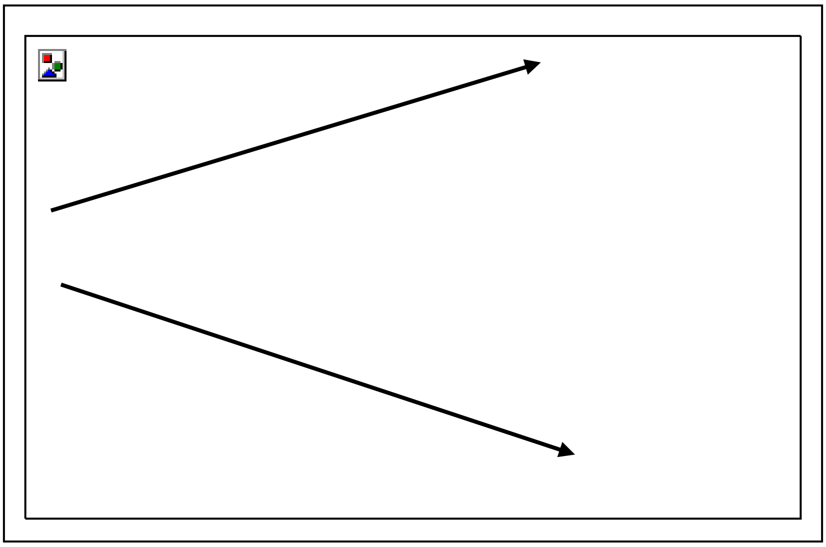


Figure 4-2. Peripheral nodes and hub node of an ingress LER

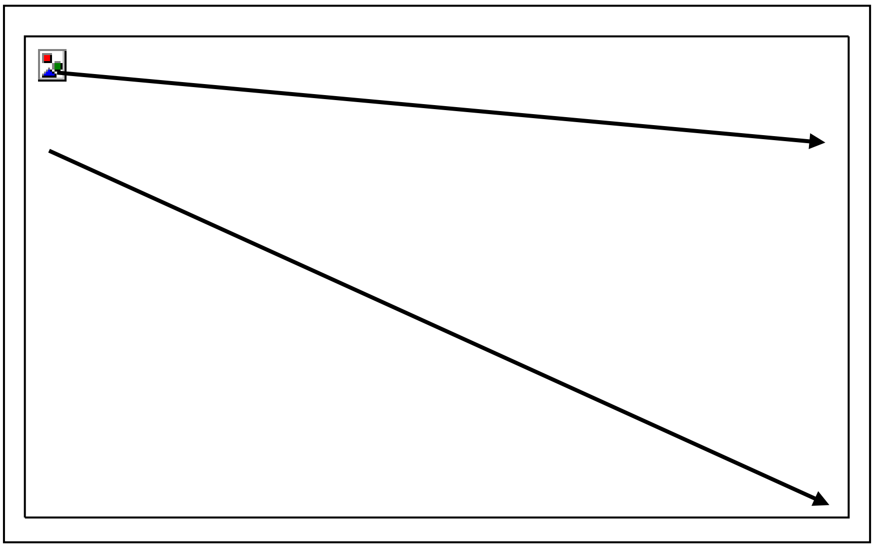


Figure 4-3. Source node, peripheral nodes, and processor of a site.

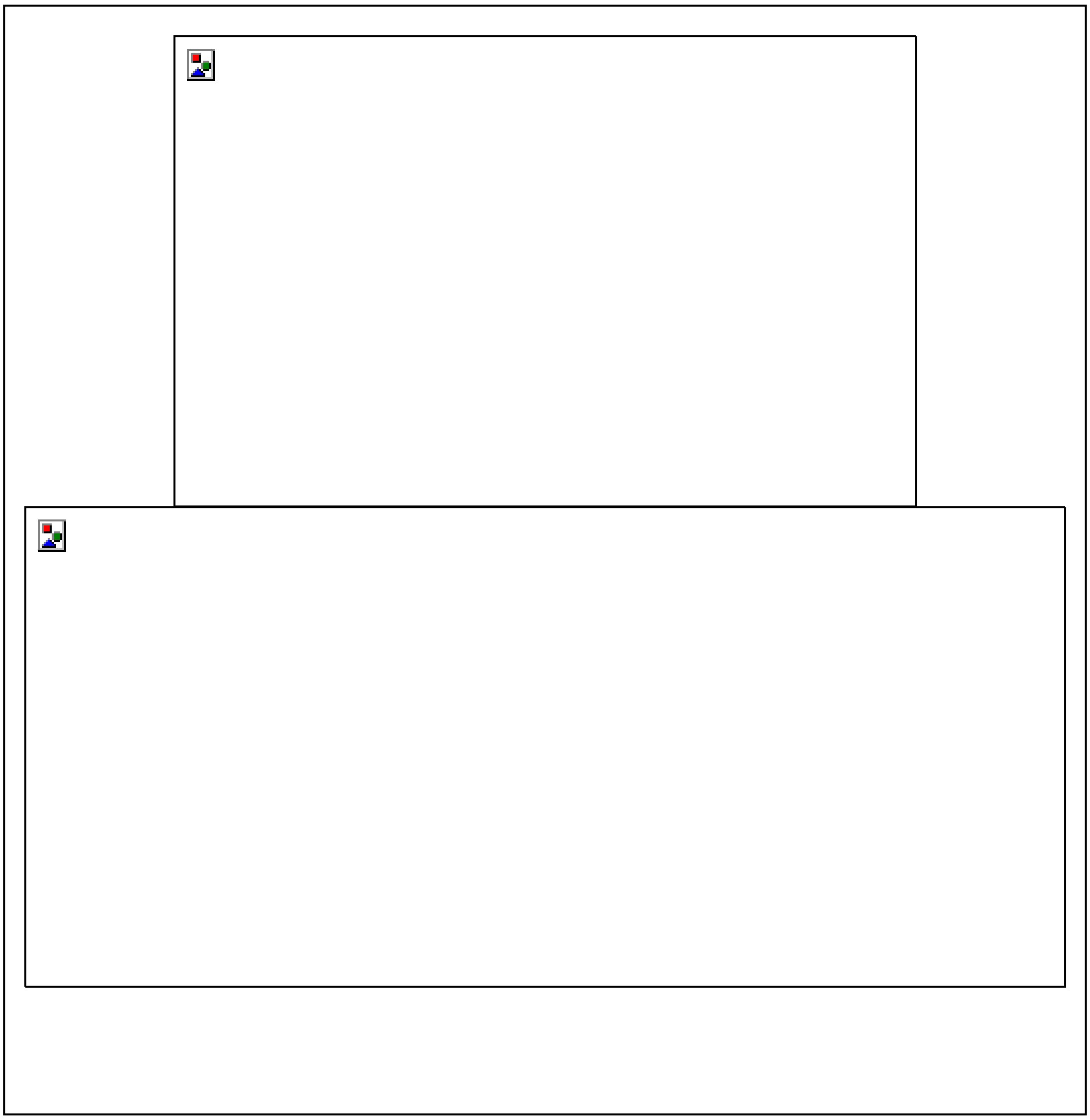


Figure 4-4. Architecture of MPLS Network mechanism.

4.3.2 Design of Dynamic and Static Label Switch Paths

OPNET tool is used to design and simulate the performance of MPLS network model. It provides a virtual network environment for the entire network models, which include its routers, switches, protocols servers, and individual applications. The goal of the simulation is to obtain results and to get an insight into other model systems by evaluating the results. MPLS model consists of configuration modules and connectivity of the nodes to generate packet-switched data transmission from point-to-point. It is designed to support the availability of resources by providing multimedia services that are sensitive to transmission to meet the requirement of the (QoS). These modules are Application Definition, Profile Definition, IP QoS Attribute Definition, MPLS Attribute Definition and Worldwide Interoperability for Microwave Access (WiMAX).

The setup of dynamic LSPs is configured manually to establish and propagate LSP information to other LSRs in the network. When the signaling protocols are enabled across the LSRs, the LSP information is transmitted throughout network. More resource utilization obtained because of the exchange and process of packets and instructions done in LSRs by dynamic LSPs than static LSPs. While Static configuration requires to explicitly configure every LSR in an LSP manually with no signaling protocol enabled. The procedure of how to configure dynamic and static LSPs is depicted in Figure 4-5.

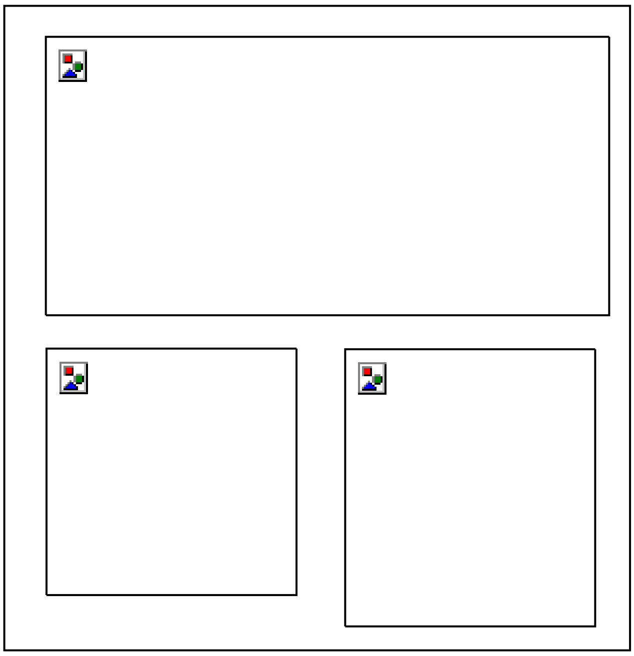
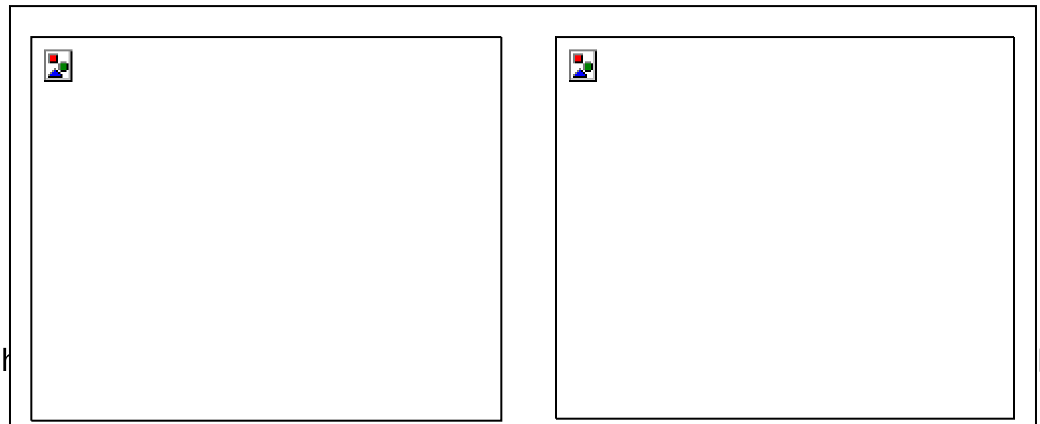


Figure 4-5. MPLS modules and flow charts of Dynamic and Static LSP models

All the routers (LERs and LSRs) along the route are defined by the LSP using MPLS_E-LSP_DYNAMIC object to provide the linkages. Then, an update of the LSP details is obtained before the simulation. This simulation uses signaling protocol (RSVP-TE) to establish an LSP from source to destination. Also, the network model for the static LSP configuration of the MPLS with the LSPs created from ingress LER1 to egress LER1 and from ingress LER2 to egress LER2. It is then compared with the scenario of the dynamic LSP configuration as shown in Figure 4-6 (a) and Figure 4-6 (b) respectively. Each connection request has a unique LSP identity (ID) assigned by either the ingress LER1 or ingress LER2.



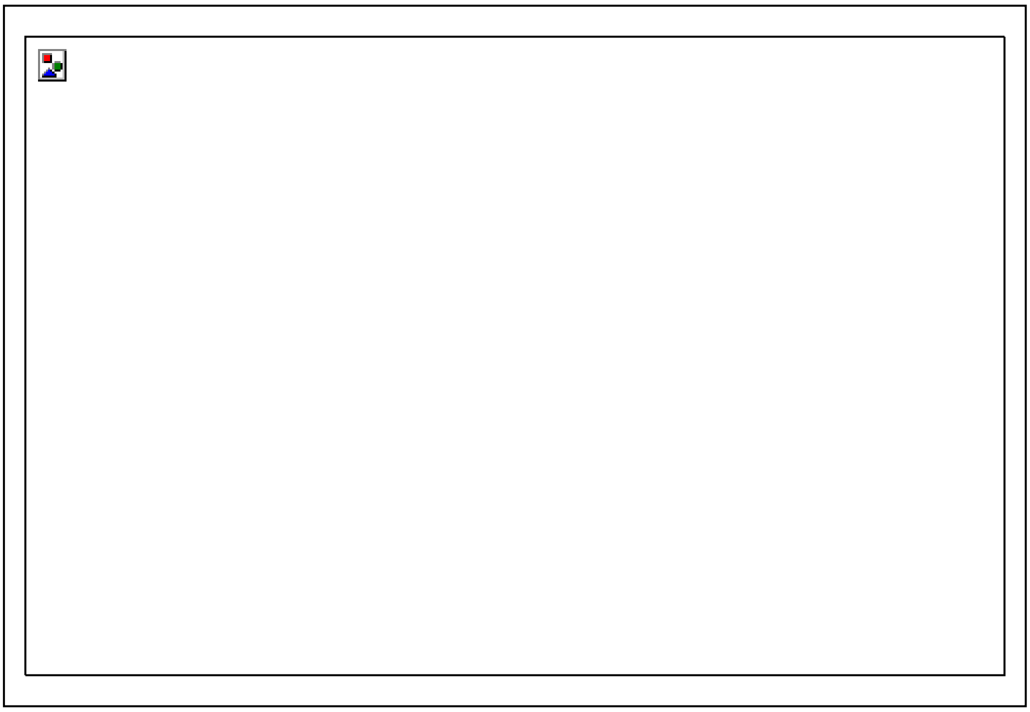
(a) Static MPLS LSP (b) Dynamic MPLS LSP

Ph ha

Figure 4-6. Static and Dynamic MPLS LSPs

4.4 Communication Service Components

Communication services are provided on a contractual basis between customers (organizations, businesses, or individual subscribers) and one or



multiple service providers. These services are to be delivered with the assurances that the exchange of information on data communication networks will be done under well-defined conditions and with guaranteed performance. A simplified communications service conceptual model is presented as shown in Figure 4-7.

Figure 4-7. Communication Service Components.

4.4.1 Quality of Service Functions

A service is a function provided by a service provider to the subscriber of that service. Quality of Service is a measure of service quality to the customer [2, 103].

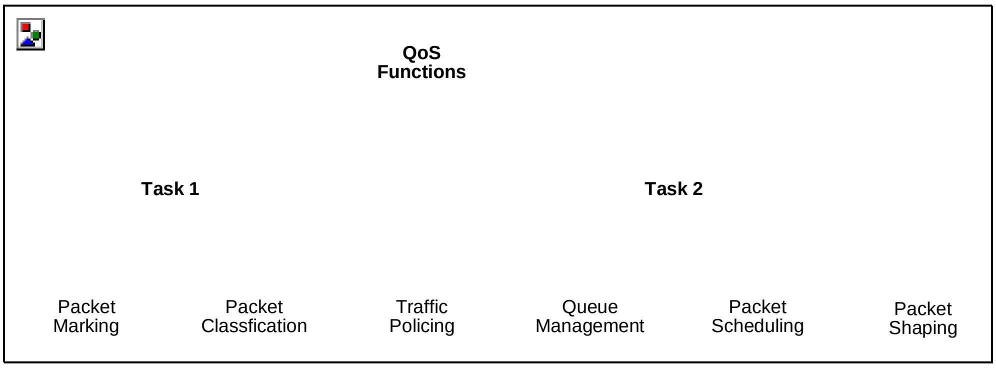


Figure 4-8. QoS functional block diagram.

In order to provide QoS over the IP network, the network must perform the following two basic tasks [104]: Firstly, the capability to differentiate between traffic or service types so that users can treat one or more classes of traffic differently than the other. Secondly, the capability to treat the different classes of traffic differently by providing resource assurance and service differentiation in a network. The sequence of the functional blocks is shown in Figure 4-8.

Packet Marking: Packet marking refer to setting the binary bits of appropriate fields in the IP header to specific values for distinguishing one type of IP packet from another. For instance, a packet may be distinguished by its source address, its destination address, or a combination of both.

Packet Classification: Packet classification is to group packets according to a classification rule. In the network, packets are selected based on the fields of the packet header that has been used for packet marking.

Traffic Policing: This is to check whether the incoming traffic at an input port conforms to the traffic rates that have been agreed upon between the customer and IP network service provider. Typically, traffic policing checks the rate of the incoming traffic concerning either a single rate referred to as the Committed Information Rate (CIR) and the Peak Information Rate (PIR).

Queue Management: This is a congestion control mechanism, and its objective is to prevent and to mitigate the TCP synchronization. The main idea of queue management is to anticipate onset congestion and act to prevent the effect of congestion. The three methods of queue management are Random Early Discarding (RED), Weighted Random Early Discarding, and Explicit Congestion Notification.

Packet Scheduling: It is a mechanism used to schedule the packets in the queues in such a way that the fixed amount of an output port bandwidth is equitably and optimally distributed among the competing classes of incoming traffic flows that are routed to that output port. Packet scheduling is at the heart of the QoS mechanism. The researchers have proposed many types of packet scheduling algorithms such First-in-First-out (FIFO), Priority Queuing (PQ), Weighted Round Robin (WRR), Weighted Fair Queuing (WFQ), and Class-based WFQ.

Packet Shaping: This is to change the rate of incoming traffic flow to regulate the rate in such a way that the outgoing traffic flow behaves more smoothly. If the incoming traffic is highly burst, it needs to be buffered so that the output of the buffer is less burst and smoother.

The authors in [105] referred to QoS in networks as a variety of measurable performance metrics such as throughput, delay, and bandwidth utilization, or too subjective notions such as fairness, anonymity, and user satisfaction. The measurement of quality is based on service parameters or QoS metrics [106, 107] [108].

4.4.1.1 Availability

Availability is the percentage of time the network, applications, or services are available when the user needs them. It is a ratio between the total time the system is used in normal conditions and a given time interval (month, year) [109].

$$A = \frac{MTBF}{MTBF + MTTR}$$

A = Availability

MTBF = mean time between failures

MTTR = mean time to restore

1. **Network Availability or Connectivity**

This is defined as the fraction of the time that network connectivity is available to a specified network ingress point and a specified network egress point. This metric has been defined by [109].

2. **Service Availability**:

It can be defined as the fraction of time the service is available between a specified ingress point and a specified egress point.

4.4.1.2 Throughput

Throughput is a measurement of the actual amount of user data/information transmitted per unit of time. Throughput depends on the amount of accompanying redundant information carried along with the transmitted data, traffic queuing/aggregation mechanisms, congestion conditions, and priority handling policies applied to data flows. It is related to data rate (always less), and it can be a fraction of the maximum available bandwidth [109].

4.4.1.3 Delay

Delay is the average transit time of packets and cells from the ingress to egress points of the network. There are the end-to-end delays and the individual delays along portions of the network. The end-to-end delay depends on the propagation rate of data in a particular communication medium (satellite or terrestrial) [109]. It can be the number and type of network elements (design, processing, switching, and buffering capabilities), routing schemas (dynamic, static, queuing, and forwarding mechanisms).

End-to-End delay = Transmission delay + Propagation delay + Processing delay + Queueing delay

$$d_{ETE} = d_{trans} + d_{prop} + d_{proc} + d_{queue} \tag{12}$$

- Propagation Delay is the time taken for a single bit of traveling from the output port on a router across a link to another router. In other words, it is the time for one bit to propagate from source to destination at propagation speed of the link. It depends on the physical medium of the link on the order of 10^{-6} seconds and negligible for two routers on the same LAN;

$$d_{propagation} = \frac{Distance}{Propagation\ speed} \tag{13}$$

- Transmission delay is the time to send out all of the packet bits. Also known as "store-and-forward" delay on the order of 10^{-6} seconds to 10^{-3} seconds and negligible for transmission rates ≥10 Mbps; significant for large packets sent over low-speed links.

$$d_{transmission} = \frac{Packet\ size}{Link\ transmission\ rate} \tag{14}$$

- Processing delay is the time required to process a packet to check for bit errors, to determine output links, at source before sending, at any intermediate router, and destination before delivering to application. It is on the order of 10^{-6} seconds or less –often negligible.
- Queuing delay is defined as the time spent waiting in a queue at any point along the route. It depends on the intensity and nature of traffic arriving at queue(s) and usually on the order of 10^{-6} seconds to 10^{-3} seconds.

4.4.1.4 Packet Delay Variation or Jitter

Jitter is one form of delay variation caused by the difference in delay exhibited by different packets that are part of the same data flow. Jitter is caused primarily by differences in queuing delays for consecutive packets and by alternate paths taken by packets because of routing decisions. For instance, Packet delay variation (PDV_k) for a packet k between source and destination is the difference between the absolute packet transfer delay (Absolute PTD_k) of the packet and a defined reference packet transfer delay, $PTD_{1,2}$, between those same measurement points:

$$PDV_k = PTD_k - PTD_{1,2} \tag{15}$$

4.4.1.5 Packet Loss

Packet loss is typically measured as a percentage of the ingress and egress traffic. Packets can be lost because they are dropped at congestion points, traffic violations (synchronization, signaling, unrecoverable errors), excessive load, or natural loss included in compression/decompression mechanisms. Voice and video communications are more tolerant to the loss of packets than in the usual data communications.

4.4.1.6 Packet Loss Rate (PLR)

The packet loss rate gives the number of voice packets that are lost in the network due to congestion. The packet loss rate is measured at the MPLS core node since the congestion is in the core network

4.4.1.7 Packet Transfer Delay (PTD)

PTD is defined for all successful and errored packet outcomes and is the time (t_2-t_1) between the occurrence of two corresponding packet reference events, ingress event at time t_1 and egress event at time t_2, where ($t_2 > t_1$) and (t_2-t_1)$\leq T_{max}$. If the packet is fragmented, t_2 is the time of the final corresponding egress event. The end-to-end packet transfer delay is the one-way delay between the measurement points at the source and destination. Therefore, Table 4 gives detail of the QoS requirement for multimedia services.

Table 4. Network QoS requirement for multimedia services [109]

QoS class	Packet Transfer Delay	Packet Delay Variation	Packet Loss Ratio	Packet Error Ratio	Packet Reorder Ratio	Typical Application
0	100ms	50ms	0.1%	0.01%	Undefined	Streaming of live content e.g. TV speech.
1	400ms	50ms	0.1%	0.01%	Undefined	Streaming of video and audio content.
2	100ms	Undefined	0.1%	0.01%	Undefined	Interactive multimedia e.g. web pages, messages
3	400ms	Undefined	0.1%	0.01%	Undefined	Download and upload of video content.
4	1s	Undefined	0.1%	0.01%	Undefined	Download of data, e-mail etc.
5	Undefined	Undefined	Undefined	Undefined	Undefined	Streaming of live TV content
6	100ms	50ms	0.0001%	0.0001%	0.0001%	Streaming of TV content.
7	400ms	50ms	0.0001%	0.0001%	0.0001%	Streaming of video content.

4.4.2 Class of Service

Class of Service (COS) is a business-based grouping of common applications and users with similar service requirements into one of several broader services or priority classes. Classes of Service are used by service providers to differentiate services along with p with performance, cost, and other offered criteria. Class of service assignment has different meaning according to the type of network environment [103].

In a packet-switched public network, such as the Internet, the initial designs were based on a “best-effort” delivery premise. Packets were given equal treatment even when the network became congested. With the advent of time-sensitive and bandwidth-sensitive applications, such as VoIP, videoconferencing, IPTV, and others, the need to handle some packets differently from others, for example when traffic congestion occurs, became apparent. In UMTS, four QoS classes are defined [33]:

Conversational Class: This class is for the most delay-sensitive traffic. This class is used for voice over IP, video conference, or any type of real-time interactive traffic. The transfer delay and delay variation are very strict. However, there are loose requirements for error tolerance.

Streaming Class: This class is used for real-time voice and video streaming applications. Because it is unidirectional, it does not have a stringent transfer delay compared with the Conversational Class. However, a maximum bound on delay variation is given to this class. There is no strict upper limit for the packet loss rate.

Interactive Class: This class is used for web browsing, database retrieval and any kind of human interaction with remote equipment's applications. Short response time is expected for interactivity thus the round-trip delay time is important in this class. This class requires a low bit error rate transport.

Background Class: This class is reserved for most delay insensitive applications. This is because the destination does not have to accept data within a certain time limit. The class is mainly used for email and database download. It requires a low bit error rate transport.

4.4.3 Service Level Specifications

Service Level Specifications (SLS) of the target level of service quality, including explanations of how the QoS metrics are collected, calculated, and reported [103]. These metrics include:

- Average availability per service period (month, year) and lowest availability;

- Average response time and percentage of time when the average response time is met;
- Average throughput;
- Guaranteed bandwidth or committed information rate;
- End-to-end latency and maximum jitter for real-time interactive voice/video communication;
- Maximum packet loss.

4.5 QoS Service Models

Internet Engineering Task Force (IETF) proposed several models for providing QoS assurances to internet-based applications. These models are called QoS models or QoS architectures. There are three major QoS models that have been proposed. These are Integrated Services (IntServ), Differentiated Services (DiffServ) and lastly Multi-Protocol Label Switching (MPLS). DiffServ is an improvement on IntServ while MPLS is a packet-forwarding scheme having advantages of both IntServ and DiffServ. The applications requirements in terms of bandwidth, delay, jitter, and loss are provided in Table 5.

Table 5. Application specific QoS requirements [110].

Applications	Bandwidth	Delay	Jitter	Loss
Email	Low	Low	Low	Medium
File Transfer	High	Low	Low	Medium
Telnet	Low	Low	High	Low
Streaming Media	High	Low	High	Low

Videoconferencing	High	High	High	Low
Voice over IP	Low	Low	High	Low

4.5.1 Traffic Models

The design of robust and reliable networks and network services is becoming increasingly difficult in today's world. The only path to achieve this goal is to develop a detailed understanding of the traffic characteristics of the network. Analysis of the traffic provides information like the average load, the bandwidth requirements for different applications, and numerous other details. Traffic models enable network designers to make assumptions about the networks being designed based on experience and enable prediction of performance for future requirements.

Traffic models can be continuous-time or discrete-time. In continuous time, we are interested in a stochastic process to represent the time-varying source rate X(t) or the set of packet arrival times $\{t_1, t_2, \ldots\}$. In traditional queueing theory, the Poisson arrival process has been a favourite traffic model for data and voice. The Poisson arrival process has several properties that make it appealing for analysis.

- It is memoryless in the sense that, given the previous arrival occurred T time ago, the time to the next arrival will be exponentially distributed with mean $1/\lambda$ regardless of T. In other words, the waiting time for the next arrival is independent of the time of the previous arrival. This memoryless property simplifies analysis because future arrivals do not need to consider the past history of the arrival process.

- The number of arrivals in any interval of length t will have a Poisson probability distribution with mean λt.
- The sum of two independent Poisson arrival processes with rates λ_1 and λ_2 , is a Poisson process with rate $\lambda_1 + \lambda_2$. This is convenient for analysis because traffic flows are multiplexed in a network.
- Suppose a Poisson arrival process with rate λ is randomly split, meaning that each arrival is diverted to a first process with probability p or a second process with probability 1-p. The first process is a Poisson arrival process with rate $p\lambda$ and the second process is another Poisson process with the rate $(1- p) \lambda$.

4.5.2 Integrated Services

The Integrated Services (IntServ) model is also known as hard QoS model. It's a model based on flows, i.e., source and destination IP addresses and ports [111]. With the IntServ model, applications ask the network for an explicit resource reservation per flow. Network devices keep track of all the flows traversing the nodes checking if new packets belong to existing flow and if there are enough network resources available to accept the packet.

By reserving resources on the network for each flow, applications obtain resources guarantees and predictable behaviour of the network. IntServ model performs deterministic Admission Control (AC) based on resources requests vs. available resources.

The implementation of this model requires the presence of IntServ capable routers in the network and uses RSVP for end-to-end resource reservation.

RSVP enables a host to establish a connection over connectionless IP Internet:

- Applications request some level of service to the network before sending data.
- The network admits or rejects the reservation (per flow) based on available resources.
- Once cleared, the network expects the application to remain within the requested traffic profile.

The scalability of this model is limited by the fact that exists a high resource consumption on network nodes caused by per-flow processing and associated state. Remember that network nodes need to maintain the reservation state for each flow traversing the node.

The fact that RSVP is a soft state protocol with continuous signaling load only aggravates the scalability problem.

IntServ Advantages:

- A good solution for managing flows in small networks.
- Intserv enables hosts to request per-flow, quantifiable resources, along end-to-end data paths and to obtain feedback regarding admissibility of these requests.

IntServ Disadvantages:

- Poor scalability.
- High resource consumption on the network nodes.
    - Per-flow processing (CPU): signaling & processing load.

- Per-flow state (memory): to keep track of every flow traversing the node.
- Continuous signaling (RSVP is a soft state protocol).

- It is very difficult to implement.

4.5.3 Differentiated Services

Differentiated Services (Diffserv) model is also known as a soft QoS model. It's a model based in-service classes and per-hop behaviours associated with each class [111]. It does not use RSVP, but instead uses Per-Hop Behaviour (PHB) to allow each router across the network to examine the packet and decide what service level it should receive? In a multi-service network environment, with best-effort priority traffic, normal priority traffic, and key-priority traffic sharing the same network, MPLS bandwidth allocation and DiffServ/priority queuing are both needed. Normal-priority and key priority traffic use MPLS capabilities to receive bandwidth allocation while the best effort traffic gets no bandwidth allocation.

Table 6. PHB and DSCP [111].

<table>
<tr><th colspan="4">Per-Hop Behaviours (PHB)</th><th colspan="3">DiffServ Code Point (DSCP)</th></tr>
<tr><td colspan="4">Expedited Forwarding (EF)</td><td colspan="3">46 (101110)</td></tr>
<tr><td colspan="7">Assured Forwarding</td></tr>
<tr><td>Class1</td><td>AF11=
001010</td><td>AF12=
001100</td><td>AF13=
001110</td><td>10
001010</td><td>12
001100</td><td>14
001110</td></tr>
<tr><td>Class2</td><td>AF21=
010010</td><td>AF22=
010100</td><td>AF23=
010110</td><td>18
010010</td><td>20
010100</td><td>22
010110</td></tr>
<tr><td>Class3</td><td>AF31=</td><td>AF32=</td><td>AF33=</td><td>26</td><td>28</td><td>30</td></tr>
</table>

	011010	011100	011110	011010	011100	011110
Class4	AF41= 100010	AF42= 100100	AF43= 100110	34 100010	36 100100	38 100110
Best Effort (BE)				000000		

All MPLS QoS implementations use the differentiated services (DiffServ), model. Routers use three bits, called Experimental bits for historical reasons, in the MPLS header of each packet transported across the MPLS network to differentiate the traffic [112, 113]. This allows eight traffic classes to be implemented; though one is usually reserved for the default traffic class, leaving only seven actual classes. Table 6 illustrates the logical operation of PHB and DSCP. Four traffic classes should be enough to cover the needs of most service providers [112].

Table 7. Mapping 6-bit DSCP field into 3bits experimental (EXP) field in MPLS [112].

		DSCP Value (6 bits)			EXP Value (3 bits)
Expedited Forwarding (EF)		**101110**			101
Assured Forwarding					
Class1	AF1	001010	001100	001110	001
Class2	AF2	010010	010100	010110	010
Class3	AF3	011010	011100	011110	011
Class4	AF4	100010	100100	100110	100
Best		000000			000

Effort (BE)		

In this case, there is no need for an explicit request for resource reservation by applications to the network. Differentiated Services is based on statistical preferences per traffic class. DiffServ allows end devices or hosts to classify packets into different treatment categories or Traffic Classes (TC), each of which will receive a different Per-Hop-Behaviour (PHB) at each hop from the source to the destination. Each network device on the path treats packets according to the locally defined PHB [114]. Table 7 shows mapping 6-bit DSCP field into 3bits experimental (EXP) field in MPLS.

PHB defines how a node deals with a TC. Network service policies can be specific to an entire QoS domain, some part of a network or even a single node. Priorities are marked in each packet using DSCP for traffic classification. This marking is performed per packet usually at the QoS domain boundary. The marking can be done at several levels of the networking layers (MPLS EXP, CoS). DiffServ model implements a statistic, class-based, AC.

DiffServ Advantages:

- Highly scalable QoS mechanism.
- Does not require any resource reservation mechanism on end hosts.
- Easy configuration, operation, and maintenance.
- Support complex traffic classification and conditioning at the edge.
- Can aggregate multiple applications flow into a limited number of TCs.
- Reduced overhead associated with the maintenance of policies on a per flow basis.

- Diffserv nodes can process traffic more easily than Intserv devices.
- Diffserv is a distributed QoS service model. Resource allocation is distributed among all the routers of a Diffserv domain, allowing for greater flexibility and efficiency in the routing process.

DiffServ Disadvantages:

- Coordination between domains in the QoS end-to-end service.
- SPs QoS customization may affect the guaranteed QoS end-to-end service.

4.5.4 Best Effort

The Best Effort (BE) QoS model is the simplest of the three. It is the QoS model used for the Internet, and it does not implement any QoS mechanism at all, that is the reason why there is not any complexity associated with this QoS model. BE does not allow for resource reservation or any other mechanism related to asking for special treatment to the network. For this reason, the BE model does not work very well will any emerging application with real-time (RT) traffic demands.

This model should not be used when the network resources are not enough to fulfill the QoS application requirements in terms of the main indicators as bandwidth, delay, jitter, etc. In these cases, with applications competing for resources, the quality of the end-user experience could be very poor if there is no other mechanism in place to manage the unfairness.

Table 8. Differences between the three QoS models

QoS Service	Best Effort	IntServ	DiffServ
Isolation	No isolation	Per flow isolation	Per aggregation isolation
Guarantee	No guarantee	Per flow	Per aggregation (Traffic Class)
Service Scope	End-to-end	End-to-end	Per domain
Complexity	No setup	Per flow setup	Long term setup
Scalability	Highly scalable	Not scalable (each router maintains per flow state)	Scalable (edge routers maintain per aggregate state; core routers per class state)
Suitable for Real Time traffic	No	Yes, resource reservation.	Yes, LLQ.
Admission Control	No	Deterministic based on flows.	Statistic based on Traffic Classes.
Applicability	Internet Default	Small networks and flow aggregation scenarios.	Networks of any size.
Resource Reservation	Not available	Per flow on each node in the source-destination path.	Per Traffic Class on every node in the domain.
Complexity	Low	High	Medium

4.6 Concluding Remarks

This chapter has looked at the various techniques of bandwidth management in an IP/MPLS environment, which fall primarily under the categories of dynamic and static bandwidth management. The critical drawbacks with the existing techniques are that they only consider monitoring and control of network traffic, do not consider bandwidth utilization with regards to different traffics in the core network. Survey of the existing routing algorithms as proposed are provided with their performance evaluation. The chapter also presented network models of proposed bandwidth management using MPLS technology with the aim of studying the performance of the network in terms of throughput and delay. Followed by the design of dynamic and static LSP models, which are compared with respect to bandwidth distribution and bandwidth utilization. Then the details of communication service components such as QoS functions and QoS metrics are discussed. Furthermore, this chapter presents traffic engineering models that include the IntServ model, the DiffServ model, and Best Effort service. Hence, there is a need for new schemes that would solve the problem of bandwidth starvation and reduce bandwidth wastage due to overutilization within the core network.

5 Proposed Intelligent based Packet Scheduling Framework

5.1 Introduction

This chapter introduces the proposed intelligent-based packet scheduling framework as it uses FIFO to serve as the baseline and then followed by WFQ to schedule non-real-time traffics. Further work used Fuzzy and Neuro-Fuzzy algorithms to schedule non-real traffics to improve on the conventional packet scheduling. Initially, Fuzzy-based algorithm is applied and then followed by Neuro-Fuzzy algorithm to improve on the conventional packet scheduling at the interface of LSR routers. In the proposed intelligent-based packet scheduling framework, this is done using the following procedures as shown in Figure 5-1.

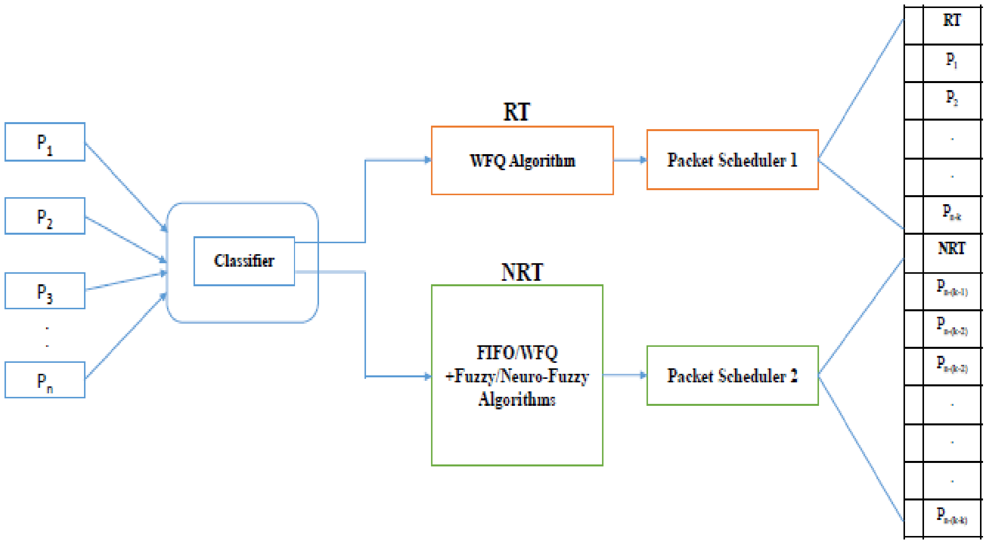


Figure 5-1. Conventional Packet Scheduling and Proposed Intelligent-based Packet Scheduling.

The traffic requires to be transmitted are divided into two groups as shown below;

- RT class - These include VoIP, live streaming, video call, and real-time gaming. This type of services is delay-sensitive.
- NRT class - These are called variable bit rate (VBR) and include buffered streaming, and transmission control protocol (TCP) based services like web browsing, email, file transfer protocol (FTP), and point to point (p2p). This type of services is delay tolerant to some extent.

In Figure 5-1, the critical blocks of the proposed framework are discussed in detail as follows: The primary function of the scheduling algorithm usually consists of two main components namely: a classifier and a scheduler. A classifier allocates packets into different queues according to their different requirements and a scheduler that selects packets to send from the queues. Hence the scheduler must be dynamic. The purpose of scheduling is to increase the utilization of resources most notably in the communication networks.

5.2 Properties of Packet Scheduling

Recently there has been tremendous growth of real-time services such as video conferencing and VoIP over packet-switched networks besides non-real-time services such as file transfer and web browsing. Packet scheduling at network nodes is a significant factor that can render service guarantees needed by applications. QoS need may vary from one application to other [99]. For real-time audio/video applications, delay, and jitter are essential QoS requirements whereas, for non-real-time applications such as FTP, throughput is a necessary QoS requirement. If the internet treats each packet equally (based on the concept of best-effort service) it will render unacceptable, if not wholly unusable service. Therefore, the modern Internet architecture should incorporate efficient scheduling mechanism that segregates packets into different service classes, allocate the suitable resources for all classes and enable the treatment of individual packets.

Scheduling algorithms are required anywhere where contention may occur. Usually, these are implemented at the output queues of each traffic class/flow. Different classes/flows of traffic need different levels of QoS from the networks.

To be more specific, most of the packet scheduling algorithm proposed is intended to achieve the following desired properties [115]:

1. Efficiency: The essential function of packet scheduling algorithms is scheduling the transmission order of packets queued in the system based on the available shared resource in a way that satisfies the set of QoS requirements of each user. A packet scheduling algorithm is generally said to be more efficient than others if it can provide a broader capacity region.

2. Protection: Besides the guarantees of QoS, another desired property of a packet scheduling algorithm to treat the flows like providing individual virtual channels, such that the traffic characteristic of one flow will have a small effect to the service quality of other flows as possible. This property is sometimes referred to as flow isolation in many scheduling contexts. Flow isolation can

significantly facilitate the system to provide flow-by-flow QoS guarantees which are independent of the traffic demand of other flows.

3. Flexibility: A packet scheduling algorithm shall be able to support users with widely different QoS requirements. Providing applications with a vast diversity of traffic characteristic and performance requirements is a typical case in the most effective integrated system nowadays.

4. Low complexity: A packet scheduling algorithm should have reasonable computational complexity to be implemented. Due to the fast-growing of bandwidth and transmission rate in today's communication system, the processing speed of packets becomes more and more critical. Thus, the complexity of the packet scheduling algorithm is also of paramount concern.

5.3 Packet Scheduling Algorithms

Packet scheduling is the core component in many recent innovations to optimize network performance and utilization. Packets entering router are classified based on their source address/port and destination address/port. After that, they are forwarded to the output interface of the router, where they experience the scheduling mechanism. It is required to take into consideration the QoS requirements of each packet flow. Packet scheduler is responsible for the order in which the packets of the various queue are dequeued and transmitted in the network. A packet flow may require a minimum throughput, maximum end-to-end delay, or maximum packet delay jitter. Scheduling of packets involves allocation of bandwidth resources and time in a manner that specific performance requirements are met. A scheduling system must be able to react to the requests within a fixed period. It is a generally attractive issue for many researchers.

In packet-switched networks, packet scheduling is an essential aspect of radio resource management. It interacts with other RRM control functions to ensure that the user QoS requirements are respected. The nature of a scheduling framework can significantly impact the QoS levels that can be provided in the system. At the fundamental level, three scheduling algorithms are referred to [116] and Broadband global area network (BGAN):

First-In-First-Out (FIFO): The FIFO is the default queueing mechanism in the absence of any specific packet scheduling algorithm. This is where packets are transmitted in the same order in which they arrived at the output link queue. As shown in Figure 5-2 , FIFO treats all packets equally and it best suitable for the best-effort network. However, it does not distinguish traffic classes, which served as the main disadvantage.

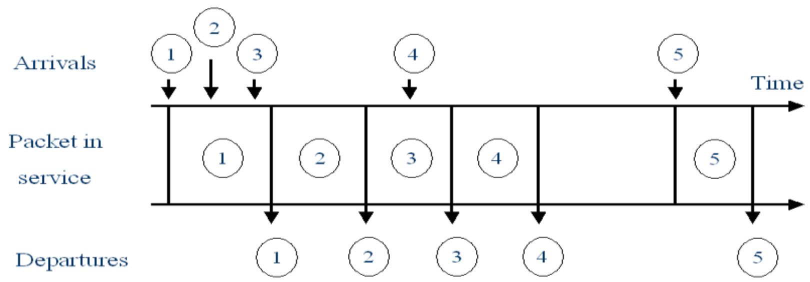


Figure 5-2. First-In-First-Out (FIFO) packet scheduling.

Priority Queuing (PQ): Packets are slotted into different queues according to their QoS requirements. These queues have different priorities and packets in the higher priority queues have a higher chance to be transmitted than the packets in the lower priority queues as shown in Figure 5-3. In PQ, N queues are created with priority ranking from 1 to N. The main drawback of PQ is the cause of starvation of low priority queues.

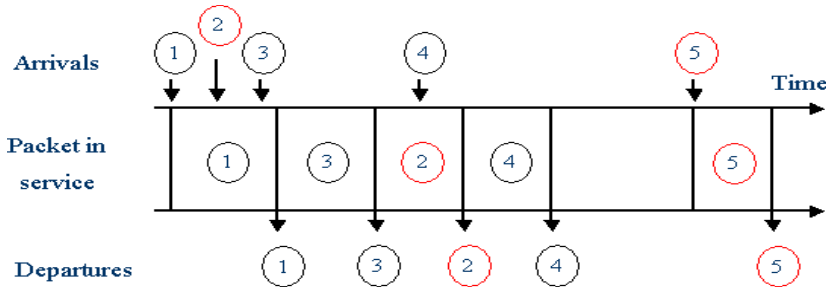


Figure 5-3. Priority Queueing (PQ) packet scheduling.

Round Robin (RR): Packets are stored in different classes. In Figure 5-4, packets of each class have the same chance to be transmitted in a scheduling period. The scheduler selects packets from among different traffic queues, and the packets in each class have the same chance to be transmitted in around.

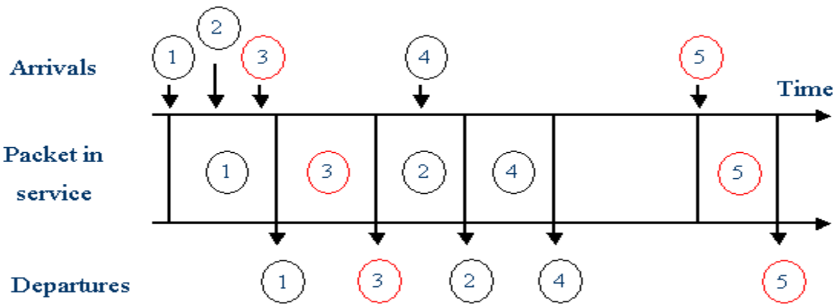


Figure 5-4. Round Robin (RR) packet scheduling.

The fairness measure of the protocol is given by the temporal average of fairness states across all departing packets. Since this metric quantifies the degree of out-of-order transmissions, a higher value implies a more significant violation of temporal fairness and, consequently, lower fairness. Let the i^{th} departing packet belong to source S_i. Let N be total number of packets and random variable $\zeta(i)$ denote the number of packets that arrived after this packet from a source different from, but departed before it as per the scheduling policy ψ [105]. Then, the Temporal Fairness-index of the scheduling policy is defined as

$$\mathrm{TFI}(\psi) = \liminf_{N \to \infty} \frac{\sum_{i=1}^{N} \mathsf{E}_{\psi}[\zeta(i)]}{N}$$

(16)

where $E_{\psi}[.]$ represents the expectation operator in the probability space determined by the arrival process and scheduling policy ψ.

Dynamically weighted low complexity fair queuing (DWLC-FQ) is proposed in [117], which is an improvement over-weighted fair queuing (WFQ) and worst-case fair weighted fair queuing+ (WF2Q+). The proposed algorithm incorporates dynamic weight adjustment mechanism to cope with the dynamics of data traffic such as burst and overload. It also reduces the complexity associated with virtual time update and hence makes it suitable for high-speed networks.

We investigate some popular fair queuing schedulers and analyze their basic properties. WRR, SFQ, and WFQ are described below.

5.3.1 WRR Algorithm

The WRR is a fundamental scheduling scheme yielding differentiated fairness among queues. It divides the output port bandwidth to input traffic classes according to their bandwidth requirement. This makes it more suitable for today's high-speed backbone networks. In WRR scheme, packets from different flows are queued in separate queues and the scheduler polls each queue in a cyclic manner in proportion to a weight pre-assigned to each queue. WRR supports flows with significantly different bandwidth requirements.

The primary limitation of WRR queuing is that it cannot provide delay bounds, unlike WFQ and WF2Q+. Secondly, it provides the correct percentage of bandwidth to each service class only if all the packets in all queues are the same size or when the mean packet size is known in advance. In WRR scheduling, each session is assigned fractional weight φ_i, and the sum of weights of all N sessions is unity or 100%. That is

$$\sum_{i=1}^{N} \varphi_i = 1 \tag{17}$$

Since the WRR server needs to determine the number of integer packets to be served from each session queue, the fractional weight is multiplied by constant integer β. The product is rounded off to next higher integer to get integer weight. So, integer weight of any session i is denoted as ω_i and is given as

$$\omega_i = [\varphi_i * \beta] \tag{18}$$

The server visits all the sessions in a predetermined order and serves ω_i packet when it visits the i^{th} service queue. The maximum length of the Round Robin cycle W is

$$W = \sum_{i=1}^{N} \omega_i \tag{19}$$

WRR guarantees fraction $\alpha_i = \omega_i / W$ of output link bandwidth to session i. If session i and j are continuously backlogged over a partial Round Robin cycle, then session i(j) can lead session j(i) by at most $\omega_i(\omega_j)$ packets. Hence proportional fairness of WRR is better than WFQ and WF2Q.

5.3.2 SFQ Algorithm

SFQ is a class of queue scheduling disciplines that are designed to allocate many separate FIFO queues. Increasing the number of queues to a large extent helps to achieve fairness. It cannot limit traffic at all; its main idea is to equalize traffic flows (TCP sessions or UDP streams) when the link is full [10]. However, managing so many separate (and dynamically changing) input classes is computationally intensive. The fairness of SFQ is ensured by hashing and round-robin algorithms.

SFQ uses a hashing algorithm which divides the traffic over a limited number of queues. It is not so efficient than other queues mechanisms, but it also requires less calculation while being almost entirely fair. It is called "Stochastic" due to the reason that it does not assign a queue for every session.

5.3.3 WFQ Algorithm

The WFQ or packetized version of GPS (PGPS) algorithm is the packet-by-packet equivalent of GPS; that is, it derives the virtual system time from the background simulation of a GPS server. The system virtual time v(t) of WFQ evolves as that of the corresponding GPS given in equation (19). WFQ schedules packets in increasing order of their virtual finish times (or departure times) [114, 118]. The virtual finish time of a packet is a function of arrival time, length of packet and weight of flow to which it belongs and can be expressed as:

$$F_i^k = max\left(F_i^{k-1}, v(t) + \frac{L^k}{\varphi_i}\right) \tag{20}$$

where i and k are indices of flow and packet respectively, v(t) is a virtual time of system, L is the packet length, and φ is the weight of flow.

The packet with the shortest finish time is chosen to send to the output port. Based on flows, WFQ classifies packets automatically, with each flow being placed into a separate queue [114, 119]. Rather than selecting all packet at the server as in WFQ, WF2Q+ system [120] considers only those packets that have started receiving services in the corresponding GPS and selects a packet that would complete service first. So, it uses both start time and finishes time of packets to emulate GPS more accurately.

In WFQ, a flow means all packets with the same source IP address, destination IP address, transport layer protocol, TCP or UDP source port, and TCP or UDP destination port. Because WFQ puts packets of different flows in different queues, it transmits a packet in its entirety before transmitting the next packet and selects the only packet that is currently in its queue It ensures that each flow gets output bandwidth proportional to assign a weight, which should add up to 100%.

Moreover, it addresses the drawback of WRR by considering the impact of the packet on bandwidth share. In the WFQ, the scheduler sends out packets from the queues based on calculated order of the packet finish time.

w_f = the (normalized) weight of flow j

p_j^i = the last packet of flow j that has arrived at the router

$V(p_j^i)$ = Virtual Finish Time of packet p_j^i

p_j^{i+1} = the next (new) packet of flow j that will arrive

A_j^{i+1} = the arrival time (real-time) of packet p_j^{i+1}

C = the transmission capacity of the communication link; b_f = the bandwidth guarantee for flow f (= reservation)

Normalized weight (guarantee) w_f:

$$w_f = \frac{b_f}{C} \tag{21}$$

$l\,(p_f)$ = the length (bytes) of packet p_f of flow f

A normalized packet length of a packet p_f:

$$\text{Normalized Packet Length}\, p_f = \frac{l(p_f)}{C}$$

Packet p_j^{i+1} starts transmission immediately after the packet p_j^i has finished transmission.

The virtual time now that packet p_j^i finishes transmission = $V(p_j^i)$

The amount of virtual time needed to transmit p_j^{i+1} is equal to:

$$\frac{L(p_j^{i+1})}{w_j} \tag{22}$$

The virtual time when packet finishes transmission is equal to:

$$V(p_j^i) + \frac{L(p_j^{i+1})}{w_j} \tag{23}$$

Therefore, the virtual time when p_j^{i+1} finishes transmission) is equal to:

$$V(p_j^{i+1}) = V(p_j^i) + \frac{L(p_j^{i+1})}{w_j} \tag{24}$$

5.3.4 PPA Algorithm

The packet processing algorithm (PPA) in MPLS is shown in Figure 5-5. Starting with the arrival of IP packets into the MPLS core routers where label can be imposed on the packet or not. Decisions are performed at the point of

LSP setup and Label Information Base (LIB) entry to determine LDP and swapping of shim header or outgoing label (olabel). This process will continue until the label is being disposed and forwarding of packets follow sequentially.

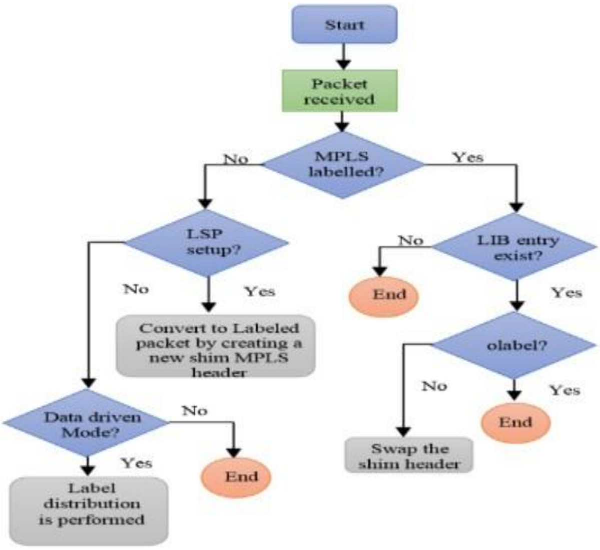


Figure 5-5. Packet processing algorithm.

5.4 Fuzzy Logic

Fuzzy logic is one of the tools of what is commonly known as Computational Intelligence (CI). CI is an area of fundamental and applied research involving numerical information processing. The idea for fuzzy logic bore in 1965. Lotfi Zadeh has published one seminar for fuzzy which was the beginning for fuzzy logic [14]. Fuzzy logic is tolerant in imprecise data, nonlinear functions and can be mixed with other techniques for different problems solving. The main principle of fuzzy logic is using fuzzy groups which are without crisp boundaries.

5.4.1 Fuzzy Logic Operation

Fuzzy logic provides a set of mathematical methods for representing information in a way that resembles natural human communication, and for handling this information in a way that is similar to human reasoning. By using fuzzy logic, a designer can blend qualitative linguistic expressions favoured by human experts in the structure of control systems. A fuzzy logic controller can be conceived as a nonlinear controller whose input-output relationship is described in linguistic terms that can be better understood and easily modified (tuned). It is independent of mathematical models of the system to be controlled, thus achieving inherent robustness and reducing design complexity.

5.4.1.1 Operation of Fuzzy Sets

There are four initial fuzzy set operations defined, namely union, intersection, complement and difference of sets. The OR (union) of $\mu_A(x)$ and $\mu_B(x)$ is denoted by $\mu_{\overline{A} \cup \overline{B}}(x)$ while the AND (intersection) of $\mu_{\overline{A}}(x)$ and $\mu_{\overline{B}}(x)$ is denoted by $\mu_{\overline{A} \cap \overline{B}}(x)$. The compliment of $\mu_{\overline{A}}(x)$ is defined as the value of membership that does not belong to $\mu_{\overline{A}}(x)$ alternatively, $1 - \mu_{\overline{A}}$. These operations are shown as follows:

Union:

$$\mu_{\overline{A} \cup \overline{B}}(x) = \mu_{\overline{A}}(x) \vee \mu_{\overline{B}}(x) \; or \; \mu_{\overline{A} \cup \overline{B}} = max\left(\mu_{\overline{A}}(x), \mu_{\overline{B}}(x)\right) \tag{25}$$

Intersection:

$$\mu_{\overline{A} \cap \overline{B}}(x) = \mu_{\overline{A}}(x) \wedge \mu_{\overline{B}} \; or \; \mu_{\overline{A} \cap \overline{B}} = min\left(\mu_{\overline{A}}(x), \mu_{\overline{B}}(x)\right) \tag{26}$$

Complement:

$$\mu_{\overline{A}}(x) = 1 - \mu_{\overline{A}} \tag{27}$$

5.4.2 Structure of Fuzzy Logic Controller

The fuzzy controller of the proposed scheme takes seven inputs: (i) bandwidth, Bw; (ii) packet delay, Pd; (iii) reliability, R; (iv) throughput, Thr; (v) packet loss rate, Plr; (vi) cost, (C); and (vii) utilization rate, Ur while the output is service or QoS decision, Serv. Below is the description of the structure of the proposed fuzzy-logic controller.

Having introduced the concepts that form the fuzzy logic theory, we summarize in this section, the use of these concepts in designing a fuzzy logic Mamdani-type control system (Mamdani & Assilian, 1975). A flow diagram of a fuzzy logic control system is given in Figure 5-6.

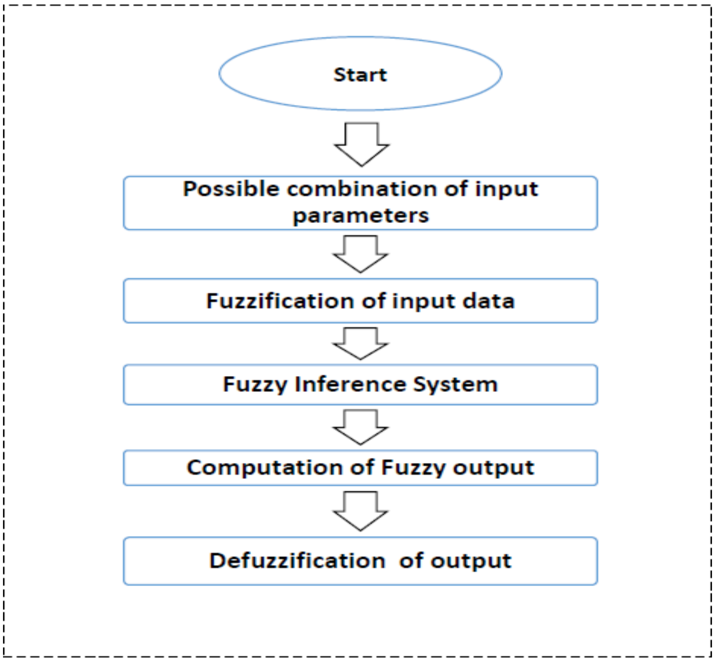


Figure 5-6. Fuzzy logic flow diagram.

The fuzzy logic controller (FLC) is composed of the following four main components:

- A rule base holds the knowledge of how best to control the system in the form of a set of IF-THEN rules. It contains a fuzzy logic quantification of the expert's linguistic description of how to achieve good control.
- An inference mechanism (also called a "Fuzzy Inference System (FIS)") emulates the expert's decision making in interpreting and applying knowledge about how best to control the system. It performs and evaluates the appropriate fuzzy operation based on the fuzzy rules to obtain a set of resultant fuzzy decisions.
- A fuzzification interface converts the fuzzy logic controller's inputs into information that the inference mechanism can use to activate and apply rules. Fuzzifier is a membership function to map the real input values of QoS criteria into fuzzy sets which have a varying degree of each membership in a set.
- A defuzzification interface converts the conclusions reached by the inference mechanism into crisp input(s) for the system. Defuzzifier converts the fuzzy decision set into a precise quantity. This crisp value can be used for deciding whether a label switch path's QoS is satisfied.

5.4.2.1 Membership Values and Functions

The membership of a fuzzy set is rather similar to the characteristic function of a crisp set. Fuzzy sets are mapped to a universe of membership values using function form. The most commonly used range of membership function is the unit interval [0, 1]. The triangular membership functions are chosen for the proposed algorithm. The membership functions for input and output linguistic parameters are shown in Figure 5-7. The values of the membership functions have been chosen based on commonly used values of membership functions in various literature. For the fuzzy controller, the membership sets (μ) for Bw, D, R, PLR, Thr, C and UR are, defined as follows:

(i) $\mu_{(Bw)}$ = {Low, Medium, High} (L, M, H)

(ii) $\mu_{(D)}$ = {Low, Medium, High} (L, M, H)

(iii) $\mu_{(R)}$ = {Low, Medium, High} (L, M, H)

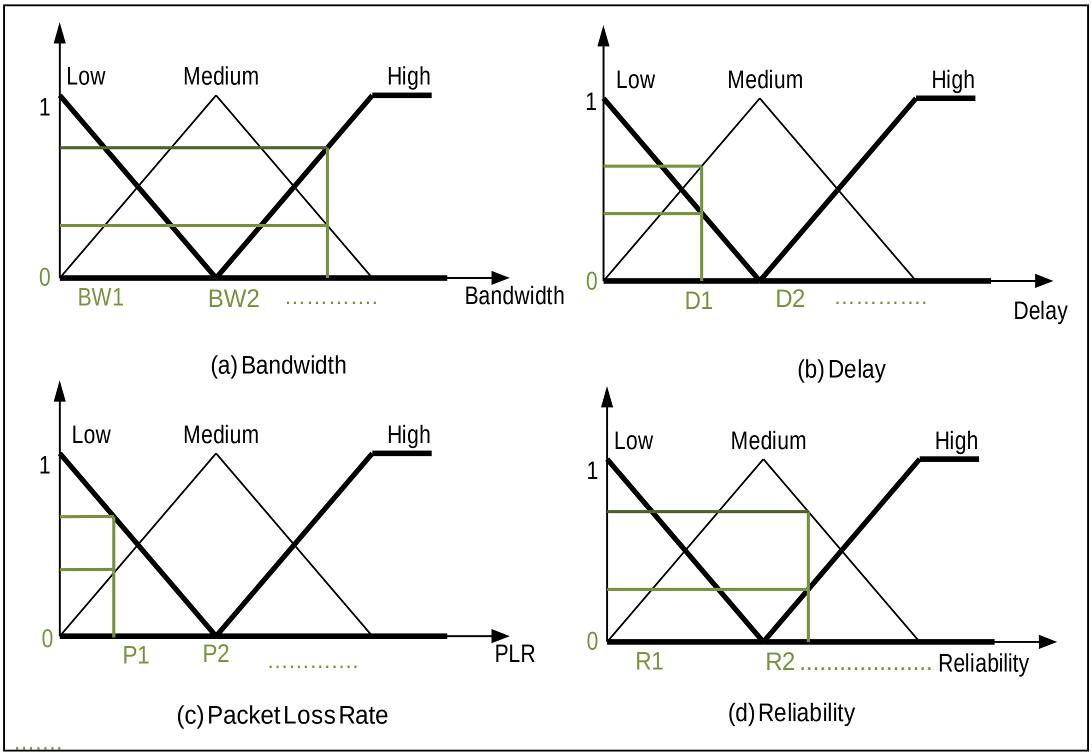
(iv) $\mu_{(PLR)}$ = {Low, Medium, High} (L, M, H)

(v) $\mu_{(Thr)}$ = {Low, Medium, High} (L, M, H)

(vi) $\mu_{(C)}$ = {Low, Medium, High} (L, M, H)

(vii) $\mu_{(UR)}$ = {Low, Medium, High} (L, M, H)

(viii) $\mu_{(Serv)}$ = {Yes, Probably Yes, Probably No} (Y, PY, PN)



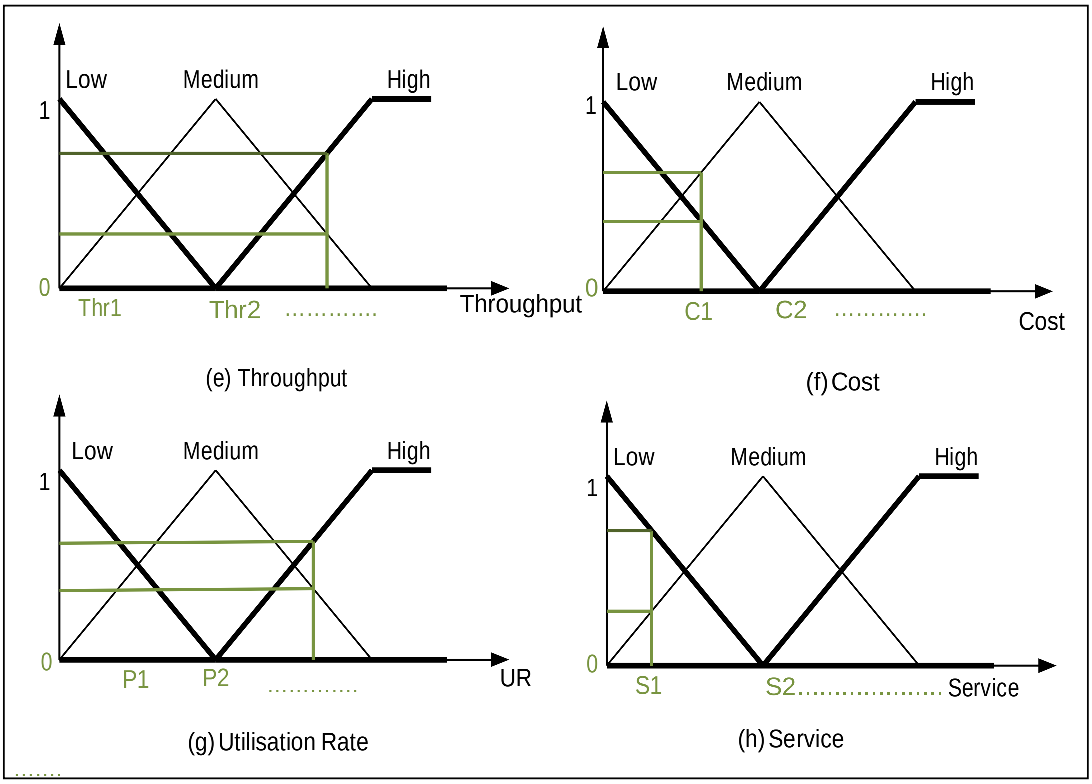


Figure 5-7. Input parameters and Membership functions

In the first stage, the values of QoS criteria are fed into the fuzzifier, which maps the periodical measurements into fuzzy sets. Fuzzy sets contain elements that have a varying degree of membership in a set. This converts a crisp value into imprecise values. For example, the delay can only be presented as either low or high in a crisp set, whereas in a fuzzy set it can be presented as low or acceptable, medium or tolerable and high or intolerable. The values of the fuzzy sets can be obtained by mapping the measurements onto a membership function, which is a curve or line. An example of how membership functions for bandwidth, delay, reliability, throughput, packet loss rate, cost, and utilization rate can be defined as shown in Figure 5-7 (a), (b), (c), (d), (e), (f) and (g) respectively. Three membership sets are defined for each membership function, namely the low set, the medium set, and the high set. In all the figures, the vertical axis represents the membership value, which lies in the range between 0 and 1, and the horizontal axis represents the measured values of the QoS parameters (bandwidth, delay, reliability, throughput, packet loss rate, cost, and utilization rate.

In terms of throughput as shown in Figure 5-7 (e), the lower the measured value, the higher its membership value will be in the low set and the lower its membership value in the medium set. Thus, in Figure 5-7 (a), when the measured bandwidth value Bw1, the membership value of Bw1 will be in the low set but the membership value of Bw1 will be in the high set. This indicates that there is a high probability that the measured bandwidth value belongs to the low set but a low probability of it belonging to the high set.

5.4.2.2 Fuzzy Rule base

It is the appropriate fuzzy operation based on the fuzzy sets that used IF-THEN rules. The fuzzy rule base consists of a series of fuzzy rules from the rule base to admit the suitable selected path with the best service to a suitable network. These predefined set of rules are the series of 'IF-Then' statements of rules. As shown above, there are seven different input parameters, and each input parameter has three-member functions such as low, high and medium. There are 3^7 (2187) rules in the rule base which are being utilized by the inference system of the fuzzy controller. Assuming that there are 'r' "IF-Then" rules, and each fuzzy condition in set A_k consists of seven input elements $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$. These control rules are of the following form: IF 'condition,' THEN 'action.'

5.4.3 Defuzzification Methods

The input for the defuzzification process is a fuzzy set (the aggregated output fuzzy set), and the output is a single number. As the aggregated fuzzy set encompasses a range of output values, a single output value must be resolved from the set. Thus, the aggregated fuzzy set constitutes a range of the maximum membership values of those implied fuzzy sets. The final crisp output result is the process of defuzzification, which converts the fuzzy reasoning output, which is a fuzzy set, into a crisp value that represents the whole inference process outcome. There are various methods for defuzzification purposes such as centroid method, bisector method, weighted average method and middle, smallest and largest of maximum methods also known as Middle of Maximum (MOM), Smallest of Maximum (SOM) and Largest of Maximum (LOM) respectively. The most popular method is the centroid method, which returns the centre of the area under the curve that represents the aggregated output fuzzy set. The centroid defuzzification method has one drawback which is computationally intensive.

$$p = \frac{\sum_{i=1}^{k} y_i \mu_C(y_i)}{\sum_{i=1}^{k} \mu_C(y_i)} \tag{28}$$

where $\mu_C(y)$ is the membership degree of y in the aggregated output fuzzy set C.

In this thesis, the weighted average method (Equation (28)) is used in the proposed fuzzy-based service-aware LSP selection technique as it gives results very close to the centroid method and requires fewer calculations or computation resources as compared to the centroid method. Figure 5-8 to Figure 5-15 illustrate the fuzzy LSP selection or packet scheduling decision for specific fuzzy rules in the fuzzy logic controller using different fuzzy input variables on the x and y-axis of the 3-D graph. These figures show how the value of service-aware LSP selection factor varies concerning the variations in the different input parameters.

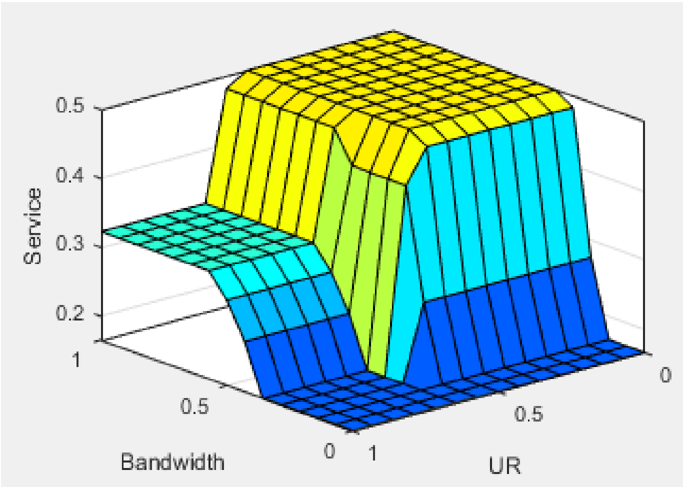


Figure 5-8. Bandwidth versus Utilization Rate (UR).

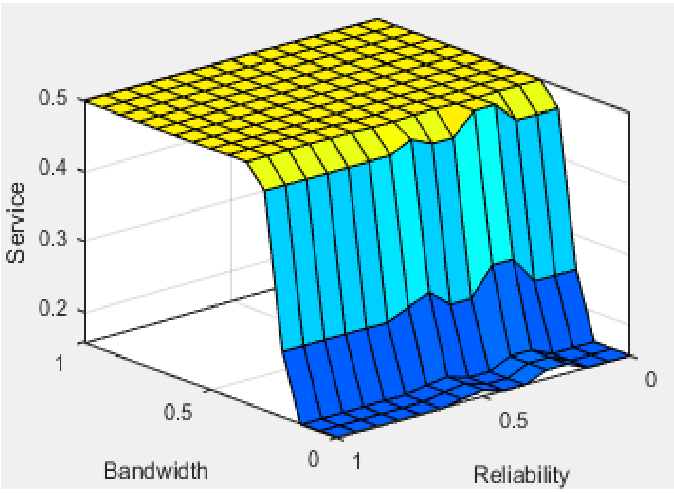


Figure 5-9. Bandwidth versus Reliability

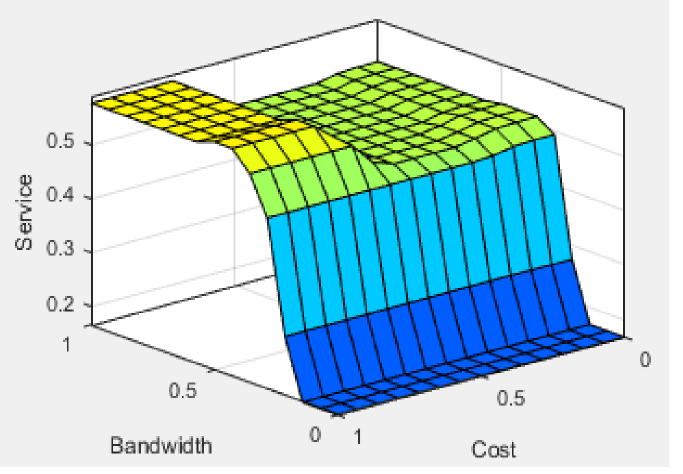


Figure 5-10. Bandwidth versus Cost

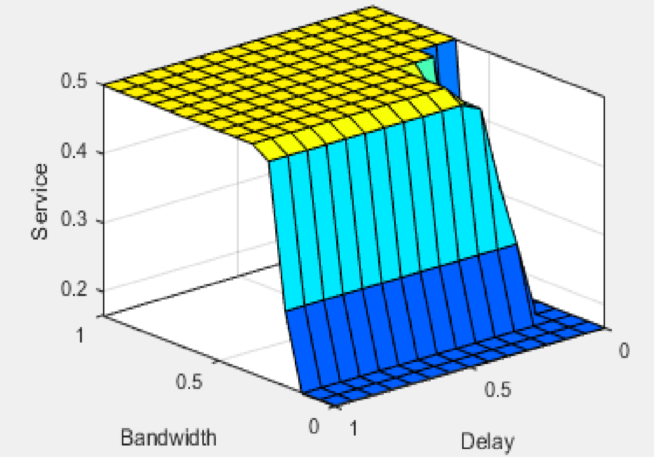


Figure 5-11. Bandwidth versus Delay

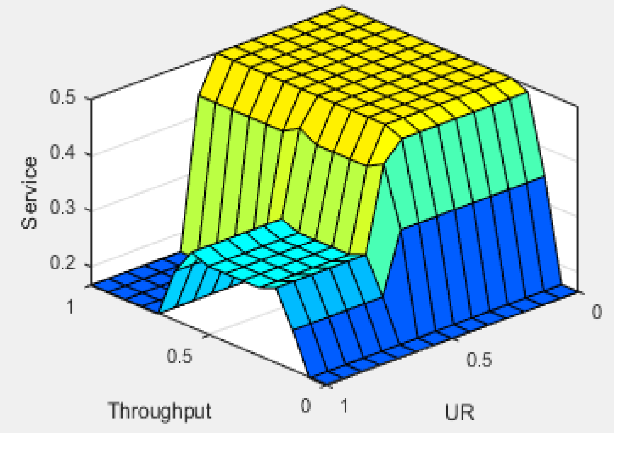


Figure 5-12. Throughput versus Utilization Rate (UR)

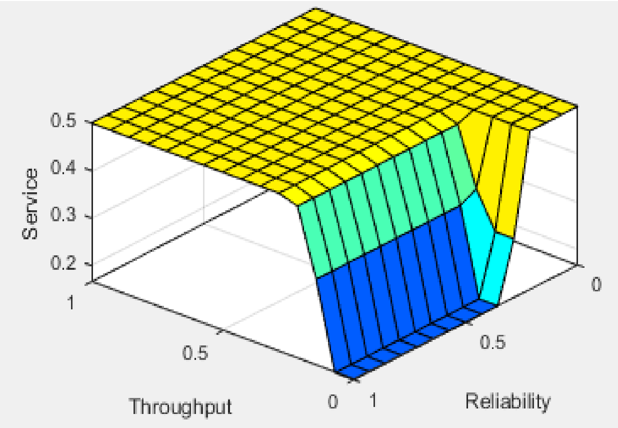


Figure 5-13. Throughput versus Reliability

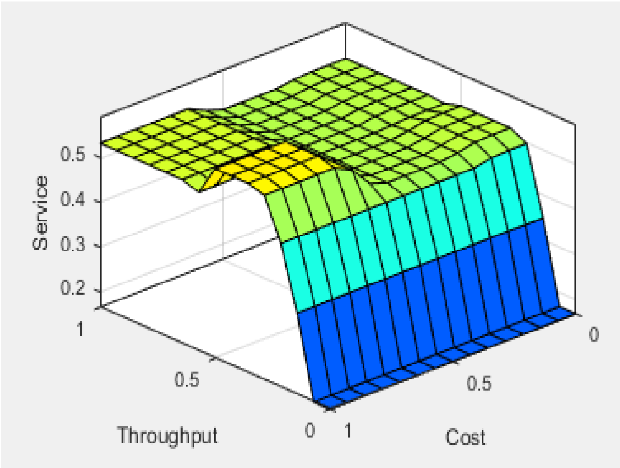


Figure 5-14. Throughput versus Cost

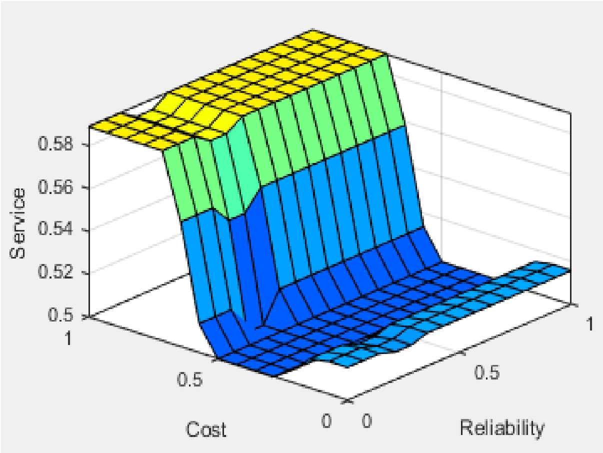


Figure 5-15. Reliability versus Cost

5.5 Proposed Fuzzy and Neuro-Fuzzy based Packet Scheduling Schemes

The fuzzy scheduling algorithm utilizes the fuzzy logic concept to derive the operation of QoS satisfaction. The core part of the fuzzy expert system is the Fuzzy Logic Controller (FLC). It periodically collects the information of each path on its user profile and perceived QoS. Based on the information, a decision can be made by the FLC to identify whether a path can give maximum service satisfaction or minimum service satisfaction. Seven different criteria are used: bandwidth, delay, cost, reliability, throughput, packet loss rate, and utilization rate. The functional entities inside the FLC are introduced as follows:

To identify an LSP's QoS satisfaction, the operation of the FLC is separated into three different stages: Fuzzification, Inference, and Defuzzification.

Based on the vast experience of successful implementations of FLC in the design of control algorithms, as indicated above, the reported strength of fuzzy logic in controlling complex and highly nonlinear systems had been used in the IP world. To the best of our knowledge, fuzzy logic, in the concept of active queue management in TCP/IP networks, has been firstly introduced for providing congestion control by [121]. This novel research demonstrated that the application of fuzzy control techniques to the problem of congestion control in TCP/IP networks is worthy of further investigation. The main reasoning can be attributed to the difficulties in obtaining a precise enough mathematical model, using conventional analytical methods, while some intuitive understanding of congestion control is available.

Fuzzy logic control has also been used, beside Active Queue Management (AQM), in other fields concerning today's Internet. Mallapur in [122] has proposed fuzzy-based bandwidth management for wireless multimedia networks to maintain the aggregate loss QoS.

Figure 5-16 shows the proposed Fuzzy based Packet Scheduling framework diagram consisting of various modules as described below. The main objective of this work is to propose an intelligent system for packet scheduling in an MPLS network that makes use of fuzzy control. Fuzzy control is based on a relatively simple idea of a fuzzy set that is a generalization of an ordinary set by allowing a degree of membership for each element. The scheduler can be implemented as a fuzzy controller with inputs such as bandwidth, delay, cost, reliability, packet loss rate, and resources utilization rate. The output defines the LSP that queue from which the next packet will be transmitted. The implementation of fuzzy-based packet scheduling using Mamdani FIS system is shown in Figure 5-17.

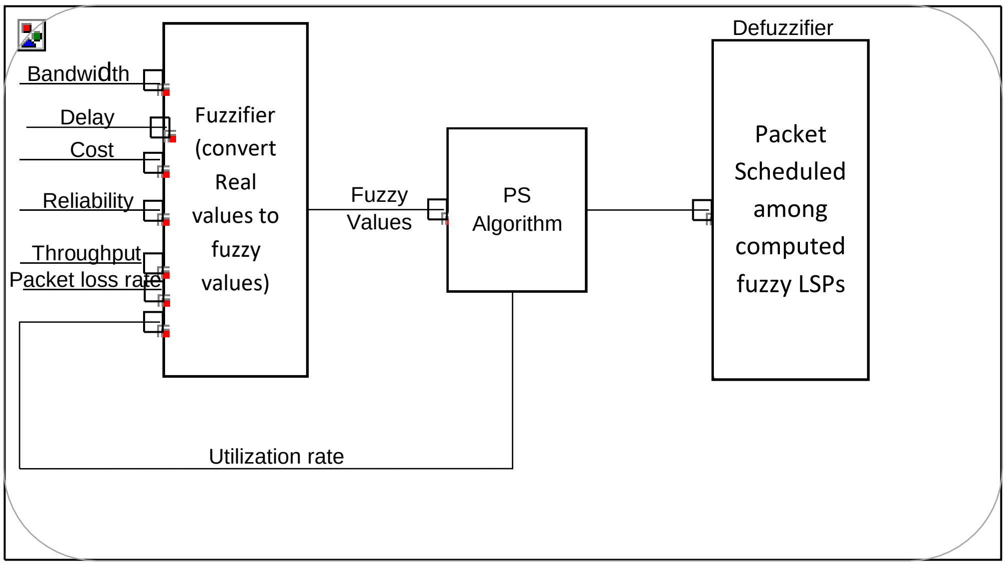


Figure 5-16. Proposed Fuzzy based Packet Scheduling framework.

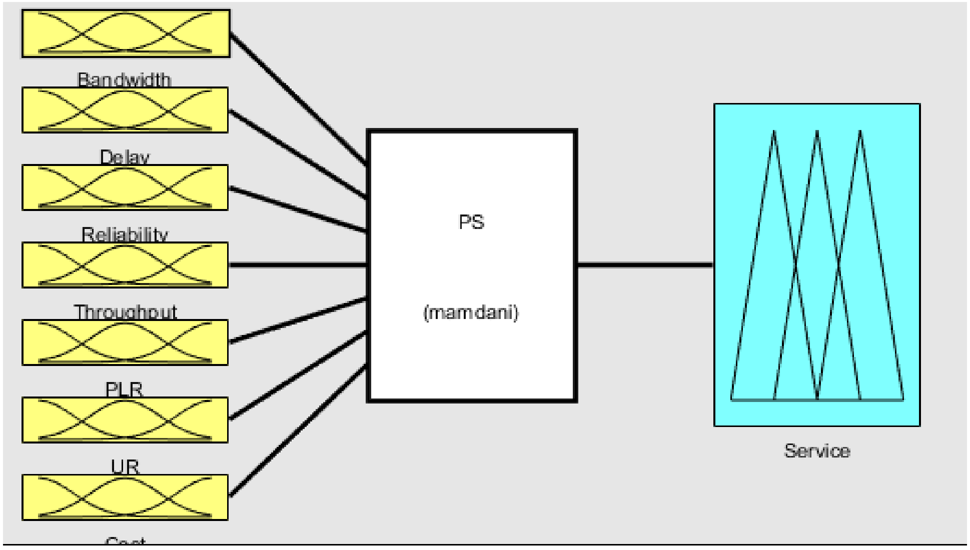
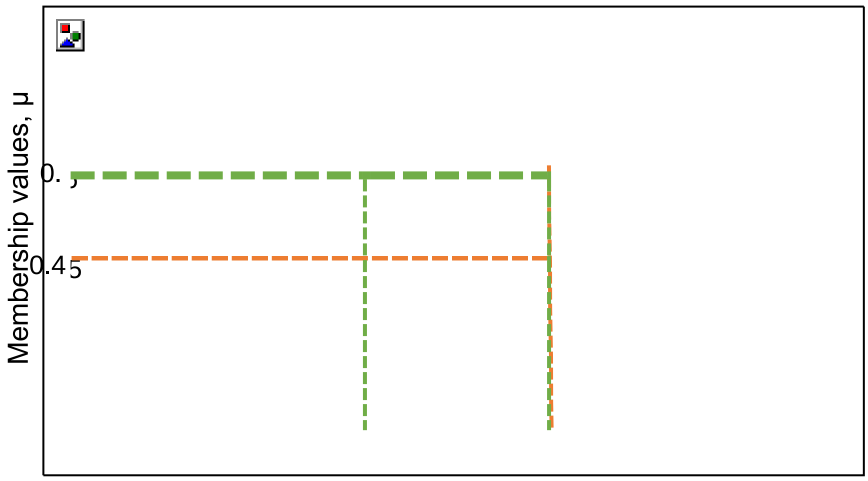


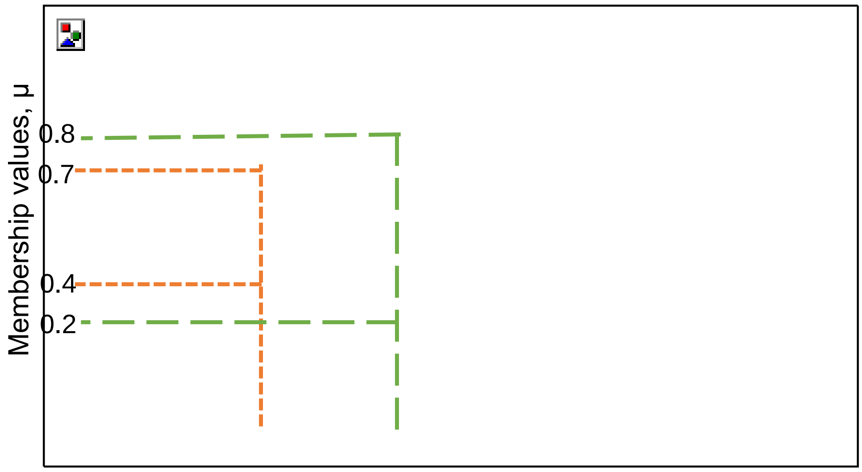
Figure 5-17.Implementation of Fuzzy based Packet Scheduling using Matlab.

5.5.1 Service-Aware LSP Selection Example

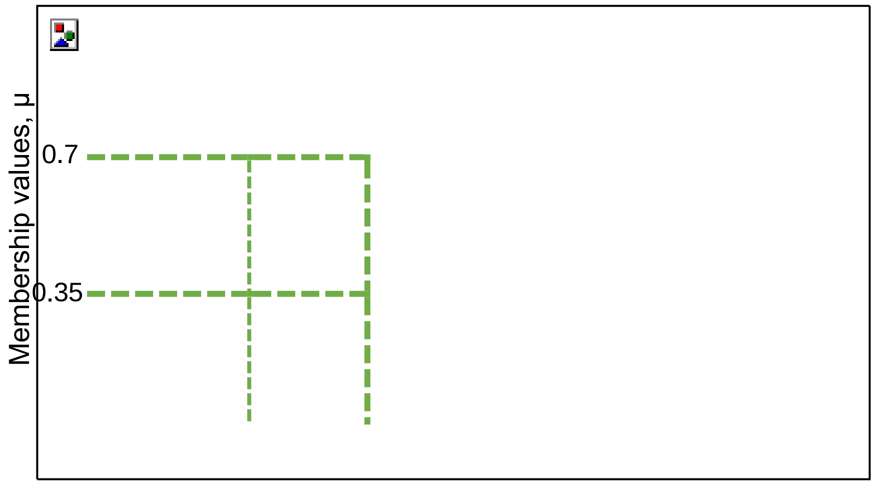
In this section, the procedure of fuzzy logic-based Service-Aware LSP selection is described with the given an example. Assume that there are two new paths detected by the source node which are, Label Switch Path 1 (LSP1) and Label Switch Path 2 (LSP2).



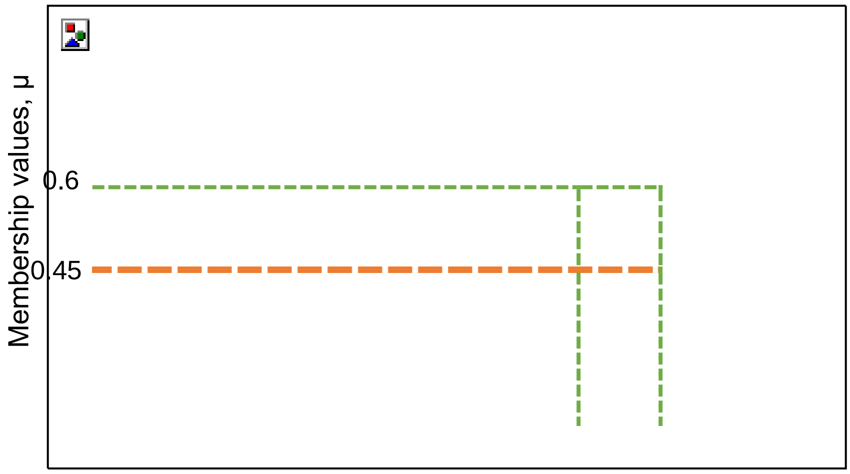
(a) Bandwidth (Mbps)



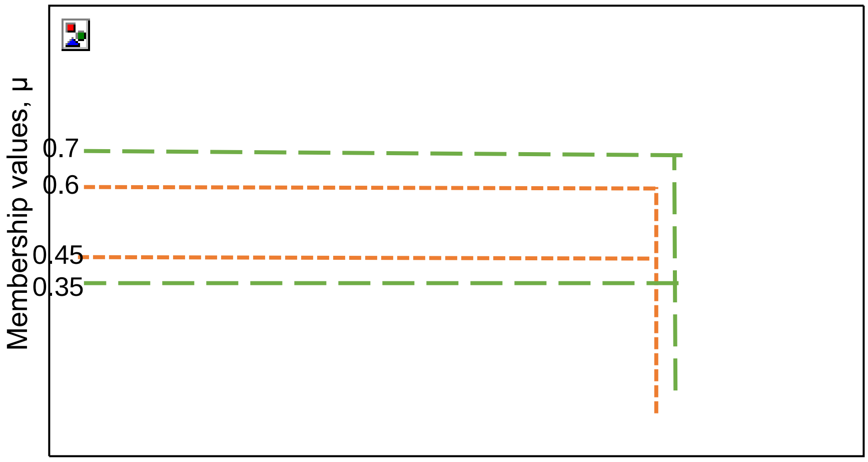
(b) Delay (ms)



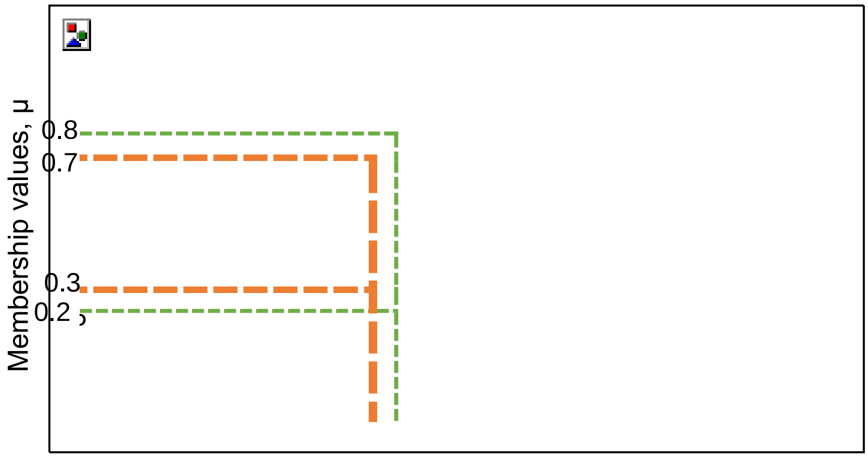
(c) Cost



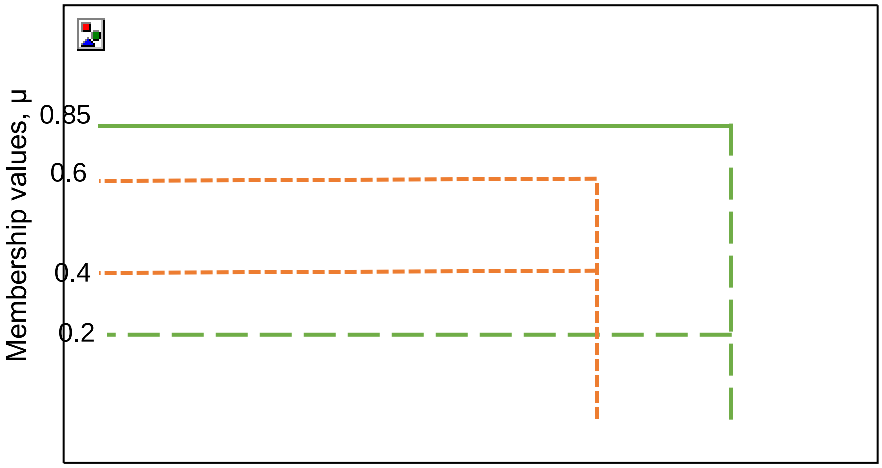
(d) Reliability



(e) Throughput (kbps)



(f) Packet loss rate (%)



(g) Utilization rate

Figure 5-18. Membership functions for (a) Bandwidth, Bw (b) Delay, D (c) Cost, C (d) Reliability, R, (e) Throughput, Thr, (f) Packet loss rate, Plr and (g) Utilisation rate, Ur.

The membership function values for each fuzzy input variable of these paths, shown from Figure 5-18 (a) to Figure 5-18 (g). The membership values of LSP1 fuzzy input variables are shown in green lines with large dashes, and that of LSP2 are shown in orange lines with small dashes. At the router node, all the input parameters values are fuzzified, and their degree of memberships have been measured for all three membership functions such as L, M and H. In this example, LSP1 and LSP2 fuzzy input values with their membership degrees are shown in Table 9 and Table 10 respectively.

Table 9. Degree of Membership Function for LSP1.

Criteria	Low (L)	Medium (M)	High (H)
Bandwidth (Bw)	0	0.45	0.6
Delay (D)	0.2	0.8	0
Cost (C)	0.35	0.7	0
Reliability(R)	0	0.45	0.6
Throughput (T)	0	0.35	0.7
Packet loss rate (PLR)	0.25	0.8	0
Utilisation rate (UR)	0	0.2	0.85

Table 10. Degree of Membership Function for LSP2.

Criteria	Low (L)	Medium (M)	High (H)
Bandwidth (Bw)	0.6	0.45	0
Delay (D)	0.7	0.4	0
Cost (C)	0.35	0.7	0
Reliability(R)	0	0.45	0.6
Throughput (Thr)	0	0.45	0.6
Packet loss rate (PLR)	0.3	0.7	0
Utilisation rate (UR)	0	0.6	0.4

Once the membership values of all input variables have been assigned, then the set of these measured membership values are compared against the logical lookup table of r rules in the fuzzy rule base. As there are seven different input variables and for this example, each having two different membership values, therefore results of IF-Else rules gives us $2^7 = 128$ different combinations. In this scenario, the UNION operation of the fuzzy set will be used in determining the Service-aware LSP selection factor. In this example, the Average Weighted method has been utilized to defuzzify the obtained fuzzy decision values. Eventually, to obtain the LSP selection factor, there is a need to construct another weighting matrix which defines the weighting of each decision element. The maximum of each decision values would be considered, i.e. Yes (Y), Probably Yes (PY), and Probably No (PN). According to the last column for LSP 1 in Table 11, there are six (Y), seven (PY), and three (PN). Then, maximum decision values are: 0.2, 0.25, and 0.7 respectively. If the weightings assigned to each possibility are Y = 0.61, PY = 0.8, and PN = 0.21 as shown in Figure 5-19 (a). While the last column for LSP 2 in Table 12, there are six (Y), seven (PY), and three (PN). This is followed by maximum decision values: 0.6, 0.45, and 0.4 respectively. If the weightings assigned to each possibility are Y = 0.8, PY = 0.76, and PN = 0.38 as shown in Figure 5-19 (b). Packet scheduling among selected paths is a decision-making problem in a fuzzy network environment. Therefore, LSP2 having 67% outperforms than LSP1 with 39% in terms of service satisfaction with an increase of 28%.

In this example the Z^* is the LSP selection factor, $\mu_c(z)$ is the membership value of each element in decision set, and z is the weight assigned to each particular decision element such as Y, PY and PN respectively. Substituting these values to the above equation gives us the following value as the LSP selection factor.

Service satisfaction factor for LSP1 = $\frac{(0.2 \times 0.61) + (0.25 \times 0.8) + (0.7 \times 0.21)}{0.2 + 0.25 + 0.7} = 0.39$

Service satisfaction factor for LSP2 = $\frac{(0.6 \times 0.8) + (0.45 \times 0.76) + (0.4 \times 0.38)}{0.6 + 0.45 + 0.4} = 0.67$

The second stage involves feeding the fuzzy sets (membership values) into an inference engine by applying the fuzzy rules to obtain fuzzy decisions. The fuzzy rules can be defined as a series of IF-THEN rules, which decides whether a connection’s QoS is satisfied. An example table of IF-ELSE rules (used in the simulation model) is shown in Table 13. If the decision criterion is based on four QoS parameters, a total of up to 81 fuzzy rules can be implemented. The higher the number of rules, the more complex the algorithm is. It may be more appropriate to implement fewer rules as long as the results are within the system’s requirements.

Therefore, Table 11 and Table 12 show the results obtained from IF-ELSE rules and Mamdani fuzzy inference system (FIS) are used for LSP1 and LSP2. Due to the computational complexity, four out of seven input parameters would be used. These are Throughput (T), Delay (D), Packet Loss Rate (PLR), and Utilisation Rate (UR). The theory of Union and intersection are used as the operations of fuzzy set in this respect, which are illustrated in equations (25) and (26) respectively. The operation of intersection applies to the rows (“Min” method) in both tables. While the operation of the union would be applied to the last column (“Max” method) for final membership values in these two tables.

Table 11. Result of IF-ELSE rule for LSP1.

Thr	D	PLR	UR	Decision	Final Membership Values
M(0.35)	L(0.2)	L(0.25)	M(0.2)	PY	0.2
M(0.35)	L(0.2)	L(0.25)	H(0.85)	PY	0.2
M(0.35)	L(0.2)	M(0.8)	M(0.2)	Y	0.2
M(0.35)	L(0.2)	M(0.8)	H(0.85)	Y	0.2
M(0.35)	M(0.8)	L(0.25)	M(0.2)	PY	0.2
M(0.35)	M(0.8)	L(0.25)	H(0.85)	PY	0.25
M(0.35)	M(0.8)	M(0.8)	M(0.2)	PN	0.2
M(0.35)	M(0.8)	M(0.8)	H(0.85)	PN	0.35
H(0.7)	L(0.2)	L(0.25)	M(0.2)	Y	0.2
H(0.7)	L(0.2)	L(0.25)	H(0.85)	Y	0.2
H(0.7)	L(0.2)	M(0.8)	M(0.2)	Y	0.2
H(0.7)	L(0.2)	M(0.8)	H(0.85)	PY	0.2
H(0.7)	M(0.8)	L(0.25)	M(0.2)	PY	0.2
H(0.7)	M(0.8)	L(0.25)	H(0.85)	PY	0.25
H(0.7)	M(0.8)	M(0.8)	M(0.2)	Y	0.2
H(0.7)	M(0.8)	M(0.8)	H(0.85)	PN	0.7

Table 12. Result of IF-ELSE rule for LSP2.

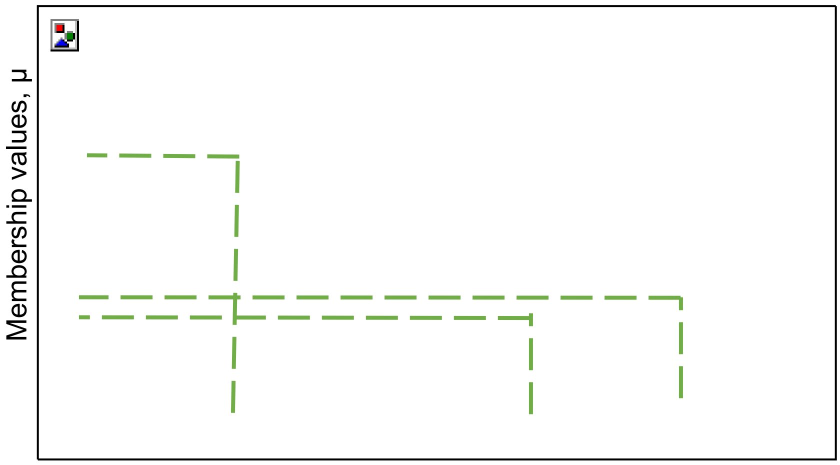
Thr	D	PLR	UR	Decision	Final Membership Values
M(0.45)	L(0.7)	L(0.3)	M(0.6)	PY	0.3
M(0.45)	L(0.7)	L(0.3)	H(0.4)	PY	0.3
M(0.45)	L(0.7)	M(0.7)	M(0.6)	Y	0.45
M(0.45)	L(0.7)	M(0.7)	H(0.4)	Y	0.4
M(0.45)	M(0.4)	L(0.3)	M(0.6)	PY	0.3
M(0.45)	M(0.4)	L(0.3)	H(0.4)	PY	0.3
M(0.45)	M(0.4)	M(0.7)	M(0.6)	PN	0.4
M(0.45)	M(0.4)	M(0.7)	H(0.4)	PN	0.4
H(0.6)	L(0.7)	L(0.3)	M(0.6)	Y	0.3
H(0.6)	L(0.7)	L(0.3)	H(0.4)	Y	0.3
H(0.6)	L(0.7)	M(0.7)	M(0.6)	Y	0.6
H(0.6)	L(0.7)	M(0.7)	H(0.4)	PY	0.4
H(0.6)	M(0.4)	L(0.3)	M(0.6)	PY	0.3
H(0.6)	M(0.4)	L(0.3)	H(0.4)	PY	0.3
H(0.6)	M(0.4)	M(0.7)	M(0.6)	Y	0.4
H(0.6)	M(0.4)	M(0.7)	H(0.4)	PN	0.4

Table 13. Fuzzy Rule Base for Fuzzy Controller.

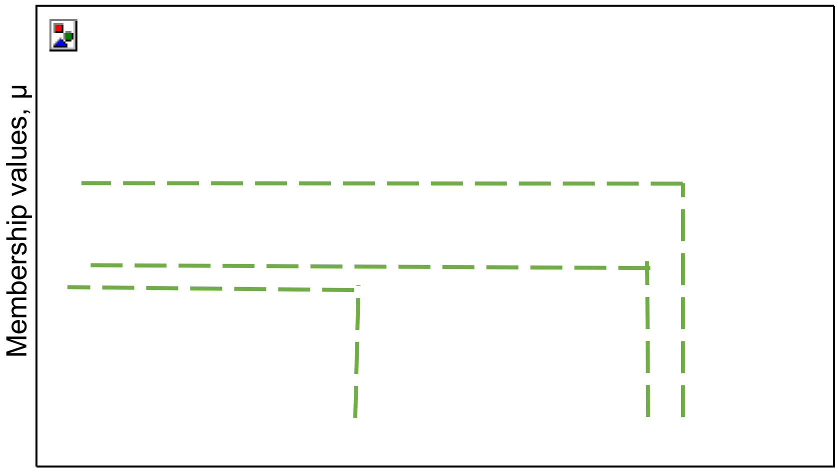
Rule	*Thr*	*D*	*PLR*	*UR*	*Decision*
1	Low	Low	Low	Low	PY
2	Low	Low	Low	Medium	PY
3	Low	Low	Low	High	Y
4	Low	Low	Medium	Low	PY
5	Low	Low	Medium	Medium	PY
6	Low	Low	Medium	High	PY
7	Low	Low	High	Low	PN
8	Low	Low	High	Medium	PN
9	Low	Low	High	High	PY
10	Low	Medium	Low	Low	PY
11	Low	Medium	Low	Medium	PN
12	Low	Medium	Low	High	PN
13	Low	Medium	Medium	Low	PN
14	Low	Medium	Medium	Medium	PY
15	Low	Medium	Medium	High	PN
16	Low	Medium	High	Low	PN
17	Low	Medium	High	Medium	PN
18	Low	Medium	High	High	PN
19	Low	High	Low	Low	PN
20	Low	High	Low	Medium	PN
21	Low	High	Low	High	PN
22	Low	High	Medium	Low	PN
23	Low	High	Medium	Medium	PN
24	Low	High	Medium	High	PN
25	Low	High	High	Low	PN
26	Low	High	High	Medium	PN
27	Low	High	High	High	PN
28	Medium	Low	Low	Low	PY
29	Medium	Low	Low	Medium	PY
30	Medium	Low	Low	High	PY

31	Medium	Low	Medium	Low	PY
32	Medium	Low	Medium	Medium	PY
33	Medium	Low	Medium	High	PY
34	Medium	Low	High	Low	PN
35	Medium	Low	High	Medium	PY
36	Medium	Low	High	High	PY
37	Medium	Medium	Low	Low	PY
38	Medium	Medium	Low	Medium	PY
39	Medium	Medium	Low	High	PY
40	Medium	Medium	Medium	Low	PN
41	Medium	Medium	Medium	Medium	PN
42	Medium	Medium	Medium	High	PN
43	Medium	Medium	High	Low	PN
44	Medium	Medium	High	Medium	PN
45	Medium	Medium	High	High	PN
46	Medium	High	Low	Low	PN
47	Medium	High	Low	Medium	PN
48	Medium	High	Low	High	PN
49	Medium	High	Medium	Low	PN
50	Medium	High	Medium	Medium	PN
51	Medium	High	Medium	High	PN
52	Medium	High	High	Low	PN
53	Medium	High	High	Medium	PN
54	Medium	High	High	High	PN
55	High	Low	Low	Low	Y
56	High	Low	Low	Medium	Y
57	High	Low	Low	High	Y
58	High	Low	Medium	Low	PY
59	High	Low	Medium	Medium	Y
60	High	Low	Medium	High	Y
61	High	Low	High	Low	PY
62	High	Low	High	Medium	PY

63	High	Low	High	High	PY
64	High	Medium	Low	Low	PY
65	High	Medium	Low	Medium	Y
66	High	Medium	Low	High	Y
67	High	Medium	Medium	Low	PY
68	High	Medium	Medium	Medium	Y
69	High	Medium	Medium	High	PY
70	High	Medium	High	Low	PN
71	High	Medium	High	Medium	PY
72	High	Medium	High	High	PY
73	High	High	Low	Low	PY
74	High	High	Low	Medium	PY
75	High	High	Low	High	PY
76	High	High	Medium	Low	PN
77	High	High	Medium	Medium	PY
78	High	High	Medium	High	PN
79	High	High	High	Low	PN
80	High	High	High	Medium	PN
81	High	High	High	High	PN



(a) Normalised Weight



(b) Normalised Weight

Figure 5-19. Membership functions for Service, $Serv$

5.5.2 Neuro-Fuzzy Packet Scheduling

The Artificial Neural Network (ANN) can be demonstrated to have the powerful capability of expressing the relationship between input-output variables. It takes 'n' input values x_1, x_2, x_3, …, x_n and process the input vectors x_N to produce output vector z_N. The output vector 'z' represents the pattern or identification group. A trained artificial neural network represents the system that maps a set of input vectors x_N: N=1,2,3, …, n to a set of target output vectors z_N: N=1,2,3, …, n. This mapping enables the neural network to make interpolations and extrapolations to correspond any input x to the output z, which best matches the input pattern [123, 124]. After training the artificial neural network, it acts as a mathematical machine that implements the algorithm specified by the input/output nodes, nodes in the hidden layers, connecting lines, transforming nodes functions and the weight associated with the connecting lines of the artificial neural network.

Both neural networks and fuzzy systems are dynamic, parallel processing systems that estimate input-output functions. They estimate a function without any mathematical model and learn from experience with sample data[125]. For an application, an ANN must be trained to acquire the suitable weights on the connecting lines so that ANN can produce the close approximation of target result. The ANN is a better option as compared to the other artificial intelligence techniques as it does not use the pre-programmed knowledge base, have no restrictive assumptions, can handle noisy/imprecise data, robust and flexible. On the other hand, it has some drawbacks too, such as it requires high-quality data for training/learning, variables must be very carefully selected a priori, risk of overfitting, requires a definition of architecture, the long processing time for training and possibility for illegal network behaviour. The ability to embed the empirical data into the fuzzy control system can be achieved by utilizing training techniques of neural networks. This can significantly widen the application of the fuzzy system as the ability to make use of both practical and expert information. The dimensionality in the fuzzy systems is its limitation or in other words a severe drawback. The term “dimensionality curse” [126] is used for fuzzy system as for a fuzzy system the cost for implementation of rule base and deriving the output increases exponentially as the input space dimension increases.

Neuro-fuzzy systems are fuzzy systems, which use Neural Networks (NNs) to determine their properties (fuzzy sets and fuzzy rules) by processing data samples [125, 127]. Neuro-fuzzy integrates to synthesize the merits of both NN and fuzzy systems in a complementary way to overcome their disadvantages. The fusion of an NN and fuzzy logic in neuro-fuzzy models possess both low-level learning and computational power of NNs and advantages of high-level human-like thinking of fuzzy systems. For identification, a hybrid neuro-fuzzy system called ANFIS combines a NN and a fuzzy system. Adaptive neuro-fuzzy inference system (ANFIS) has been proved to have significant results in modeling nonlinear functions. In ANFIS, the membership functions (MF) are extracted from a data set that describes the system behaviour. The ANFIS learns features in the data set and adjusts the system parameters according to the given error criterion. In a fused architecture, NN learning algorithms are used to determine the parameters of the fuzzy inference system.

The expert information to model with the input space relationship could be utilized to reduce the set of rules, as the expert knowledge may make the problem tractable. Relying only on expert knowledge is not enough to tune a fuzzy system for efficient and precise output. The use of training techniques based on error allows a fuzzy system to acquire the complexities hidden in the input data. A neural-fuzzy technique can be used for building a fuzzy system in multiple ways apart from training method. For example, it can be used for fuzzy membership function determination, in fuzzy rule selection and also in case of hybrid systems. In order to compensate for the disadvantages of one approach with the advantages of another approach, a hybrid method called ANFIS is proposed

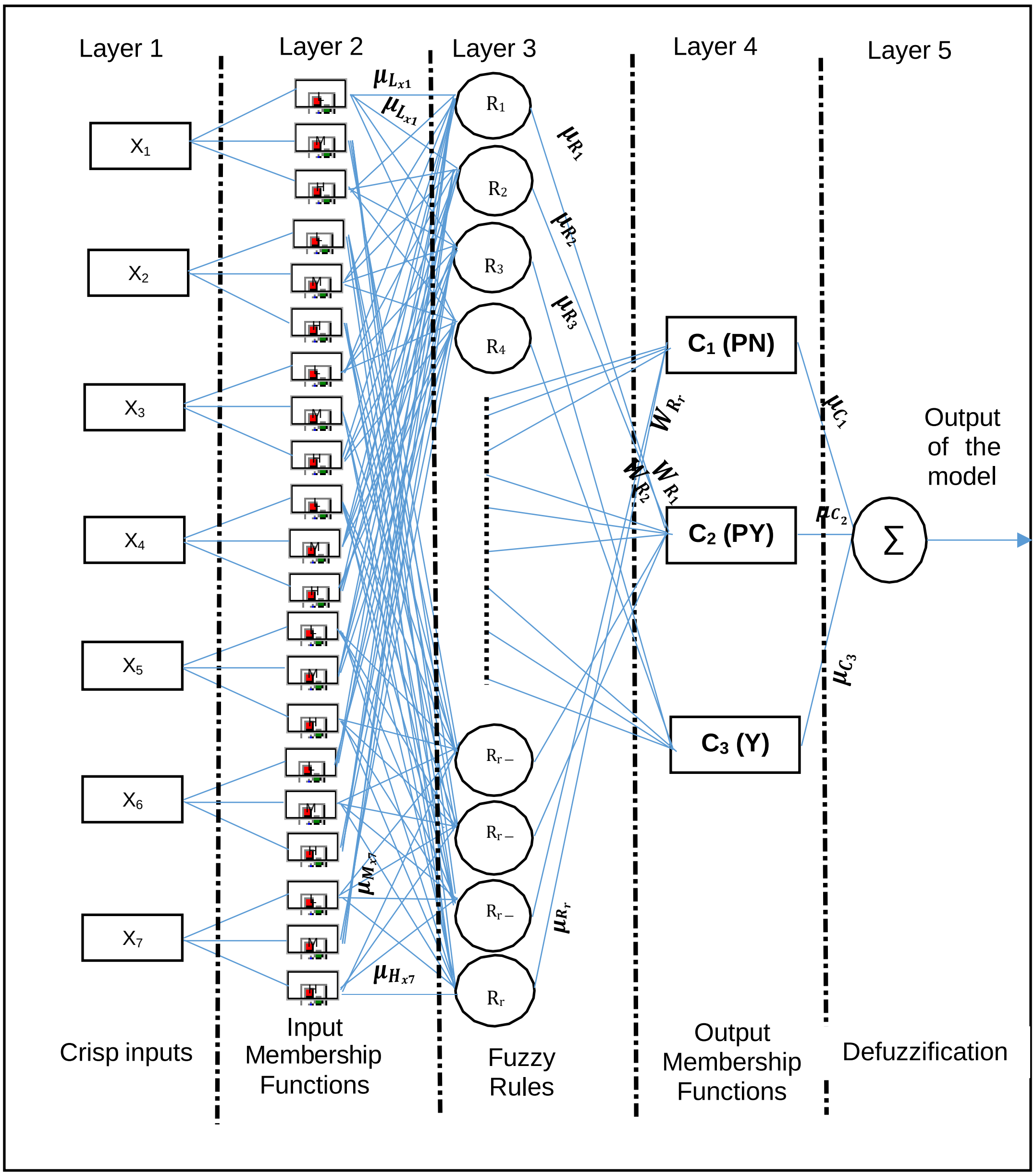


Figure 5-20. Neuro-Fuzzy System Architecture

Figure 5-20 depicts the neural network representation of the fuzzy control system. The components of the input vector consist of membership values to the overlapping partitions of linguistic properties low, medium, and high corresponding to each input feature. This provides scope for incorporating linguistic information in both the training and testing phases of the said models and increases their robustness in tackling imprecise or uncertain input specifications. An n-dimensional feature space is decomposed into 3^n overlapping sub-regions corresponding to the three primary properties low, medium, and high. Although there is an associated increase in dimension and cost, one has to offset this with the specific gains achieved.

The inputs x_1, x_2, x_3, x_4, x_5, x_6, and x_7 represent the input parameters of the neuro-fuzzy control system such as Bandwidth, Delay, Reliability, Throughput, Packet Loss Rate, Utilisation Rate, and Cost respectively. The purpose of utilizing the neural network here is to take advantage of the neural network intelligent techniques for deriving the reduced number of rules and get rid of the fuzzy logic dimensionality curse which affects the efficiency of the fuzzy algorithm due to many input parameters in the target packet scheduling system.

As can be seen in Figure 5-20, layer 1 shows the seven inputs to the system and the membership of input parameters or fuzzification is performed at layer 2. The fuzzy sets are used in the antecedents of fuzzy rules are represented by the neurons in layer 2. The fuzzification of the neuron will be followed, after receiving the input, determines the grade to which the input belongs to neuron's fuzzy sets. Layer 3 represents the fuzzy rules or can also be called the fuzzy rules layer. Each fuzzy rule is represented by a different neuron in this layer. The neuron (fuzzy rule neuron) in this layer receives input from fuzzification neurons from layer 2, which denotes fuzzy sets. The intersection operation can be implemented in neuro-fuzzy systems, using the product operator. Therefore, the output of the i^{th} neuron in layer 3 can be determined as in equation (29):

$$\mu R_i = X_{1i} \times X_{2i} \times X_{3i} \times X_{4i} \times X_{5i} \times X_{6i} \times X_{7i} \tag{29}$$

This will yield equation (30) by substituting for rule 1 (R_1):

$$\mu R_1 = \mu(L)_{X_1} \times \mu(L)_{X_2} \times \mu(L)_{X_3} \times \mu(L)_{X_4} \times \mu(L)_{X_5} \times \mu(L)_{X_6} \times \mu(L)_{X_7} \tag{30}$$

This is followed by the firing strength of the neuron, which denotes rule R_1 can be represented by the μR_1. Then, layer 4 can be referred to as output membership layer, which denotes the fuzzy sets in the output of the fuzzy rules. The neurons in this layer receive inputs from the neurons in the fuzzy rules layer. Once the inputs are received from layer three, the neurons in this layer combine all the received inputs using fuzzy union operation. The fuzzy union operation can be implemented by an exclusive OR operator which is also known as the probabilistic OR operator. The output of the neurons at this layer can be denoted by μC_i, where 'i' is the number of neurons in this layer. Equation (31) represents the general format for the firing strength of neurons in layer four.

$$\mu C_i = \mu R_1 \oplus \mu R_2 \oplus \mu R_3 \oplus \mu R_4 \oplus \ldots\ldots\ldots \mu R_n \tag{31}$$

In this case, 'n' represents the total number of rules which satisfy the output condition 'i.' Layer 5 which is also called the defuzzification layer is composed of a single neuron. The neuron in this layer represents the single output of the neuro-fuzzy system. The neuron in layer 5 receives the input from the output of the neurons in layer 4 and combines them into a single fuzzy set. A standard defuzzification method centroid, average weighted or any other can be applied in a neuro-fuzzy system. In this thesis, the average weighted expression is as shown. Equation [32] represents the output decision of the fuzzy neural system using the weighted average method.

$$Decision = \frac{\mu C_1 \times aC_1 \times bC_1 + \mu C_2 \times aC_2 \times bC_2 + \mu C_3 \times aC_3 \times bC_3}{\mu C_1 + \mu C_2 + \mu C_3} \tag{32}$$

where ‘a’ is the centre, and ‘b’ is the width of triangular activation/membership function. Many fuzzy rules are represented by neurons, which will be in high computational power on layer 2. It can be reduced to a certain extent by converging fuzzy inputs into a set of general rules, which will depend on the output of the fuzzy rules set. The operation of generalized rules can be used to reduce the total number of rules to a preferable number depending on how large the rules being formed. The implementation is shown in Figure 5-21.

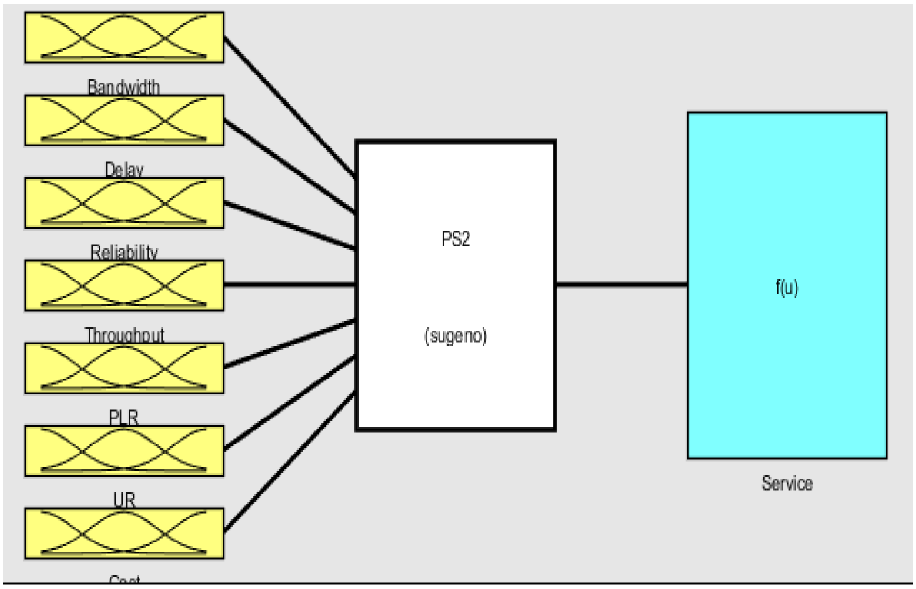


Figure 5-21. Implementation of Neuro-Fuzzy based Packet Scheduling in Matlab.

The neuro-fuzzy system, with a reduced number of rules, is trained to perform efficiently with the set of rules data in hand. This training process of the neuro-fuzzy system adjusts the weights to produce the required output for any input pattern. In the fuzzy system, the same process is carried out by tuning the membership functions, which was nearly impossible for the target packet scheduling system with such broad input parameters. Once the neuro-fuzzy system is trained, a comparison is made between neuro-fuzzy and the fuzzy inference system with the help of 400 input samples for each of the seven input parameters. Each of these inputs is randomly generated and provided to the neuro-fuzzy and the fuzzy inference system. Finally, the obtained decision factors from both neuro-fuzzy and fuzzy inference systems are plotted to compare the output of each system as shown in both Figure 5-22 and Figure 5-23. Subsequently, to evaluate further the probability distribution of both decisions are plotted with the highest probability values of 0.995 and 0.999 as shown in Figure 5-24 and Figure 5-25 respectively.

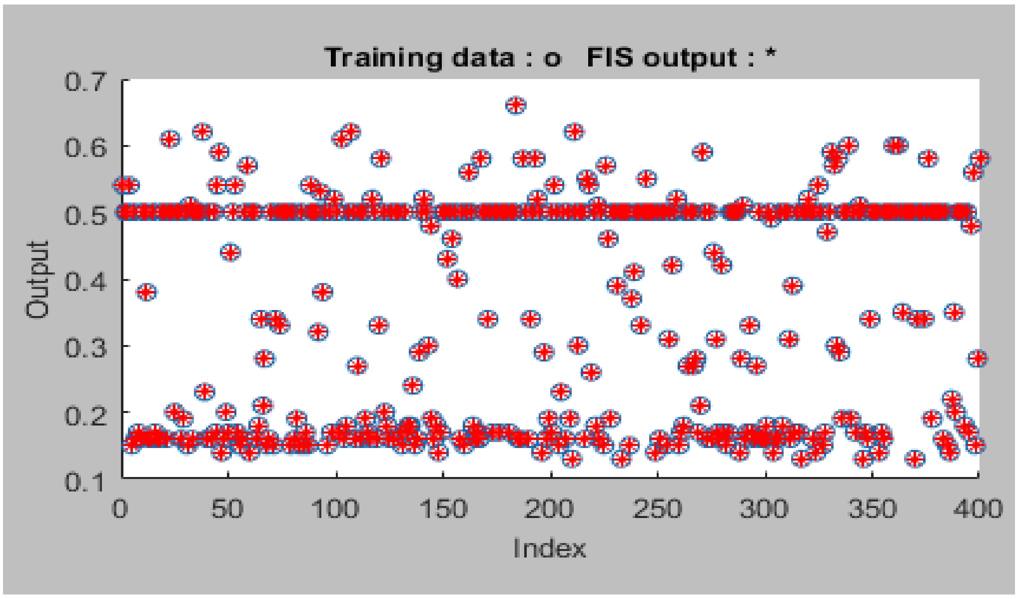


Figure 5-22. Comparison of Training Random variables and FIS output (Neuro-Fuzzy).

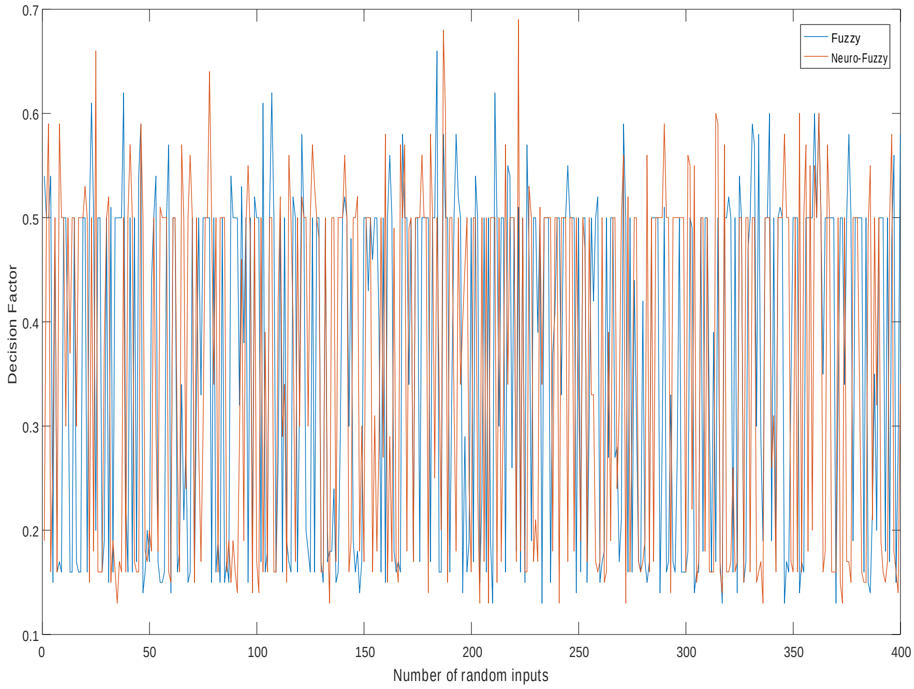


Figure 5-23. Comparison of Neuro-Fuzzy and Fuzzy Inference system with 400 random input samples

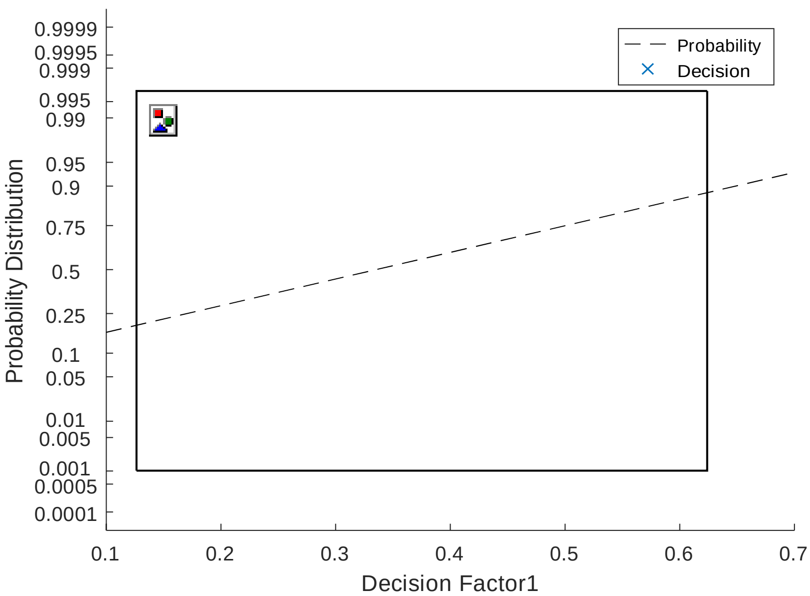


Figure 5-24. Probability Distribution of Fuzzy Inference system

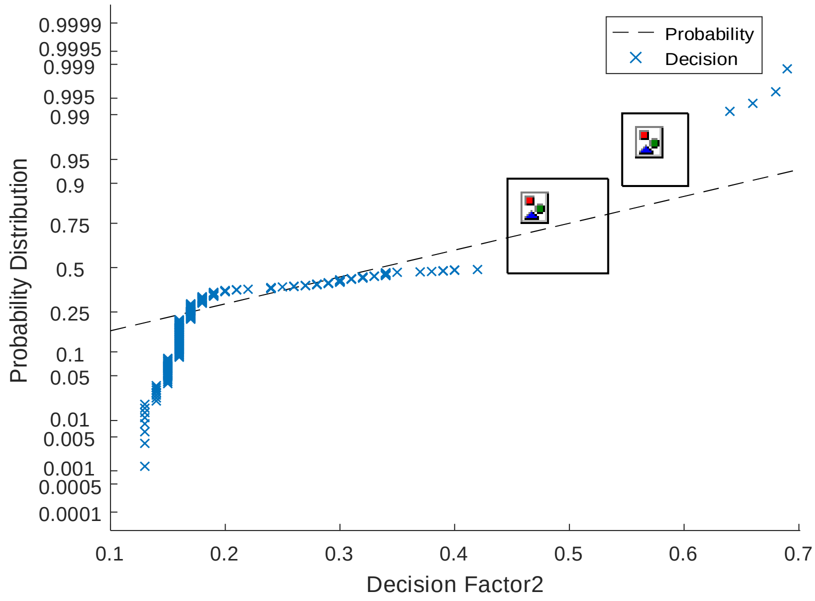


Figure 5-25. Probability Distribution of Neuro-Fuzzy Inference system

Figure 5-26 shows the training error of the neuro-fuzzy inference system. It is observed that the training error of the neuro-fuzzy inference system is decreasing as the number of iteration increases, which indicates that the module is working well in terms of performance.

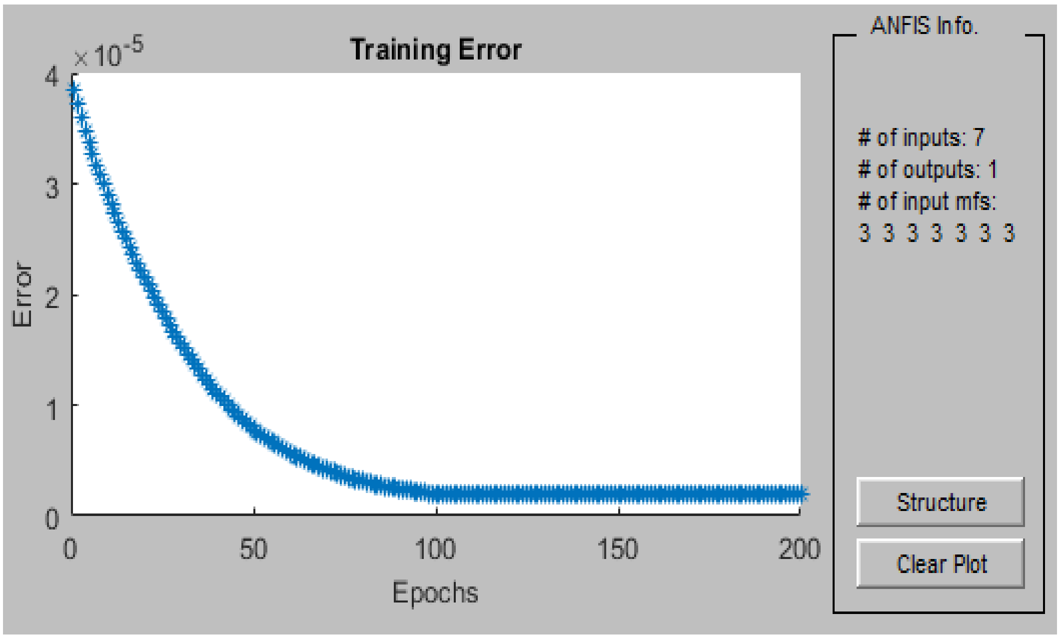


Figure 5-26. Training Error of Neuro-Fuzzy Inference system.

5.6 **Concluding Remarks**

In this chapter, we adopt fuzzy logic control due to the significant property of attaining an intuitive understanding of the way to control the process, through incorporating human reasoning in the control algorithm. It is independent of mathematical models of the system to be controlled, thus achieving inherent robustness and reducing design complexity. This is in contrast with conventional control approaches that concentrate on constructing a controller with the aid of an analytical system model that, in many cases is uncertain, nonlinear, and subject to noises. Thus, if the fuzzy logic control is designed with a good (intuitive) understanding of the system to be controlled, the limitations due to the complexity the system's parameters introduce on a mathematical model can be avoided. A novel approach in this thesis is to use both fuzzy and neuro-fuzzy systems to either ignore such complex parameters in the mathematical model or to simplify the model to such an extent to obtain some improved results on the conventional system.

6 Simulation Model and Results

6.1 Introduction

In this chapter, the simulation models that were performed using analytical tools such as OPNET, NS2, and MATLAB are presented. The objectives from the OPNET, NS2, and MATLAB simulation are specified in section 6.3.1, 6.3.2, and 6.3.3. The OPNET model focuses on the bandwidth management technique (elaborated in Section 6.2). The signaling protocols that were specified in chapter 4 will also be modeled in the OPNET simulation and the end-to-end delay performance determined. The simulation approach considers certain aspects of bandwidth management because certain functions in MPLS rely on the functional procedures for traffic engineering. However, the main aspects related to bandwidth management in the OPNET simulation are addressed in this research study.

Based on the findings from section 5, the two conventional packet scheduling algorithms identified will be simulated in NS2 and presented in Section 6.3. The FIFO and WFQ algorithms will be applied to the IP/MPLS network. Furthermore, C program is written to link the FIS module formed in MATLAB with NS2 to determine the performance metrics for comparison. This is followed by the compilation of the FIS module and then execution of compiled FIS with the input parameters in the NS2 platform. An output is generated, which served as the service utilization for the system.

The methods of simulation are essential in the research and standardization phases whenever a new technology is under development. Consequently, the most convenient way to evaluate the performance of the system is by computer simulations, which have become a widely adopted methodology [128, 129]. This simulation-based approach not only reduces the high cost of implementing a real system but also saves significant development time.

Over the last two decades, the development of powerful and inexpensive computers, together with robust software packages, has proliferated. Consequently, computer-aided design and analysis techniques are readily available and are usually referred to as simulations. An important motivation for adopting a simulation approach in this thesis is that it provides a valuable insight into system behaviour and performance before considering real expensive implementation. For this reason, the aim of this chapter is to evaluate guidelines specifying both the methodology and tools used in assessing the overall performance of the proposed bandwidth management framework and discuss simulation results.

6.2 Performance Metrics

The following performance metrics will be used for evaluating the performance of the proposed IP/MPLS framework.

1. Throughput: The data rate or throughput is defined as the number of information bits per second successfully delivered or received and is an important performance metric in terms of QoS. Throughput depends on the amount of accompanying redundant information carried along with the transmitted data, traffic queuing/aggregation mechanisms, congestion conditions, and priority handling policies applied to data

flows. It is related to data rate, and it can be a fraction of the maximum available bandwidth.

2. Delay: Delay is the average transit time of packets from the ingress to egress points of the network. There are the end-to-end delays and the individual delays along portions of the network. The end-to-end delay depends on the propagation rate of data in a communication medium (satellite or terrestrial) [94]. It can be the number and type of network elements (design, processing, switching, and buffering capabilities).
3. Packet Delay Variation or Jitter: Jitter is one form of delay variation caused by the difference in delay exhibited by different packets that are part of the same data flow. Jitter is caused primarily by differences in queuing delays for consecutive packets, and by alternate paths taken by packets because of routing decisions. For instance, Packet delay variation (PDV_k) for a packet k between source and destination is the difference between the absolute packet transfer delay (Absolute PTD_k) of the packet and a defined reference packet transfer delay, $PTD_{1,2}$ between those same measurement points:
4. Packet Delivery Ratio: Packet delivery fraction is the ratio of a total number of a data packet received to the total number of data packets sent. This is an estimation of how routing protocol is efficient and effective.

$$PDR = \frac{Number\ of\ packets\ received}{Number\ of\ packets\ sent} \tag{33}$$

5. Packet Loss: Packet loss is typically measured as a percentage of the ingress and egress traffic. Packets can be lost because they are dropped at congestion points, traffic violations (synchronization,

signaling, unrecoverable errors), excessive load, or natural loss included in compression/decompression mechanisms. Voice and video communications are more tolerant to the loss of packets than in the usual data communications.

6.3 Simulation Platforms

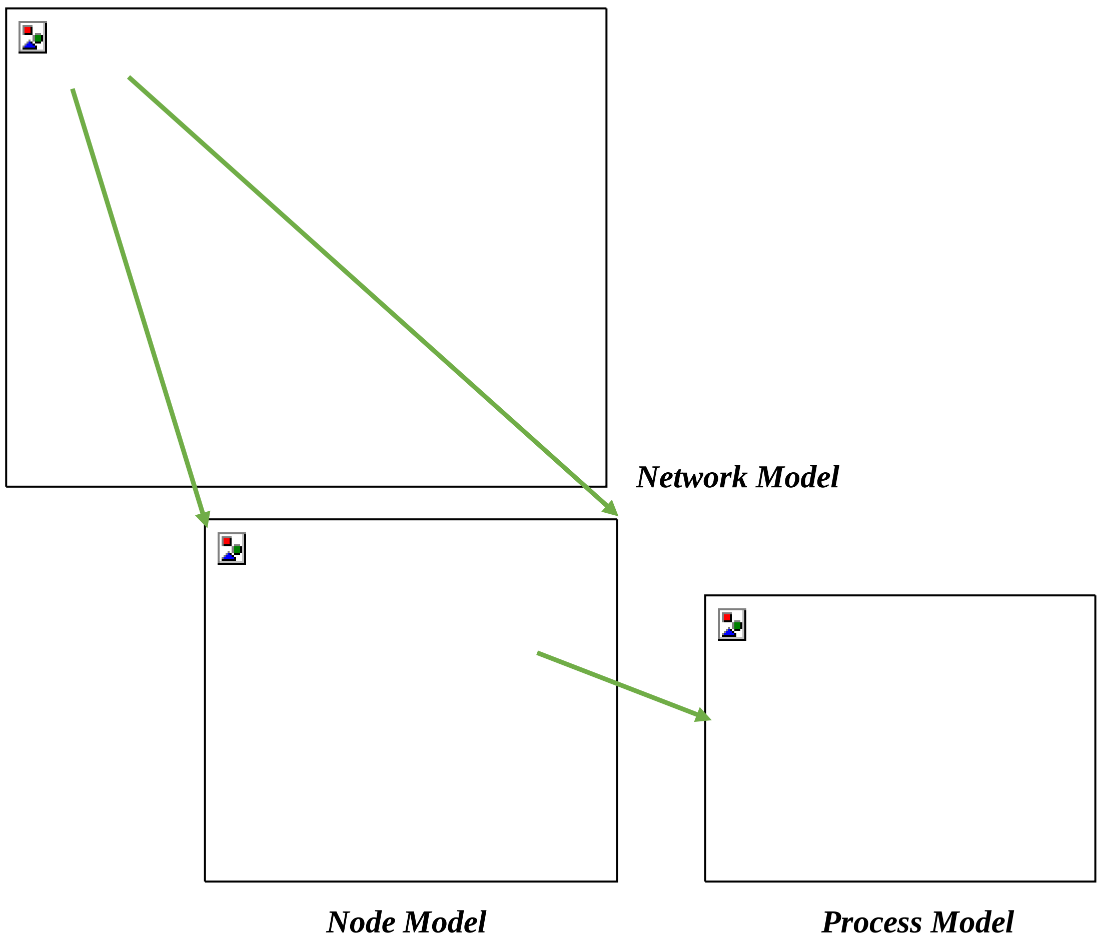
6.3.1 OPNET Simulation Platform

The OPNET simulation scenarios and parameters are defined to analyze the MPLS system performance under specific metrics such as throughput, end-to-end packet delay, and packet delay variation. OPNET is a discrete event simulator which uses a graphical interface where the network components of a real World are identified. It provides a virtual network environment for the entire network models, which include its routers, switches, protocols servers, and individual applications. The main difference with other simulators lies in its power and versatility. It performs better with the OSI model. The procedure of working with OPNET consists of three significant aspects namely: Network Model, Node Model, and Process Model as shown in Figure 6-1. The goal of the simulation is to obtain results and to get an insight into other model systems by evaluating the results [130].

Network Model: This consists of nodes, links, and subnets, which form the basis for network devices and a group of devices. The devices are linked together using either point-to-point or bus links. Here, there is the interconnection of nodes such as LERs, LSRs, and sources

Node Model: This represents the basic building block (modules) which include processors, queues, and transceivers. It is also made up of interfaces such as packet stream and statistical wires.

Process Model: In this case, we have the State transition diagrams, Block of C code, OPNET Kernel Procedure (KPs), State variables, and Temporary variables. Processes usually responded to interrupts.



```c
/*****************************************************************************/
/** This state is entered after a remote interrupt from IP has been        **/
/** received. This is a signal that IP has finished its interface table    **/
/** initialization process, and that RSVP is enabled at least on one IP    **/
/** interface.                                                             **/
/**                                                                        **/
/** The state may also be entered after a remote interrupt from MPLS Mgr   **/
/** This will occur only if RSVP is being used to setup MPLS LSPs in the   **/
/** network. If we receive an interrupt here from MPLS then this means     **/
/** that RSVP is configured as signaling protocol for Dynamic LSPs         **/
/**                                                                        **/
/** The state variables that were not                                      **/
/** initialized after BEGSIM interrupt was received, are initialized.      **/
/** RSVP then publishes in the process registry its new status.            **/
/** The process also discovers the local IP and application processes.    **/
/** Finally, a pointer to the IP routing table is obtained.                **/
/*****************************************************************************/

/* Get the interrupt details                                                */
intrpt_type = op_intrpt_type ();
intrpt_code = (RsvpT_Intrpt_Type)op_intrpt_code ();

if (IP_OR_MPLS_NOTIF)
    {
    /* Register this higher layer protocol with IP                          */
    higher_layer_protocol_id = IpC_Protocol_Rsvp;
    Ip_Higher_Layer_Protocol_Register ("rsvp", &higher_layer_protocol_id);

    /* Obtain pointers to the ip common routing table and ip iface table    */
    rsvp_ip_ucast_tables_ptr_get ();

    /* Get a pointer to multicast routing table.                            */
    rsvp_ip_mcast_table_ptr_get ();

    /* Finish the initialization procedure.                                 */
    rsvp_sv_init (intrpt_type, intrpt_code);
```

Blocks of C code

Figure 6-1. Network model, Node model and Process model of Mobile device

The packet is the information-carrying entities that circulate among the system component. This forms general data structures, organized into fields of user-defined information. Packets are dynamically created and destroyed as the simulation progresses. The procedure of moving packets from processor 1 to processor 2 that follows the pattern of the queue is to avoid congestion in the network. It can be used to transfer data between objects in the node and Network domains. Processors, queues, and modules forwarded packets within nodes via packet streams as shown in Figure 6-2. This may introduce delay and queue packets until destination module is ready to accept them.

Packets consist of three main storage areas. The first area be a list of user-defined values called packet fields; the second area is made up of a set of pre-defined values that are used by the simulation system for accounting purposes, and the third area contains transmission data attributes that are used to support models for communication links.

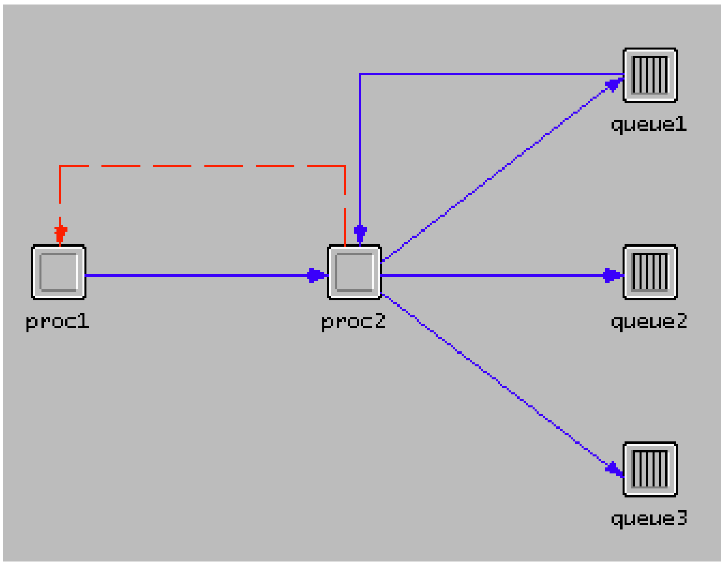


Figure 6-2. Processors and Queues modules.

Table 14 depicts the configuration parameters used in the simulation of MPLS model using multimedia services such as interactive voice and interactive video while the relationship between theoretical and simulation results in terms of R-values and Mean Opinion Score is shown in Table 15.

Table 14. Simulation Parameters for Multimedia Services

Voice		**Video**	
Attributes	**Values**	**Attributes**	**Values**
Encoder scheme	G.711	Frame per second	30
Voice Frame per packet	1	Frame size (B)	352x240 pixels
Type of Service	Interactive voice	Type of Services	Interactive video
Data rate (kbps)	120	Data rate (Mbps)	30

Table 15. Relationship between R-values and MOS

R-value (lower limit)	***Theoretical MOS***	***Simulation MOS***	***User Satisfaction***
90 < R < 100	4.34	3.86	Very satisfied
80 < R < 90	4.03	3.66	Satisfied
70 < R < 80	3.60	2.94	Some users dissatisfied
60 < R < 70	3.10	2.10	Many users dissatisfied
50 < R < 60	2.58	1.39	Nearly all users dissatisfied

6.3.2 Network Simulator NS2 Platform

NS is an event-driven network simulator that simulates a variety of IP networks. It implements network protocols such as TCP and UDP, traffic source behaviour such as FTP, Telnet, and Web. Currently, NS (version 2) written in C++ and OTcl (Tcl script language with Object-oriented extensions). It is an OTcl interpreter with network simulation object libraries. The traditional methods for examining the performance of TCP have been a simulation, implementations, and measurements. However, the efforts have been made to characterize the throughput of TCP analytically as a function of parameters (like packet drop rate and round trip time) [131, 132].

In this thesis, NS 2 is used to design and evaluate the performance of Packet scheduling algorithms such as FIFO (baseline) and WFQ using real-time and non-real time traffics in IP/MPLS networks. While tool command language (TCL) scripting is used in simulation scenario definition. Each link is configured with the parameters such as bandwidth, propagation delay, and queue type. Data communication between nodes is configured with transport and application layer agents that are required to be attached to both sender and receiver nodes. The procedure of NS2 to design packet scheduling in IP/MPLS is as shown in Figure 6-3.

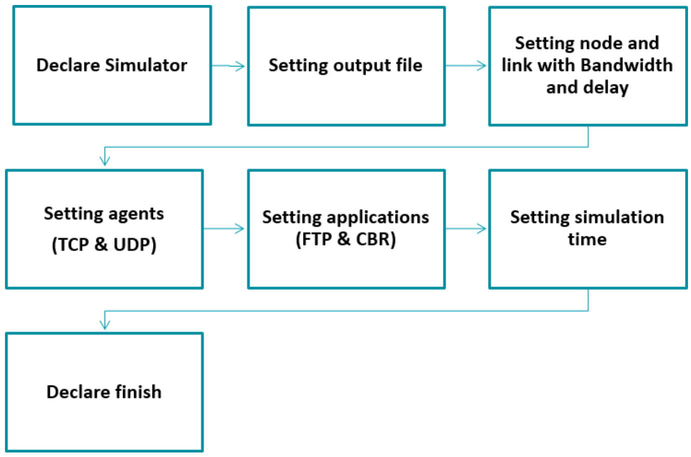


Figure 6-3. Procedure of NS2 Implementation

It is imperative to create a separate file both header and source, which has the fuzzy rules required to integrate into TCP. Then, call those rules at appropriate points after implementing the TCP code in NS2. Afterward, recompile the NS2 source code and evaluate the results with existing protocols. There is a need to be very clear on the changes observed when source code is call in NS2. The fuzzy and neuro-fuzzy algorithms have been incorporated in NS2 using new custom-made “ANSI C” library developed for packet scheduling purposes.

6.3.3 MATLAB Platform

Over the past years, a variety of software packages have been developed, which have been widely used to simulate communication systems. The graphical model builders are relatively simple to use by clicking and dropping functional blocks on the computer screen and linking them together to create a simulation model in a hierarchical block diagram form. As an alternative to using a graphical block diagram editor for model building, one could use an intermediate (pseudo) language such as the MATLAB command language, which is one of the accessible numerical computing environments and programming languages. MATLAB was chosen as the primary system modeling tool for this thesis to demonstrate original concepts, for problem-solving, and rigorous comparison with the baseline model. There are many persuasive reasons for adopting MATLAB. Firstly, it combines excellent computational capabilities with easy-to-use graphical capabilities. Additionally, MATLAB code is very concise, making it possible to express complex digital signal-processing (simulation) algorithms using relatively few lines of code and although MATLAB is relatively slow compared to the underlying C/C++ or Java, running the MATLAB codes in powerful computers can overcome these slow execution speeds. For this thesis, the computer specifications and MATLAB version used are summarised below.

Table 16. Simulation platform specifications

<table>
<tr><th>MATLAB version</th><th colspan="2">PC specification</th></tr>
<tr><td rowspan="4">MATLAB R2014a</td><td>OS</td><td>Windows 8 (64-bit)</td></tr>
<tr><td>Processor</td><td>Intel Core i5-3470 CPU @3.20GHz</td></tr>
<tr><td>RAM</td><td>8GB</td></tr>
<tr><td>Hard Disk</td><td>500GB</td></tr>
</table>

6.4 Simulation Parameters

To analyze the performance of the baseline FIFO and WFQ based packet scheduling algorithms, NS2 simulation of MPLS consists of 15 nodes layout with four sources, four destinations, and seven core LSRs is considered. The bandwidth of 2MB with a link propagation delay of 10 ms is considered, and there are up to packet size of 1000 bytes with a sending rate of about 200 kbps from the source. The number of bits transmitted per second within the MPLS network is 200,000 bits per second. The architecture of MPLS model is shown in Figure 6-4 while the establishment of label distribution, as well as explicit routing at the MPLS core, can be seen in Figure 6-5.

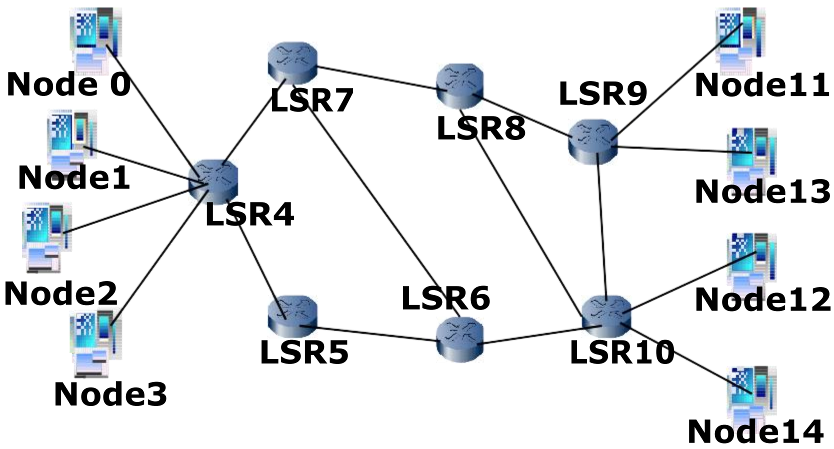


Figure 6-4. Architecture of the MPLS network

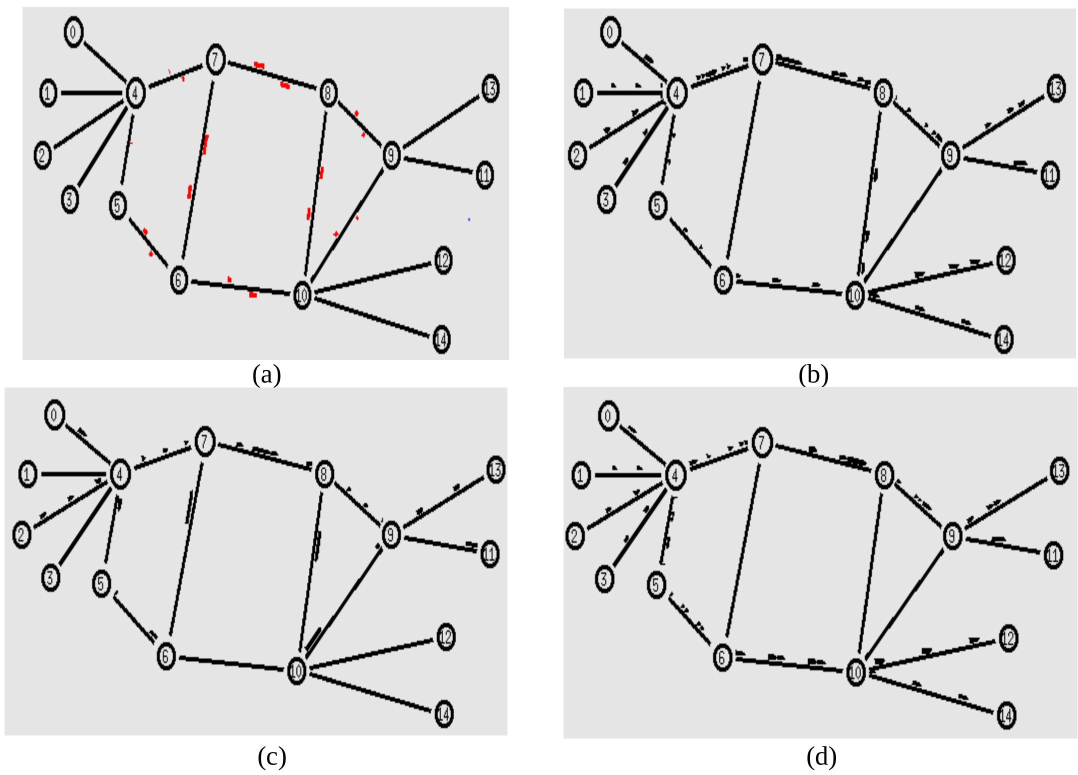


Figure 6-5. Label Mapping, Label Distribution, and Packet Forwarding.

Under the packet forwarding scheme based on the low-cost path, packets from node 0 (source) are delivered along LSR 4-7-8-9, and packets from node 1 (source) are delivered along LSR 4-5-6-10. Figure 6-5 (a) shows the initial simulated network. At 0.5 s that LDP Mapping Message is used to distribute labels based on control is driven trigger. As a result, every possible LSP in the MPLS network is established.

The event at about 34 seconds when flows of FEC 11, FEC 13 and FEC 12, FEC 14 are aggregated into a flow of FEC 10 is illustrated in Figure 6-5 (b). Subsequently, Figure 6-5 (c) shows an event at 70 seconds that Constraint-Routing LDP (CR-LDP) request message initiated by LSR4 is delivered along LSR 6-7-8-10 to create an Explicit Routing LSP(ER-LSP) between LSR4 and LSR10. Figure 6-5 (d) shows occurrence at 76 seconds that CR-LDP Mapping Message is sent by LSR 10 as the response for the LDP Request Message initiated by LSR4.

Then, an LSP for FEC 10 is terminated at 50 seconds, and node 1 and node 3 agents stop generating packet at 60 seconds. An ER-LSP of which LSPID is 3000 is established between LSR4 and LSR12 through LSR 6-7-8-10 is followed at 70 seconds. At 90 seconds, a flow of FEC 11 is bound to the established ER-LSP. At 110 seconds, the ERLSP is terminated with LDP release message. Also, then, at 120 seconds, an ER-LSP Tunnel of LSPID 3500 is established between LSR7 and LSR8 through LSR6 and LSR10. An ER-LSP of LSPID 3600 is also established between LSR4 and LSR9 through LSR 4-7-3500 at 130 seconds. The value of 3500 in the specified ER means LSPID, which is used to identify the tunnel ingress point as a next-hop. This allows for stacking new ER-LSP (that is, LSPID 3600) within an already established LSP Tunnel (that is, LSPID 3500). At 140 seconds, the flow of

FEC 9 is bound to the established ER-LSP. MPLS tables can be seen in Appendix C for more details information.

Various parameter settings are shown in the tables below in Table 17, Table 18, and Table 19, respectively.

Table 17. Simulation platform specifications for Voice

Parameters	**Values**
Simulation Length (s)	180
Number of Nodes	15
Packet size (Bytes)	1000
Sending rate (kbps)	200
Interval (s)	0.005
Link bandwidth (MB)	2
Link Propagation delay (ms)	10
Transmission delay (ms)	4
Processing delay (μs)	10-20
Queuing delay (ms)	variable
Type of source	Exponential traffic (voice)

Table 18. Simulation platform specifications for File Transfer Protocol (FTP)

Parameters	Values
Simulation Length (s)	180
Number of Nodes	15
Packet size (Bytes)	1000
Sending rate (kbps)	500
Interval (s)	0.005
Link bandwidth (MB)	2,5
Link Propagation delay (ms)	10
Transmission delay (ms)	4
Processing delay (μs)	10-20
Queuing delay (ms)	variable
Type of source	FTP traffic

Abbreviations used throughout the simulation scenarios are MPLSftp with FIFO, MPLSftp with WFQ, IPftp with FIFO, and IPftp with WFQ. These imply the application of the conventional packet scheduling algorithms using non-real-time traffic (FTP). It can be seen in Table 19 that neuro-fuzzy inference system information such as number of nodes, number of fuzzy rules, and number of training data pairs is provided.

Table 19. Neuro-Fuzzy Inference System Information

Parameters	Values
Number of Nodes	4426
Number of Fuzzy rules	2187
Number of training data pairs	2187

6.5 Simulation Scenarios

The following six scenarios will be carried out to evaluate the performance of the proposed algorithms of the framework:

1. Proposed Bandwidth Management using IP/MPLS model for Multimedia Services.
2. Conventional FIFO and WFQ Algorithms in IP/MPLS.
3. Proposed Fuzzy based Packet Scheduling Algorithms (2 MB).
4. Proposed Fuzzy based Packet Scheduling Algorithms (5 MB).
5. Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (2 MB).
6. Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB).

The following subsections describe these five scenarios in details.

6.5.1 Scenario 1: Proposed Bandwidth Management using Multimedia Services in the MPLS network

This scenario involves four schemes of MPLS model using an analytical tool of OPNET. The aim of this scenario addresses the performance of multimedia services in Multi-Protocol Label Switching (MPLS) network model using Traffic Engineering (TE) for minimizing the delay with the aid of efficient throughput. These schemes are mention as follows:

1. Bandwidth Management using baseline IP/MPLS Model for Mobile Wireless Network.
2. Proposed Resource Reservation Protocol Tunnelling Extension in MPLS for sustainable Mobile Wireless Networks.
3. Proposed Approach to Label Distribution Protocol Signalling using Multimedia Services for Bandwidth Allocation.

4. Proposed Bandwidth management techniques using LDP signaling and LSP for bandwidth allocation in IP/MPLS model.

These scenarios involve bandwidth management techniques using traffic engineering only at the core of the MPLS network. The aim of this scenario is to show how the proposed schemes can be compared with the conventional IP network in terms of the throughput, packet delivery ratio, end-to-end delay, and packet delay variation. Here the results of the proposed schemes are obtained and will be presented in the results section. The performance metrics used for this purpose are considered in the following expression for mean delay and relationship between many sources and service utilization in the network. More details about the delay and jitter models can be seen in Appendix B.

Let pps = packets/s, bps = bits/s

Therefore,

$$pps * average\_packet\_size = bps$$

$$Bandwidth\ requirement\ (video) = frame / s * frame\_size * pixel\_resolution$$

$$Bandwidth\ (voice) = total\_packet\_size * pps$$

$\square$ = utilization,

$\square$ = demand for the resource per unit time,

$\square$ = supply of the service provided or capacity of the system.

d = demand of the resource.

Let the Average rate and utilization be A_r and ρ, respectively.

C = buffer service capacity (bits/s).

n = number of sources generating background traffic.

$$n = \frac{\square * C}{A_r}$$

but, $$C = \frac{\square}{\square}\left(bits/_{s}\right) \tag{34}$$

$$\square = \frac{no\_of\_packet}{mean\_arrival\_time} = \left(packets/_{s}\right)$$

(35)

$$Mean\_delay(w) = \frac{\square 1}{\square C - \square} = \frac{\square 1}{\frac{C}{1/\square} - \square}$$

$$n = \left(\frac{d}{\square}\right) * \left(\frac{1}{A_r}\right) * \left(\frac{\square}{\square}\right)$$

(36)

6.5.2 Scenario 2: Conventional FIFO and WFQ Algorithms in IP/MPLS

This involves the comparative study of the baseline FIFO and WFQ packet scheduling schemes using real-time traffic. The motivation for these two schemes is to be able to compare them and select the most suitable one for the proposed architecture. The aim of this scenario is to show the effects of these schemes on the number of sources and number of receivers, average throughput, and average end-to-end delay. These schemes will be evaluated in both IP and MPLS environments to see better performance.

6.5.3 Scenario 3: Proposed Fuzzy based Packet Scheduling Algorithms (2 MB).

After the performance of the (scenario 2) and separate schemes of baseline FIFO, WFQ, and fuzzy-based packet scheduling (scenario 3) using non-real-time traffic have been evaluated. In this case, they are also compared for further variation in the performance with the primary aim to show the difference between the two scenarios. Moreover, there is a need to show an effect on average network throughput and end-to-end delay. These schemes will also be considered in both IP and MPLS environments and compared to view the effects on the performance metrics.

6.5.4 Scenario 4: Proposed Fuzzy based Packet Scheduling Algorithms (5 MB).

The previous scenarios 1 to 2 were based on conventional schemes, which may be affected in terms of performance as the demand for the services increases when a specific limit is met. However as explained previously, it is imperative that a more sophisticated Packet Scheduling Algorithm is used to make the most use of the benefits of the proposed framework. In addition to this, an expert system operates intelligently to cope with the complexity of the networks. Based on the simulation results of scenarios 3 and 4, which will be explained in Section 6.6.2 and 6.6.3. It was observed that when fuzzy is introduced to the scheme, WFQ outperforms FIFO using non-real traffics in terms of average throughput and average end-to-end delay. Therefore, it is the most suitable one for our system model, and in this scenario, the proposed Fuzzy based packet scheduling is applied at the interface of core routers in IP/MPLS networks, which is chosen for its best performance. The main deficiency of pure fuzzy systems is their incapacity for learning because they have no memory. This can be solved by associating fuzzy systems with techniques that can perform learning using a neural network.

6.5.5 Scenario 5: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (2 MB).

In this scenario, only Neuro-Fuzzy based Packet Scheduling can take the merits of both the neural network and fuzzy system to overcome the complex computational problem exhibited by Fuzzy system alone. Furthermore, the neuro-fuzzy system's operation is faster as compared with the previous scenario. The more complex the network becomes, the more difficult for the fuzzy system. However, the performance of the neuro-fuzzy system is to improve the efficiency of the neuro-fuzzy based packet scheduling algorithm for IP/MPLS network service for LSP selection.

6.5.6 Scenario 6: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB).

In this scenario, the advance in the expert system will be presented, where link bandwidth is increased to 5 MB. This will be compared with the previous scenario to see the effect of increasing link bandwidth on the attributes of the network. Similarly, it has been stated that when the network is involved, the fuzzy system is inefficient due to the multiple input variables.

6.6 Results Evaluation

The results of the proposed MPLS framework and performance analysis are provided in this section. All the results of the five scenarios introduced in Section 6.5 will be presented in this section.

6.6.1 Results for Scenario 1: Bandwidth Management using Multimedia Services in the MPLS network.

6.6.1.1 *Impact on network throughput*

Figure 6-6 shows that the overall throughput can stabilize for both voice and video traffics. It appears that a wide gap between video conferencing traffic (THRvc) sent and video traffic received with the effect of more packets transmitted than packets received. In the case of voice traffic (THRv), a slight difference occurs between many packets transmitted and packets received which indicates both have low throughput within the range of 0 and 500000 packets/s. In general, from Figure 6-6, it provides a valuable opportunity to advance the understanding of the throughput of video conferencing traffic, which is considerably higher than that of voice traffic due to the level of bandwidth consumption. However, voice transmission is quite easily manageable as compared to video transmission. Subsequently, voice transmission is also used in both IP and MPLS for monitoring the performance in terms of how fast is the packet forwarding at the core of the network. The comparison of these networks using voice is shown in Figure 6-7.

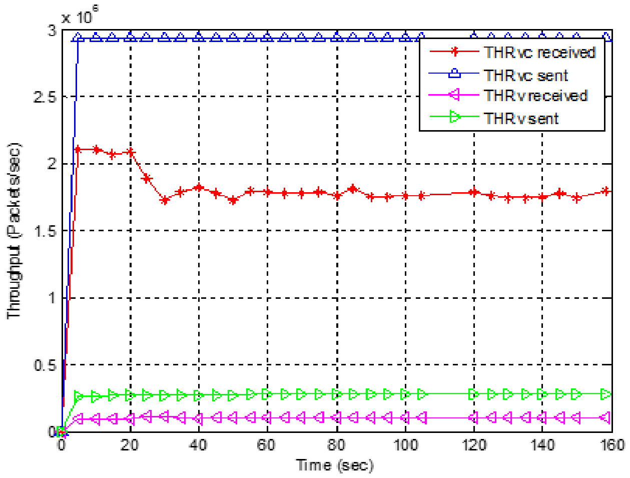


Figure 6-6. Comparison of Throughput sent and received for voice and video traffics

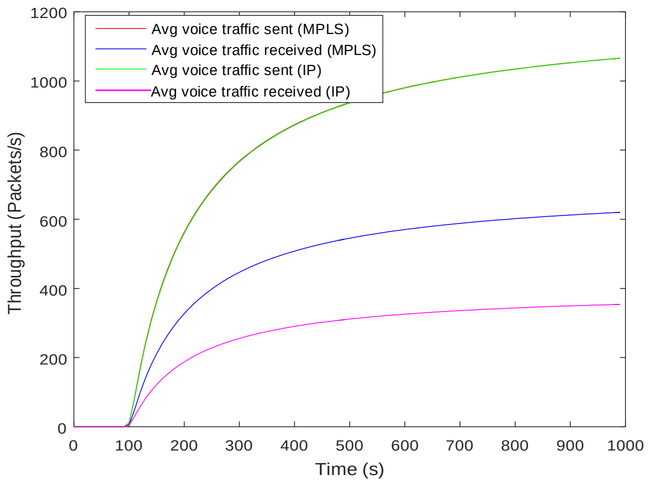


Figure 6-7. Comparison of Throughput sent and received for voice in both IP and MPLS networks

6.6.1.2 Impact on end-to-end delay and delay variation

Figure 6-8 illustrates the effects of end-to-end packet delay and delay variation (jitter) on both video and voice traffics. As for the packet delay variation, there is an uprising to the peak of about 2.8 s at the initial stage of the video traffic which later drop down by fluctuation to an approximate value of 1.1 s while a remarkable constant delay exists for voice with a minimum of 0 and maximum of about 0.75 s. This is because the data rate of the voice is lower than that of the video. Besides, the end-to-end delay appears to follow a different pattern in which that of the voice able to reach up to 1.2 s and video has a peak of 0.8 s with a small variation. The comparative study of the effect of end-to-end delay in IP and MPLS networks is shown in Figure 6-9.

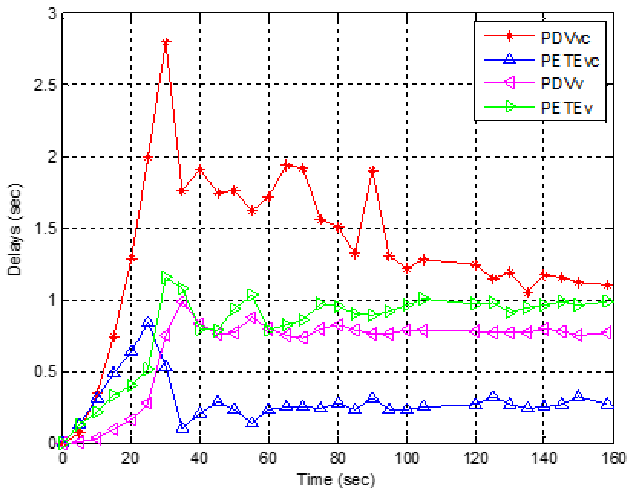


Figure 6-8. Comparison of both Packet delay variation and End-to-End delay for voice and video traffics

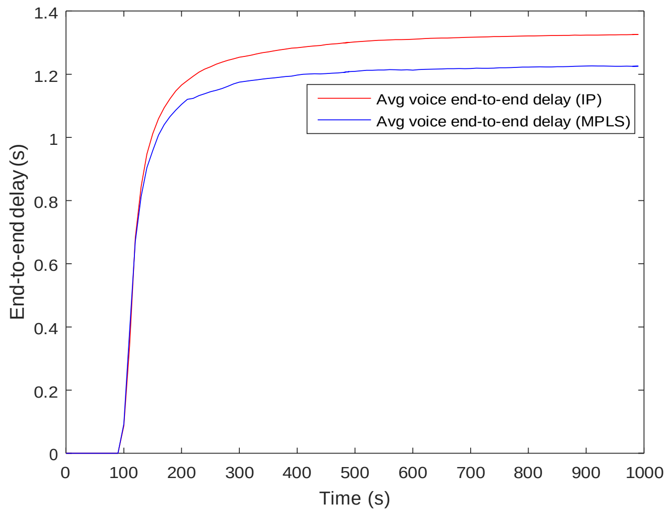


Figure 6-9. Comparison of End-to-End delay for voice in IP and MPLS networks.

6.6.1.3 Impact on the number of sources

As indicated above, the mean delay is inversely proportional to the number of packets per second in equation (34) while equation (35) shows that number of sources is directly proportional to the ratio of several packets/s and mean service requirement. In Figure 6-10, the mean service utilization is inversely proportional to the number of sources. There is variation in the service supplied to the customers at different utilization percentages (25%, 50%, and 75%).

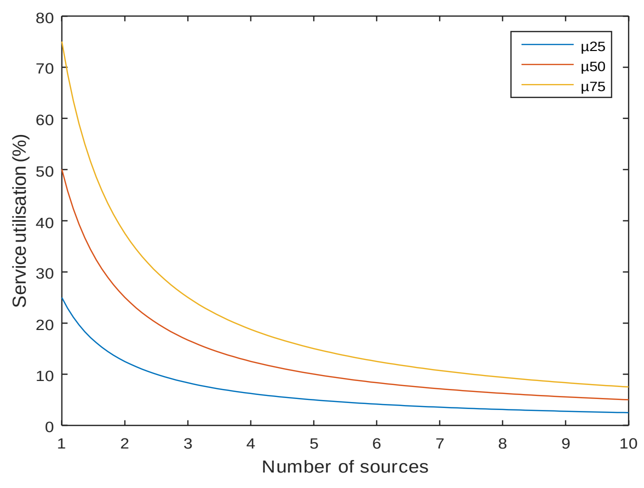


Figure 6-10. Service utilization versus some sources.

6.6.1.4 *Impact on service utilization*

All the signaling messages generated by request will contain the identification (ID): the reply to the signaling messages will also have this ID. It can be observed that the estimated bandwidth value of 51.498 MB is evenly distributed and still have a reservation on the links as shown in Figure 6-11.

Similarly, the percentage of bandwidth utilization using mesh LDP configuration is shown in Figure 6-12. From the results of utilization of bandwidth, it shows that better utilization of bandwidth with reservation of 28.57 % as compared with reservation of 16.67% in the first and second circle view respectively. This indicates that moderate bandwidth utilization can be used to control congestion in the network. The analysis of bandwidth estimation and allocation can be seen in Appendix A.

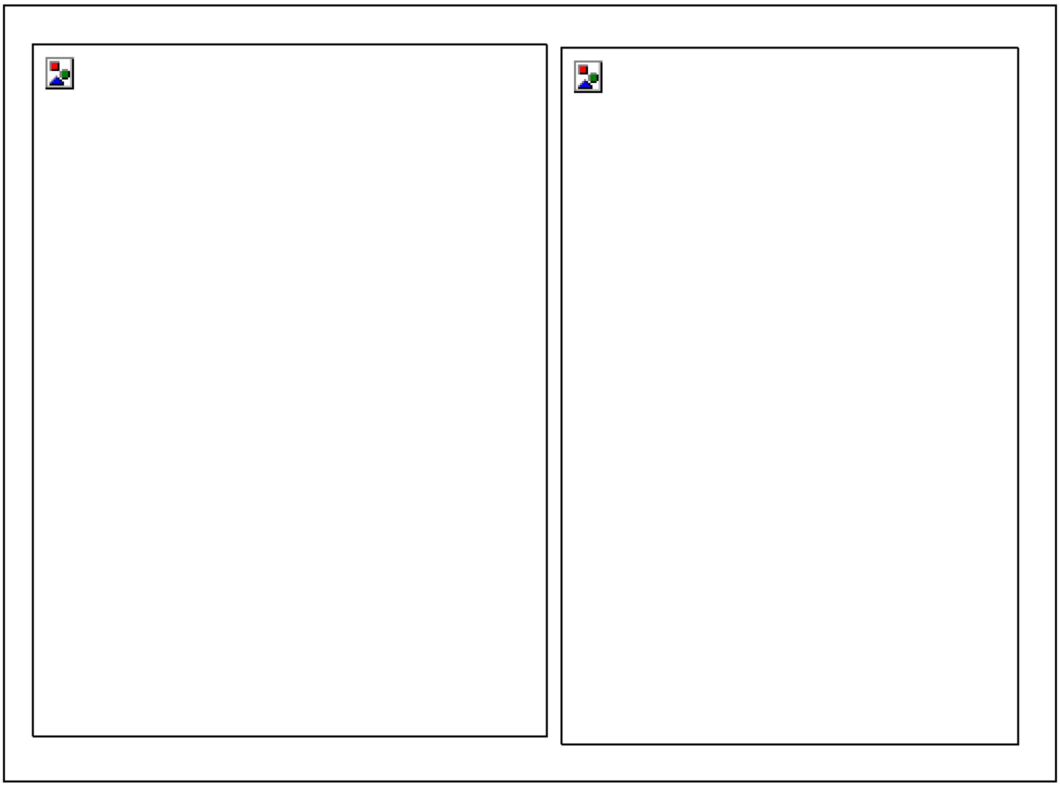


Figure 6-11. Percentage of Bandwidth utilization using MPLS LSP

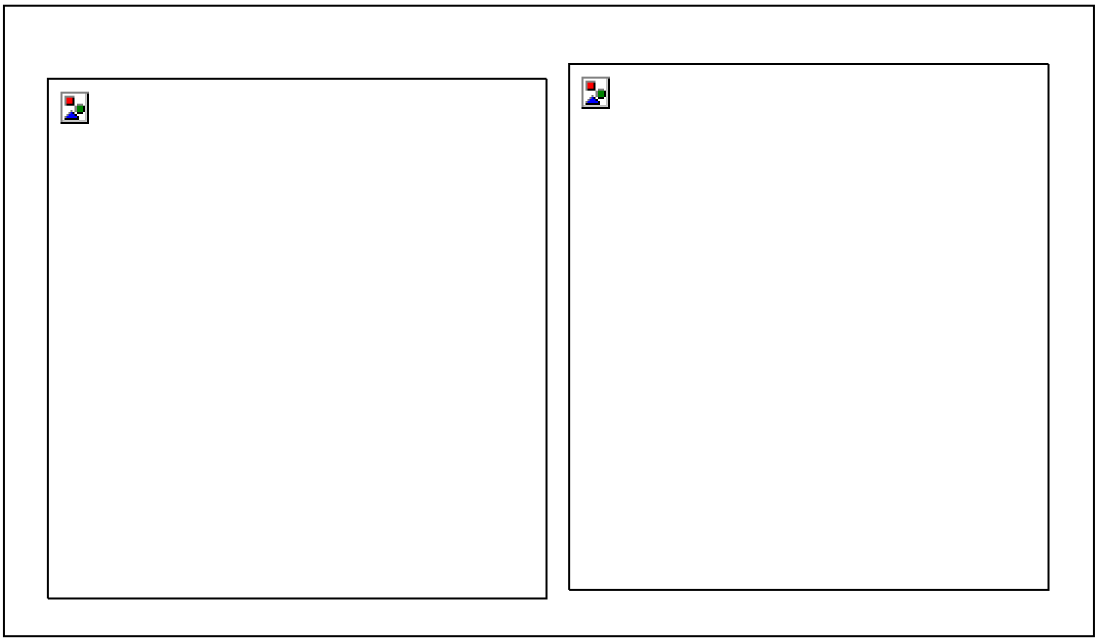


Figure 6-12. Circle view of bandwidth distribution.

It can be observed that the MPLS model was simulated in order to verify its performance with LDP configuration using multimedia services. This served as the baseline for further change in configurations. The result of the performance indicates that there is an absolute packet delivery from ingress operating point to the egress endpoint. As for the results of the implementation, the MPLS baseline and modified MPLS networks with two random seeds scenario (seed 128 and seed 110) using configurations of voice and video conference are used which yielded results as shown in Figure 6-13 and Figure 6-14. There is an infinite variation in the result of traffic received at the destination (Egress) point. Higher throughput is experienced in the modified network. This is due to the label distribution protocol being configured at the core routers (LSRs) to allocate bandwidth uniformly as shown in Figure 6-11. As can be seen from Figure 6-14, all the throughput received increase rapidly to an average of about 550 kbps and 490 kbps for video conferencing configuration. There exists a considerable difference in the received traffic with an average value of 440 kbps and 240 kbps respectively. This is an indication of constant traffic flows.

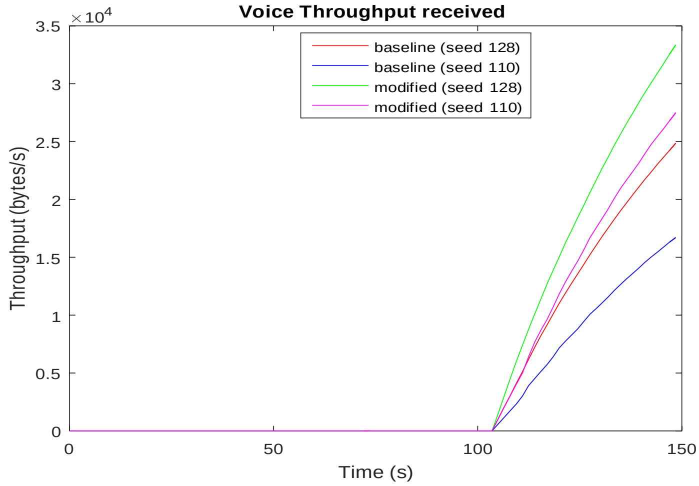


Figure 6-13. Average Throughput for Voice in MPLS.

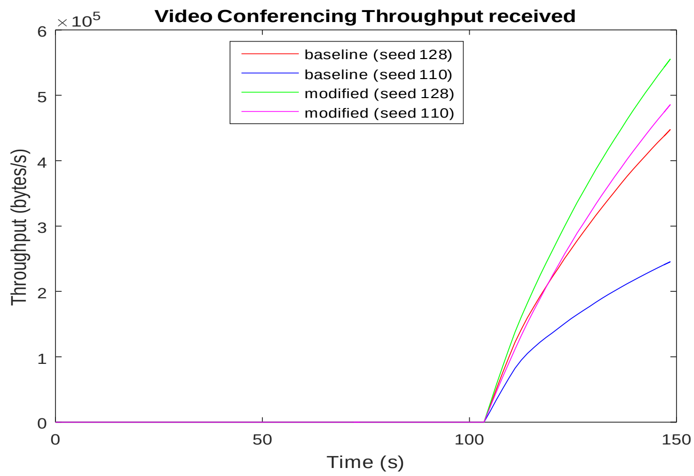


Figure 6-14. Average Throughput for Video Conferencing in MPLS.

Both Figure 6-15 and Figure 6-16 show packet end-to-end delay for both video and voice traffic. As for the packet delay variation, there are steady and low values resulting from the modified network with LDP of the average peak of about (1.0 s / 0.18 s) for voice and (0.6 s / 0.3 s) for video as compared with the baseline without LDP configuration. However, the baseline result for the video appears to decrease sharply. The analysis of delay and jitter models can be seen in Appendix B.

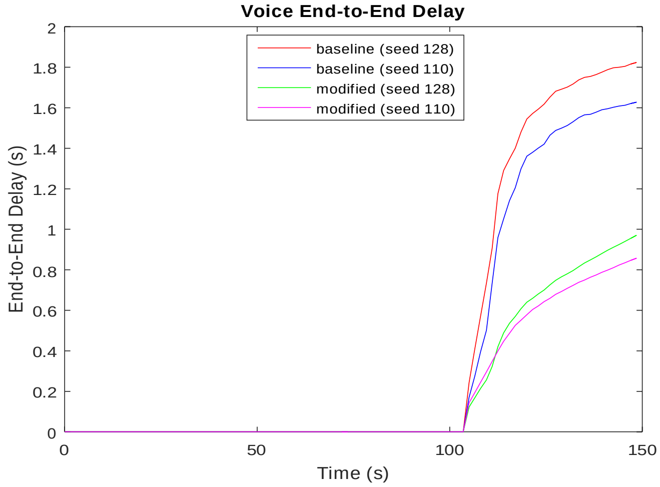


Figure 6-15. Average End-to-End delay for Voice.

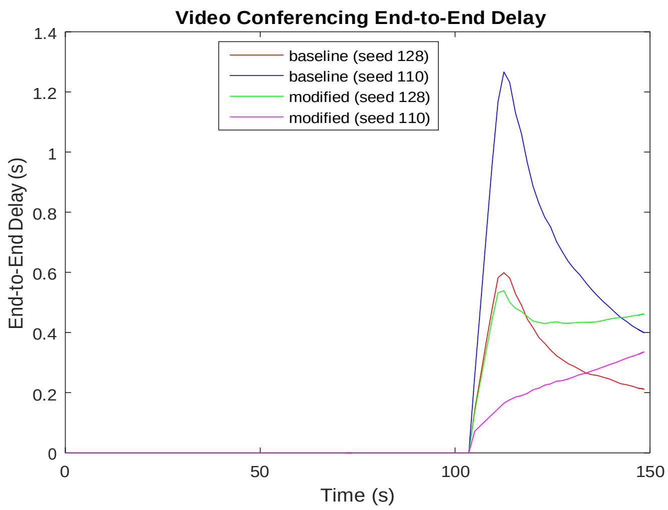


Figure 6-16. Average End-to-End delay for Video Conferencing.

6.6.2 Results for Scenario 2: Conventional FIFO and WFQ Algorithms in IP/MPLS.

6.6.2.1 *Impact on average network throughput.*

It can be seen from Figure 6-17 that there is no significant change in both IP and MPLS networks with the introduction of scheduling algorithms using VoIP. This is due to the constant sending rate of data from source to destination as well as not commonly impacted by packet re-ordering. Therefore, the impact of delay on bandwidth requirement for each flow could be used to distinguish this scenario.

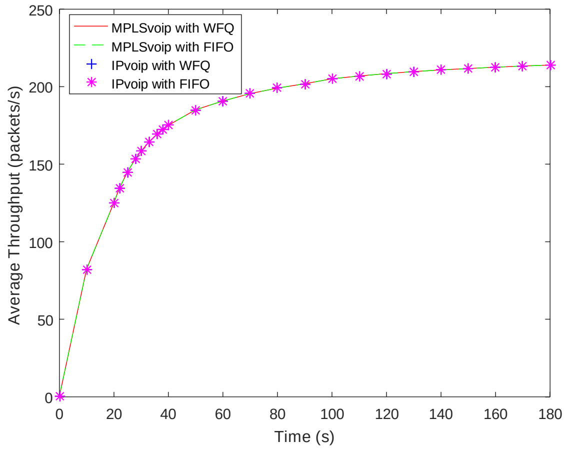


Figure 6-17. Average Throughput for VoIP in IP/MPLS networks using FIFO and WFQ Algorithms.

6.6.2.2 *Impact on end-to-end delay.*

Figure 6-18 shows the effects of average end-to-end delay on the baseline FIFO and WFQ in IP/MPLS networks. There is an absolute difference of close to 250 ms observed between WFQ in MPLS and WFQ in IP using the real-time application of VoIP. Furthermore, decrease in the value of delay approximate to 410 ms between FIFO in MPLS and FIFO in IP respectively.

High level of delay can be attached to the sensitivity of real-time traffic such as VoIP.

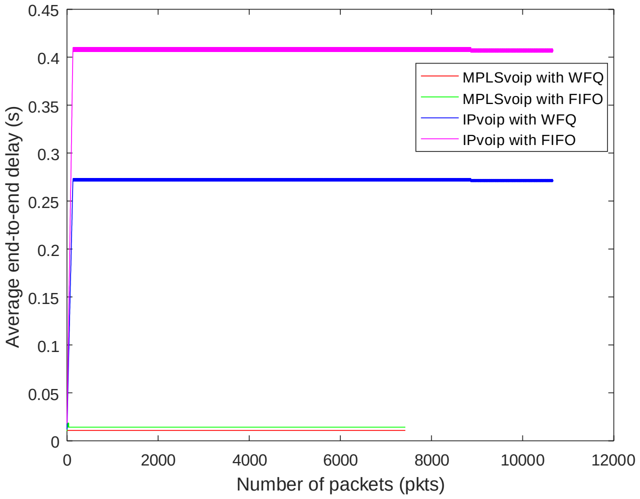


Figure 6-18. The average end-to-end delay for VoIP.

6.6.2.3 *Impact on packet loss ratio.*

The packet loss ratio for schemes is shown in Figure 6-19. The packet loss ratios for both schemes using VoIP in the MPLS network are 12%, 7% while for IP network are 13%, 11%. The differences in the values are resulting from loss in application end system of the MPLS. Therefore, WFQ in MPLS (MPLSwfq) performs better than other scenarios (MPLSbaseline, IPbaseline, IPwfq).

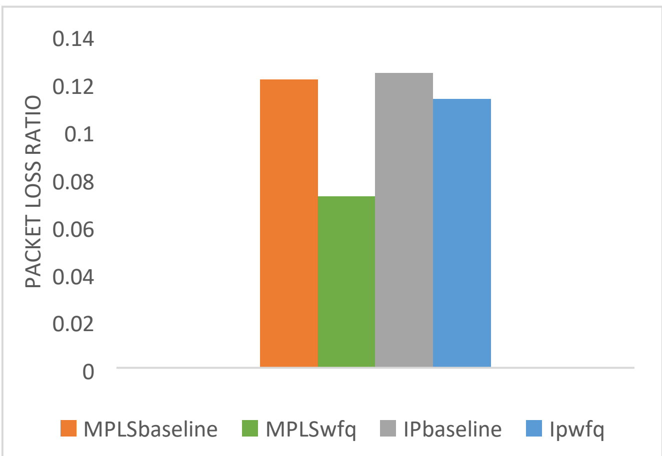


Figure 6-19. Packet Loss Ratio for VoIP.

6.6.3 Results for Scenario 3: Proposed Fuzzy based Packet Scheduling Algorithms (2 MB).

6.6.3.1 *Impact on average network throughput.*

The comparison of average throughput versus simulation time in IP/MPLS with the incorporation of FIFO (baseline) and WFQ algorithms using FTP traffic is shown in Figure 6-20 and Figure 6-21. For both schemes, the average throughput increases exponentially because more traffic utilizes more resources. FTP traffic is created on top of a TCP connection between node 0 and node 10. MPLS performed better than IP with both algorithms because of its fast packet forwarding mechanism. There is an indication of a 43% increase with WFQ and 49% increase in the case of FIFO. In the fuzzy scenario using link bandwidth of 2 MB, an increase of about 46% with WFQ and an approximation of 56% increment with FIFO are illustrated.

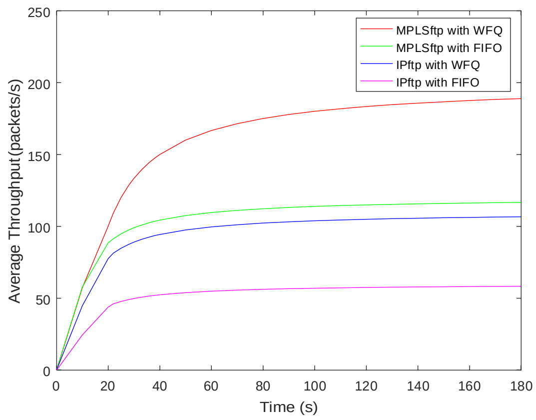


Figure 6-20. Average Throughput of IP/MPLS with baseline FIFO and WFQ for 2 MB.

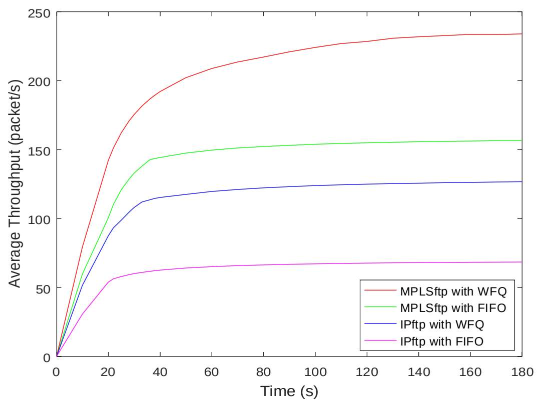


Figure 6-21. Average Throughput of IP/MPLS (Fuzzy) with FIFO and WFQ for 2 MB.

6.6.3.2 *Impact on average end-to-end delay.*

Figure 6-22 shows the impact of average end-to-end delay in the IP/MPLS network with packet scheduling algorithms using a non-real-time application of FTP. A steady delay observed at 1.9 ms with WFQ in MPLS environment as compared with a variation of about 86% with WFQ in IP network. This implies that WFQ has an absolute low delay than FIFO due to fast packet forwarding mechanism. Furthermore, the delay still has a similar relationship in the performance of FIFO scheduling algorithms within IP and MPLS environments. However, there is small difference as compared with MPLS operations. This is due to the non-sensitive behaviour of FTP applications over TCP. Lower delay is seen after introducing fuzzy to the system in Figure 6-23. The MPLSftp with WFQ, MPLSftp with FIFO, IPftp with WFQ, and IPftp with FIFO have about 1.3 ms, 1.8 ms, 5.9 ms, and 7.8 ms respectively. Therefore, MPLS operations perform better than IP with a scenario of 2 MB of the link bandwidth using conventional and fuzzy-based packet scheduling.

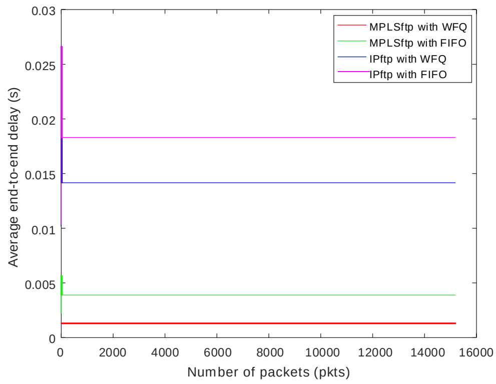


Figure 6-22. Average End-to-End delay in IP/MPLS with baseline FIFO and WFQ for 2 MB.

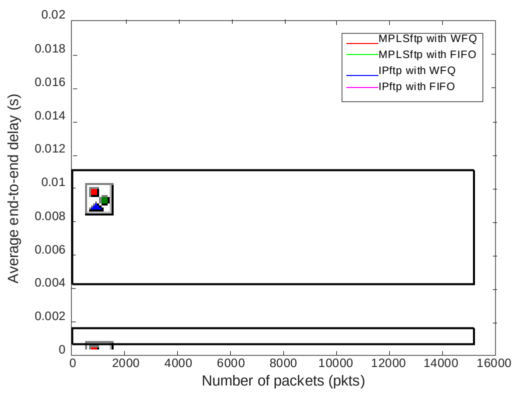


Figure 6-23. Average End-to-End delay in IP/MPLS (Fuzzy) with FIFO and WFQ for 2 MB.

6.6.3.3 Impact on the reliability of the system.

In Figure 6-24, the impact of the reliability of the system on different schemes is shown. The scheme of WFQ with fuzzy happened to be the most reliable for system. This is because of low delay with high level of service availability. For example, FIFO has lowest value of reliability showing that it is greatly affected by the high delay and low service.

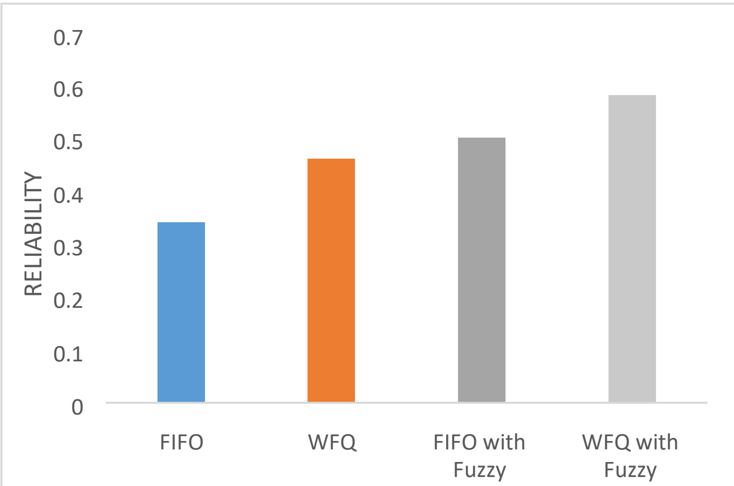


Figure 6-24. Reliability of the system.

6.6.3.4 Impact on service utilization.

Figure 6-25 indicates how the service utilization is affected by the schemes. Since the number of sources has great impact on service utilization, it implies that more packets will be transmitted and the increase in utilization would occur simultaneously. Therefore, a system with high throughput will have moderate value of service utilization, which can be illustrated with scheme of WFQ using fuzzy.

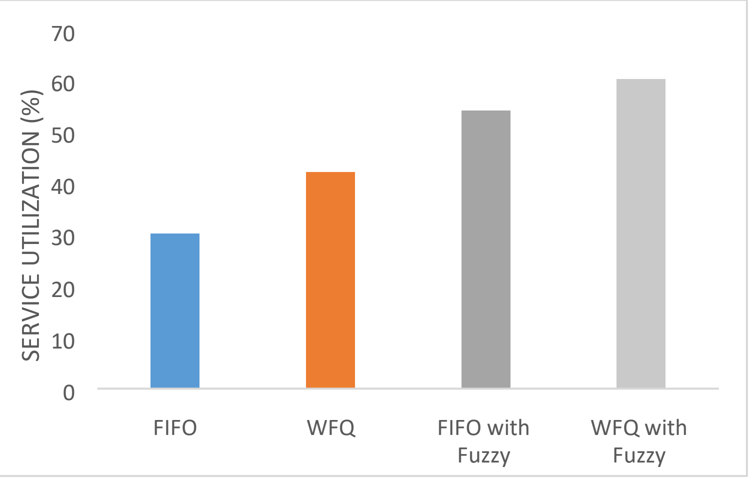


Figure 6-25. Effects of the Fuzzy on service utilization.

6.6.3.5 Impact on the queue waiting time.

The effect of the average queue waiting time on various schemes in the MPLS network is observed in Figure 6-26. There is drastic decrease in queue waiting time as the scheme improved from the conventional packet scheduling (FIFO and WFQ) to fuzzy-based schemes. This implies that when fuzzy is introduced to FIFO, the queue time reduces by 48% approximately. While in the case of WFQ, it reduces by about 47.4%.

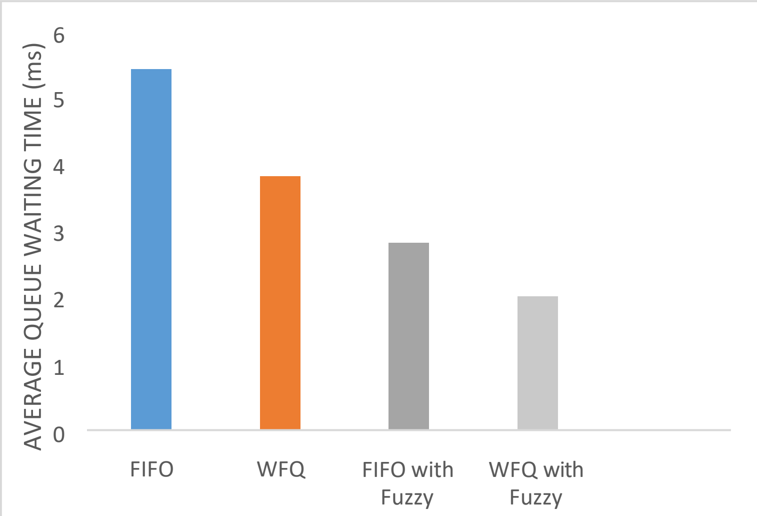


Figure 6-26. Effects of the Fuzzy on average waiting time.

6.6.3.6 Impact on link cost.

Figure 6-27 shows the effect of the link cost in IP/MPLS with the scheduling algorithms of FIFO (baseline) and WFQ. As the number of packets per second increases, the packet loss in the network also increases since the long queue leads to more packet drops. Moreover, more packet loss will be incurred over longest path of the link cost most especially if there is long queue.

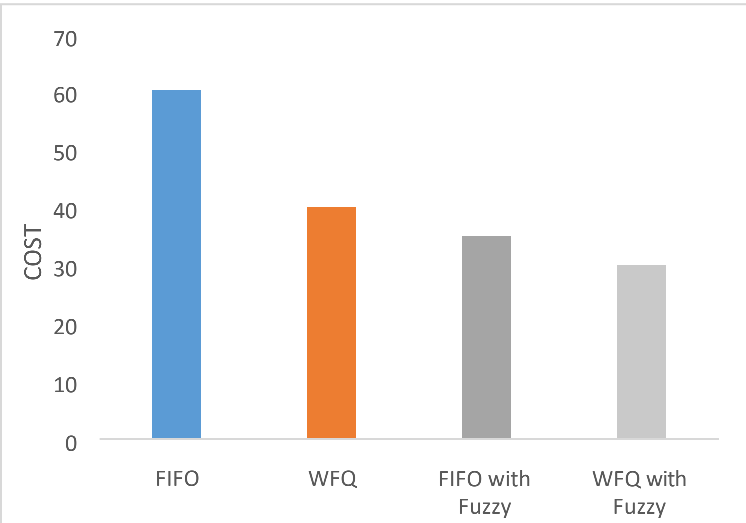


Figure 6-27. Link cost for all the schemes.

6.6.4 **Results for Scenario 4: Proposed Fuzzy based Packet Scheduling Algorithms (5 MB)**.

6.6.4.1 ***Impact on average network throughput***

The effect of packet scheduling on average network throughput is investigated by comparing the baseline with fuzzy based packet scheduling. In the conventional packet scheduling, the approximate value of 35% increase with WFQ and FIFO with 44% increment as shown in Figure 6-28. The link bandwidth of 5 MB is used in this scenario with close to 43% increment for WFQ and about 50% increase with FIFO as shown in Figure 6-29.

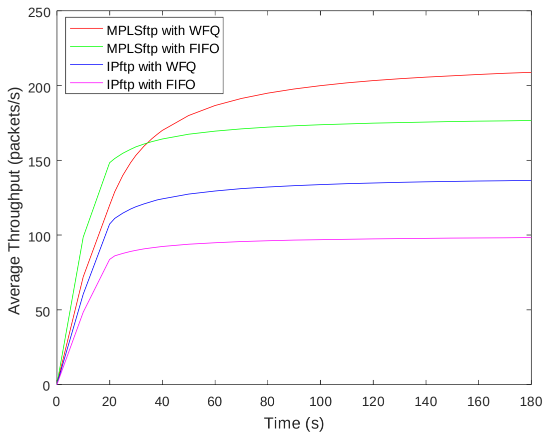


Figure 6-28. Average Throughput of IP/MPLS with baseline FIFO and WFQ for 5 MB.

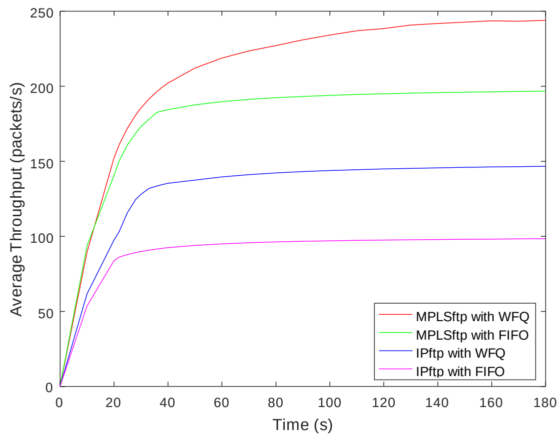


Figure 6-29. Average Throughput of IP/MPLS (Fuzzy) with FIFO and WFQ for 5 MB.

6.6.4.2 Impact on average end-to-end delay.

Figure 6-30 and Figure 6-31 illustrates the performance of IP/MPLS in terms of average end-to-end delay for link bandwidth of 5 MB. This is done by comparing the conventional packet scheduling scheme with the fuzzy-based scheme. The average end-to-end delay of WFQ in IP/MPLS and FIFO in IP/MPLS are about 0.1 ms, 3.1 ms, 0.9 ms, and 3.2 ms, respectively. There is close gap between IP scenarios using FTP application.

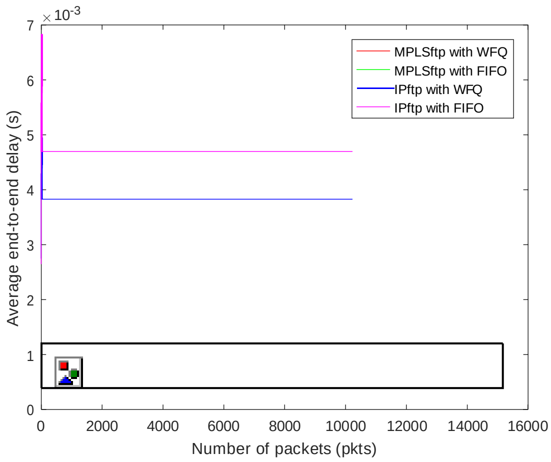


Figure 6-30. Average End-to-End delay in IP/MPLS with baseline FIFO and WFQ for 5 MB.

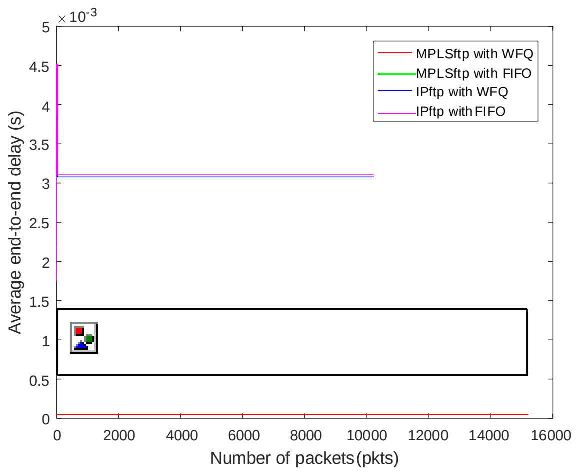


Figure 6-31. Average End-to-End delay in IP/MPLS (Fuzzy) with FIFO and WFQ for 5 MB.

6.6.4.3 Impact on the reliability of the system

Figure 6-32 shows the reliability for packet delivery in the MPLS network using a non-real-time application of FTP. The algorithms of FIFO and WFQ with the introduction of fuzzy have 66% and 85% reliabilities respectively. On the other end, FIFO and WFQ without fuzzy having about 45% and 56% reliabilities for packet delivery. The value of low reliability could be because of a failure in the link, low bandwidth allocation, and increase in packet loss.

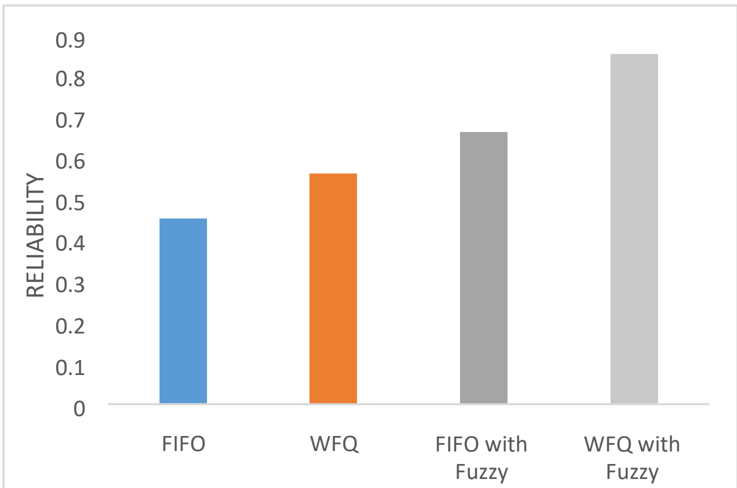


Figure 6-32. Reliability of the system (5 MB).

6.6.4.4 Impact on service utilization

Figure 6-33 shows the effects of the Fuzzy on service utilization within the MPLS. As can be seen, service utilization increases rapidly with the application of fuzzy on WFQ and FIFO scheduling algorithms within the MPLS environment. However, FIFO and WFQ without fuzzy have lower values of service utilization as compared with an improved intelligent system. It is due to the fast response in packet delivery of FTP applications over TCP.

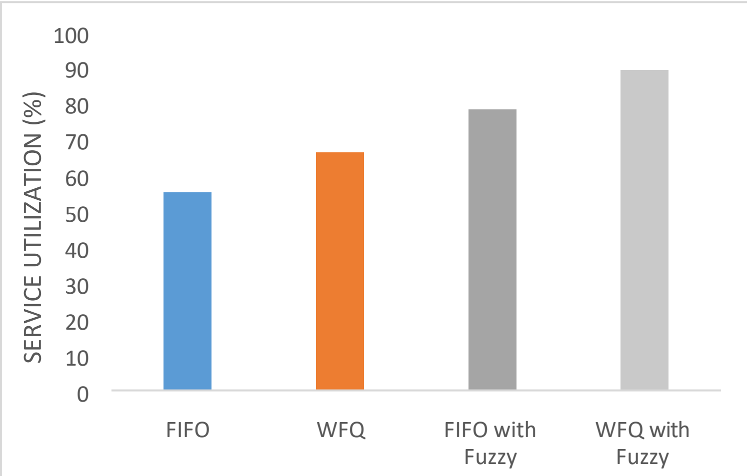


Figure 6-33. Effects of the Fuzzy on service utilization (5 MB).

Figure 6-34 gives an illustration of a scenario of an average queue waiting time in the MPLS network with packet scheduling algorithms using a non-real-time application of FTP. It is observed that there is a drastic reduction in waiting time due to the application of fuzzy to FTP. FIFO reduces by about 38% while it is 64% reduction for WFQ. This implies that WFQ has an absolute low waiting than FIFO to better scheduling and allocation of services.

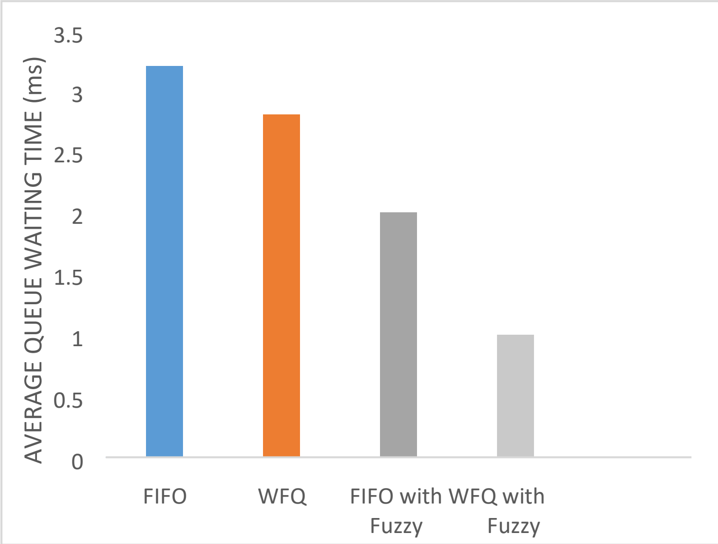


Figure 6-34. Effects of the Fuzzy on average waiting time (5 MB).

6.6.4.5 Impact on link cost

The link cost in the MPLS network with packet scheduling algorithms using FTP is as shown in Figure 6-35.The lower the link cost, the better for the fast route in the MPLS network. Fuzzy based algorithms indicate an improve link cost because of computation and selection of the lowest link cost, which served as the best route for packets delivery.

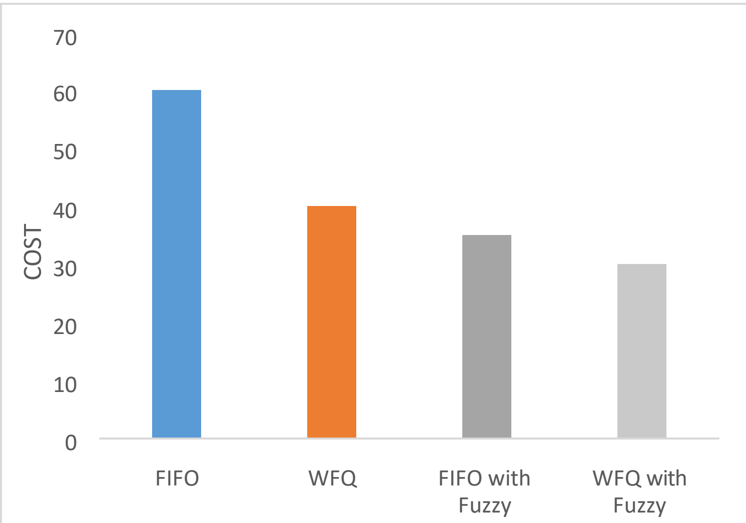


Figure 6-35. Link cost for all the schemes (5 MB).

6.6.5 Results for Scenario 5: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (2 MB).

6.6.5.1 *Impact on average network throughput*

The effects of packet scheduling on average network throughput using neuro-fuzzy with the link bandwidth of 2 MB in IP/MPLS environment is illustrated in Figure 6-36. The introduction of neuro-fuzzy in MPLS using WFQ scheme led to an increase in the maximum value of the average throughput of 21% from 189pps to 240pps. Similarly, the maximum value of average throughput is increased by about 19% from 106pps to 133pps in IP using FIFO scheme.

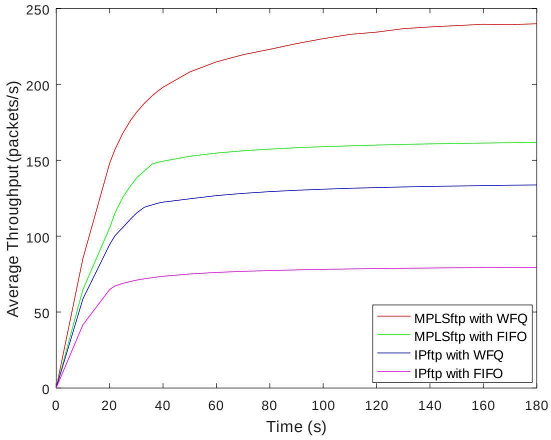


Figure 6-36. Average Throughput of IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 2 MB.

6.6.5.2 Impact on average end-to-end delay.

Figure 6-37 depicts the effect of average end-to-end delay on different scenarios (baseline and neuro-fuzzy schemes) in IP/MPLS networks. The MPLSftp with WFQ, MPLSftp with FIFO, IPftp with WFQ, and IPftp with FIFO have about 0.3 ms, 1.4 ms, 1.6 ms, and 3.1 ms respectively. Therefore, MPLS operations perform better than IP with a scenario of 2 MB of the link bandwidth using neuro-fuzzy based packet scheduling. This is due to a decrease of about 81% in average end-to-end delay for WFQ scheme and reduced by about 55% for FIFO scheme.

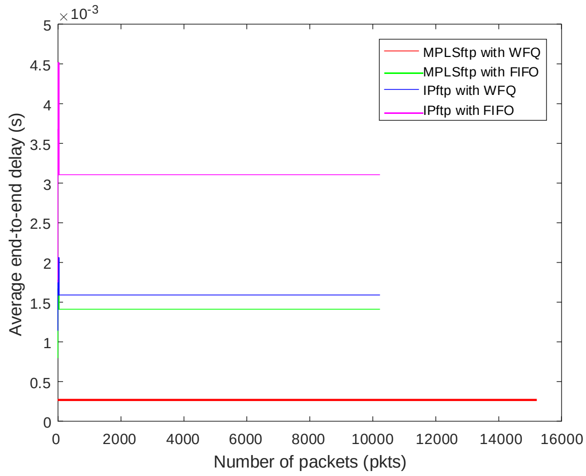


Figure 6-37. Average End-to-End delay in IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 2 MB.

6.6.5.3 Impact on the reliability of the system.

Figure 6-38 illustrates the impact of the reliability of the system on neuro-fuzzy based packet scheduling. This shows slight increase and better performance in the reliability of the system for neuro-fuzzy scheme ranging from normalized weight of 0.2 to 1. This may be due to either improvement in packet forwarding along a path (LSP) or low packet drop in the path.

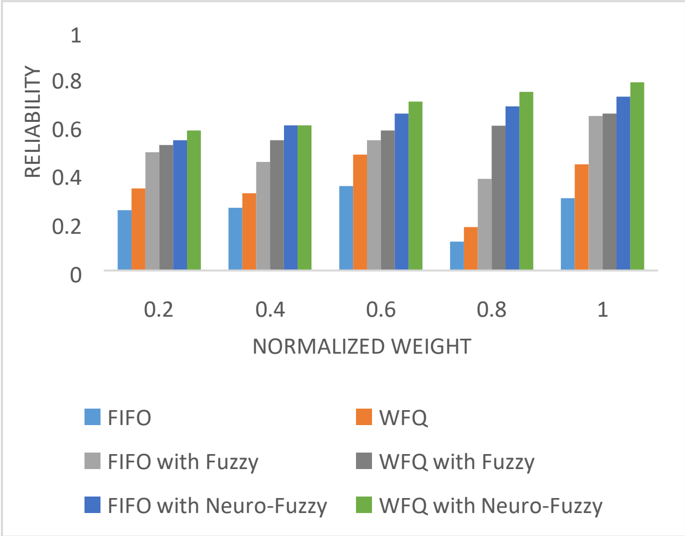


Figure 6-38. Reliability for packet delivery of the schemes (2 MB).

6.6.5.4 Impact on service utilization.

Figure 6-39 highlights the impact of service utilization on neuro-fuzzy based packet scheduling. In the example for traffic flow 4 where both FIFO and WFQ with neuro-fuzzy have the same 70% for service utilization as compared with traffic flow 5, in which the conventional FIFO and WFQ possessed the value of 45%.

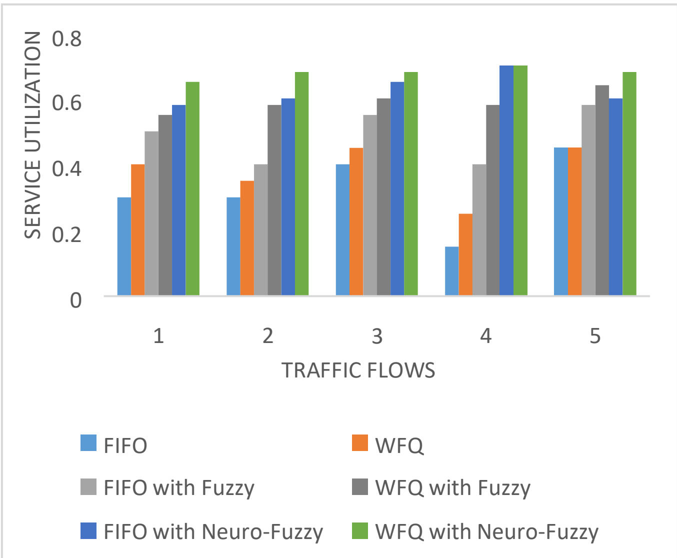


Figure 6-39. Service utilization of the schemes (2 MB).

6.6.5.5 Impact on link cost

Figure 6-40 highlights the impact of link cost on neuro-fuzzy based packet scheduling. The lowest link cost appears to be same for both WFQ with fuzzy and neuro-fuzzy in traffic flow 2, which is followed by the same values of link cost for both FIFO with fuzzy and neuro-fuzzy schemes. This could be due to intelligently computing and selecting the lowest link cost.

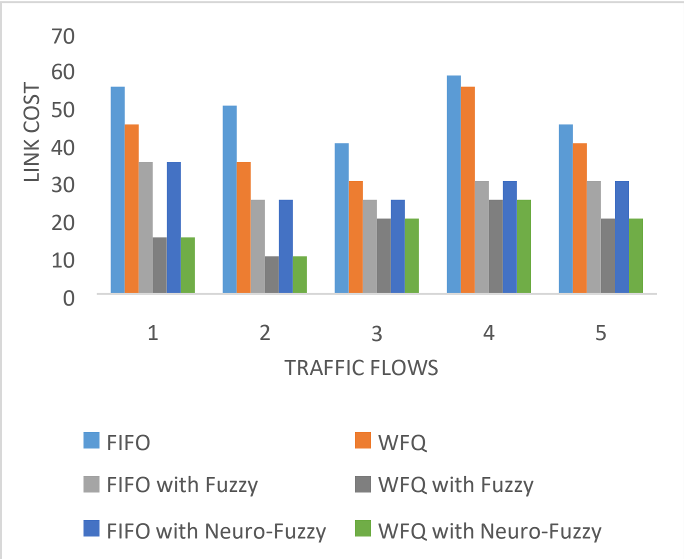


Figure 6-40. Link cost of the schemes (2 MB).

6.6.5.6 Impact on packet loss ratio

Figure 6-41 shows the impact of packet loss ratio on neuro-fuzzy based packet scheduling. It appears that there is an appreciable packet loss ratio throughout the schemes. However, neuro-fuzzy based schemes still have lowest value as compared with fuzzy-based and conventional schemes. This is because of low packet loss observed in the transmission of packets from source to the destination.

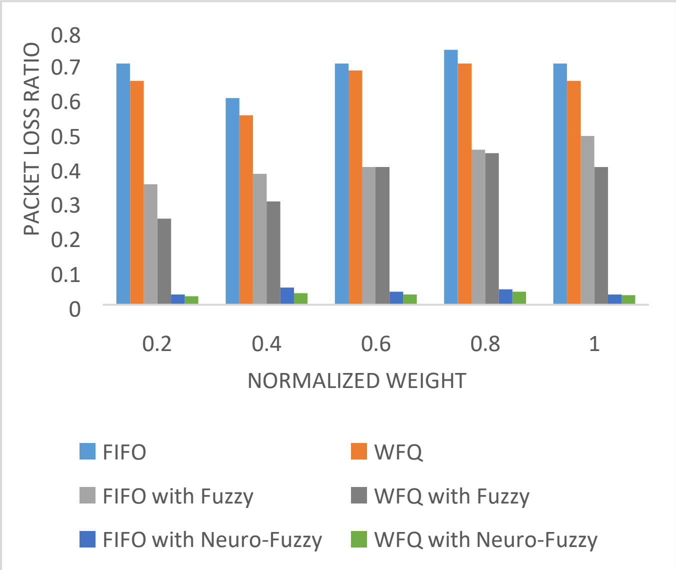


Figure 6-41. Packet loss ratio of the schemes (2 MB).

6.6.6 Results for Scenario 6: Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB).

6.6.6.1 *Impact on average network throughput.*

In Figure 6-20 , Figure 6-21, Figure 6-28, and Figure 6-29, the effect of the conventional and fuzzy-based packet scheduling on the average network throughput are already shown. As the link bandwidth increases, it shows a considerable growth in average throughput for neuro-fuzzy scheme, which is shown Figure 6-42. Furthermore, there is need to compare fuzzy-based packet scheduling scheme with that of neuro-fuzzy based packet scheduling. Considering the results from section 6.6.3 to section 6.6.6, neuro-fuzzy algorithm performs better than the fuzzy algorithm due to the inability of fuzzy-based packet scheduling scheme to support the multiple input parameters. Also, it lacks learning process in which neuro-fuzzy algorithm served as the best alternative. The results from Figure 6.36 to Figure 6.47 can be used to justify the performances both IP and MPLS environments.

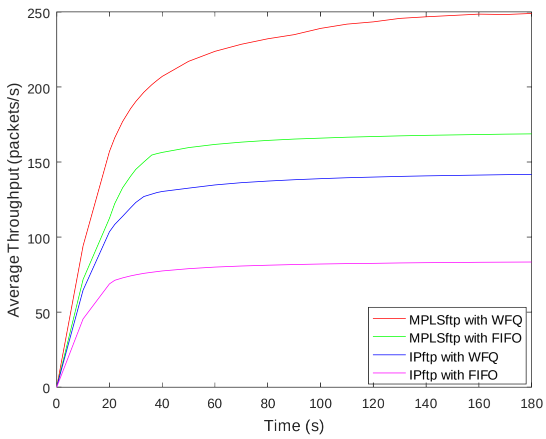


Figure 6-42. Average Throughput of IP/MPLS (Neuro-Fuzzy) with (FTP) for 5 MB.

6.6.6.2 Impact on average end-to-end delay.

As can be seen in Figure 6-43, variations in the value of average end-to-end delay were illustrated. The MPLSftp with WFQ, MPLSftp with FIFO, IPftp with WFQ, and IPftp with FIFO have steady delay of about 0.15 ms, 0.6 ms, 1.4 ms, and 2.6 ms respectively. The IP/MPLS scenarios possess difference of about 1.25 ms using FTP application with WFQ having approximate decrease of 89% while it appears that about 77% decrease with FIFO in IP/MPLS operation.

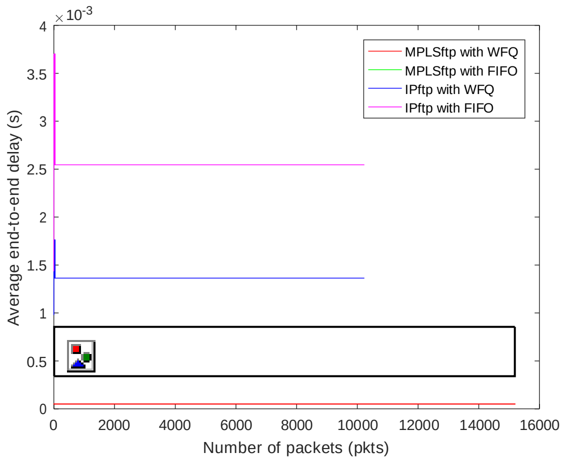


Figure 6-43. Average End-to-End delay in IP/MPLS (Neuro-Fuzzy) with FIFO and WFQ for 5 MB.

6.6.6.3 Impact on the reliability of the system.

In this case, it can be seen in Figure 6-44 that reliability has a positive effect because of introducing neuro-fuzzy at the interface of the routers. Therefore, both FIFO and WFQ are implemented with fuzzy and neuro-fuzzy schemes, respectively. For instance, at normalized weight of 0.2, about 10% of 7.5 of the overall system's reliability for neuro-fuzzy while 10% of 1.4 more for reliability at 1.0 of the normalized weight is observed resulting from low rate of failure in the overall system. Furthermore, the lowest value of reliability occurs at 0.8 of the weight indicating high packet loss or high failure rate could be the effect.

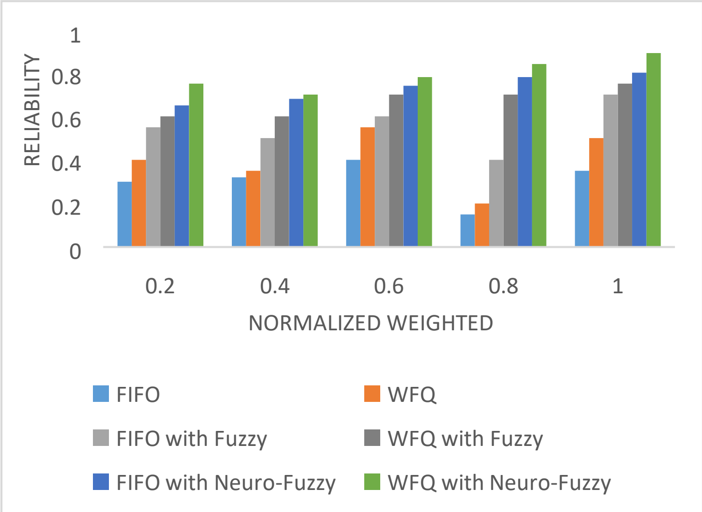


Figure 6-44. Reliability for packet delivery of the schemes (5 MB).

6.6.6.4 Impact on service utilization.

Figure 6-45 illustrates the effect of service utilization along the LSP created in the core of the network. Few flows from 1 to 5 are used to determine the consequence of underutilization and overutilization of service in a flow. It has been stated that WFQ and WRR solve the problems of bandwidth starvation for the low priority traffic, which is broadly caused by the FIFO.

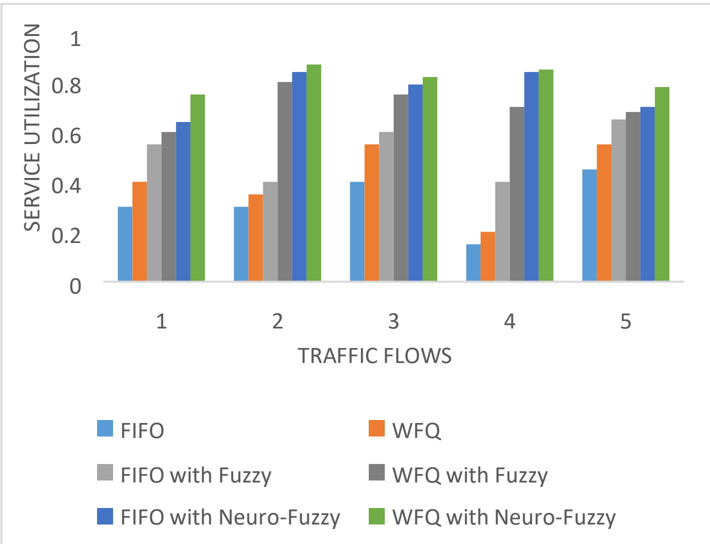


Figure 6-45. Service utilization of the schemes (5 MB).

6.6.6.5 Impact on link cost.

In Figure 6-46, the effects on link cost for the schemes are shown. This is relative to the traffic flows, which are given sequentially assuming no link failure. Flow 2 has the lowest link costs of 35, 25, and 10 for WFQ, WFQ with fuzzy, and WFQ with neuro-fuzzy, respectively. While flow 4 has the highest link costs of 20 more for WFQ and 15 more for WFQ with neuro-fuzzy respectively. Apart from FIFO, which indicates high link cost WFQ with neuro-fuzzy has appreciably improved in link cost than other schemes.

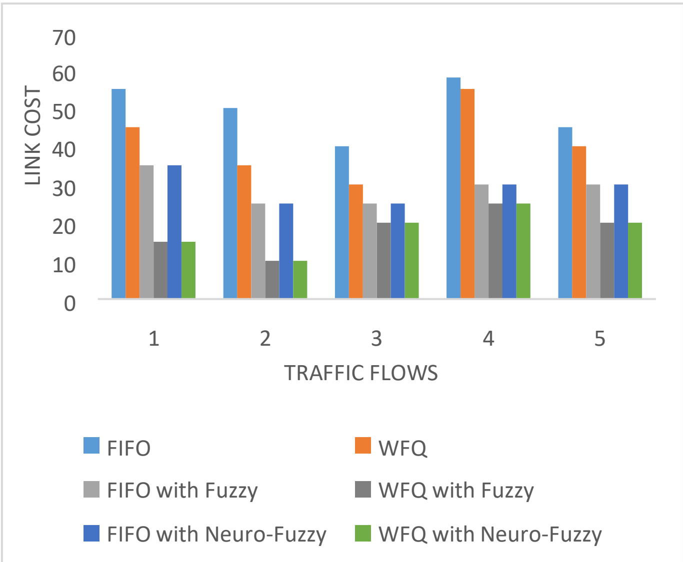


Figure 6-46. Link cost of the schemes (5 MB).

6.6.6.6 Impact on packet loss ratio.

The transmission of packets from source to destination usually can be affected by the amount of packet delivery fraction as well as congestion in the network. It is shown in Figure 6-47 that WFQ increases in packet loss ratio rapidly while a low ratio of packet loss appears for both WFQ with fuzzy and WFQ with neuro-fuzzy. This could arise with the impact of delay on the bandwidth requirements, which form bases for the merit of WFQ over other conventional scheduling schemes.

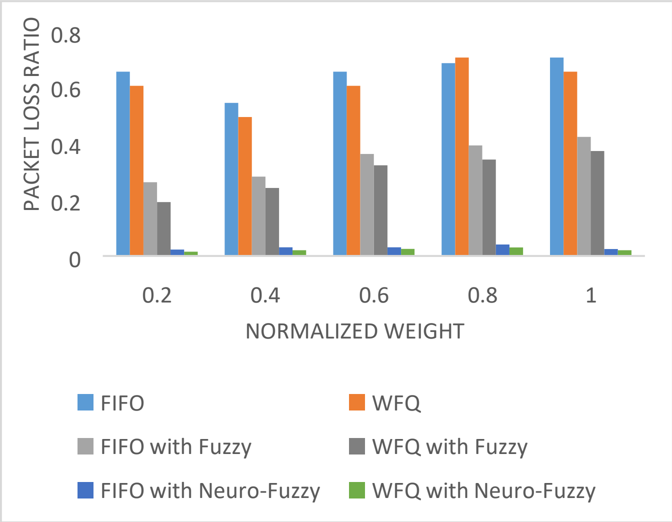


Figure 6-47. Packet loss ratio of the schemes (5 MB).

The summary of the performance analysis of all FIFO and WFQ schemes in terms of service utilization with their standard error of the mean (SEM) as well as the confidence of interval (95% CI) is highlighted in Table 20 and Table 21 respectively.

Table 20. Performance Analysis of FIFO using FTP in MPLS

Scenarios	Algorithms	Service Utilization SEM	95% CI
Scenario 3	No Fuzzy (2 MB)	0.05148	0.32 ± 0.101
	Fuzzy (2 MB)	0.03736	0.486 ± 0.0732
Scenario 4	No Fuzzy (5 MB)	0.05148	0.32 ± 0.101
	Fuzzy (5 MB)	0.03370	0.706 ± 0.0661
Scenario 5	No Neuro-Fuzzy (2 MB)	0.05148	0.32 ± 0.101
	Neuro-Fuzzy (2 MB)	0.02180	0.626 ± 0.0428
Scenario 6	No Neuro-Fuzzy (5 MB)	0.05148	0.32 ± 0.101
	Neuro-Fuzzy (5 MB)	0.03980	0.762 ± 0.0780

Table 21. Performance Analysis of WFQ using FTP in MPLS

Scenarios	Algorithms	Service Utilization SEM	95% CI
Scenario 3	No Fuzzy (2 MB)	0.03741	0.38 ± 0.0733
	Fuzzy (2 MB)	0.01480	0.59 ± 0.0291
Scenario 4	No Fuzzy (5 MB)	0.03741	0.38 ± 0.0733
	Fuzzy (5 MB)	0.02736	0.706 ± 0.0661
Scenario 5	No Neuro-Fuzzy (2 MB)	0.03741	0.38 ± 0.0733
	Neuro-Fuzzy (2 MB)	0.01800	0.678 ± 0.0157
Scenario 6	No Neuro-Fuzzy (5 MB)	0.03741	0.38 ± 0.0733
	Neuro-Fuzzy (5 MB)	0.02204	0.814 ± 0.0432

6.7 Concluding Remarks

This chapter explains concluding remarks on the proposed scenarios. In Scenario 1, the results of the proposed Bandwidth Management using Multimedia Services in MPLS network were presented where traffic engineering mechanism is implemented, which were compared in terms of their performance at acceptable QoS levels while achieving significant delay and delay variation. In scenario 2, the results of the FIFO and WFQ were presented using real-time traffic and results have shown that both schemes in MPLS compared to the schemes in IP perform better in terms average throughput and average end-to-end delay. Then, the results were presented using non-real time traffic with the two schemes in which WFQ in MPLS performs better than WFQ in IP.

Further improvement on performance is observed as the proposed Fuzzy based scheduling scheme (2 MB) in scenario 3 was introduced using non-real-time traffic with the two schemes of FIFO and WFQ provided in scenario 2 and the results show that the combination has improved in terms of performance metrics. Also, the proposed Neuro-Fuzzy based Packet Scheduling Algorithm (2 MB) in scenario 4 outperforms the one that was proposed in scenario 3 using same scheme, which was chosen for its best performance. There are three schemes being compared for their performance evaluations; i) baseline FIFO scheme which has been used in scenarios 2 to 6, ii) proposed Fuzzy-based packet scheduling iii) proposed Neuro-Fuzzy packet scheduling. The proposed Neuro-Fuzzy packet scheduling with FIFO and WFQ schemes in scenario 5 outperformed other schemes in terms, average throughput, delay, cost, reliability, packet loss ratio, and service utilization. Finally, in scenario 6, the results for the Proposed Neuro-Fuzzy based Packet Scheduling Algorithms (5 MB) were presented. The proposed Neuro-Fuzzy scheme produces better SEM than the conventional scheme and the Fuzzy scheme showing low delay, low link cost, moderate reliability, low packet loss ratio, and increased rate of packet delivery within the MPLS. This can be verified in Table 20 and Table 21, which depict the SEM performance analysis of FIFO and WFQ when both fuzzy and neuro-fuzzy were introduced.

7 Conclusion and Future Work

7.1 Conclusion

This thesis proposed an intelligent based packet scheduling scheme to improve on the conventional FIFO and WFQ performance by continuously altering important simulation parameters. The IP/MPLS was considered as the technology for the application of the proposed approach due to its flexibility in traffic engineering. Fuzzy logic and neuro-fuzzy controllers are used for decision making. Both require a set of rules to be used for the selection of LSP for better services. The essential features of MPLS to the future of networking is to have one automatic network control structure, which is the dream of every carrier. The ability to make this dream come true has appeared in the form of a new approach to routing set that comprises the framework of an intelligent based packet scheduling algorithms using MPLS technology for future generation. This allows for automatic provisioning, load balancing, provisioned bandwidth service, and bandwidth on demand.

The main objective of this research study is to design an intelligent based packet scheduling framework for IP/MPLS system which aims at minimizing end-to-end delay and at the same time having moderate bandwidth utilization while maintaining QoS. The proposed framework consists of various functionalities which are summarised below.

7.1.1 Summarised key Contributions and Simulation Results Summary

7.1.1.1 Design and Investigation of Bandwidth Management technique using MPLS technology

- Bandwidth Management using baseline IP/MPLS Model for Mobile Wireless Network.
- Proposed Resource Reservation Protocol Tunnelling Extension in MPLS for sustainable Mobile Wireless Networks.
- Proposed Approach to Label Distribution Protocol Signalling and LSP in IP/MPLS using Multimedia Services for Bandwidth Allocation.

7.1.1.2 Fuzzy Logic Algorithm

- Modeling a fuzzy-based packet scheduling for 5G system service-aware has been proposed. It will eliminate the shortcomings of conventional algorithms using input parameters for deciding service-aware scheme. The major problem with this algorithm is its complexity as a result of increase in the number of input parameters.

7.1.1.3 Neuro-Fuzzy Algorithm

- The proposed Neuro-Fuzzy based packet scheduling algorithm is a combination of two algorithms. In order to improve on the fuzzy-based scheme and eliminate its limitations by employing training of neuro-fuzzy system using a combination of fuzzy rules, there is need to propose new model of neuro-fuzzy scheme. This will provide fast forwarding mechanism at the core while sustaining bandwidth utilization and insuring minimal packet dropping rate.

7.2 Future Work

The work presented in this research work provides the foundation, which does leave room for future improvements. There is a need to consider further simulation scenarios to be similar to real-life scenarios. Some of these potential research directions are discussed in the following subsections.

7.2.1 Improving the Fuzzy based Packet Scheduling Scheme

In the proposed fuzzy-based scheme, only the FIFO and WFQ are considered. This can be improved by the extension to other conventional schemes in the IP/MPLS networks. There is needing to further use real-time traffic for the system, but the trade-off between delay and bandwidth in terms of packet loss ratio and reliability of the system should be taken into consideration.

7.2.2 Improving Neuro-Fuzzy based Packet Scheduling Scheme.

In this thesis, the introduction of neuro-fuzzy to FIFO and WFQ was presented and considered. This can be improved by incorporating more input parameters for optimal performance. Furthermore, the scheme can also be improved with the incorporation of neuro-fuzzy into baseline scheme using real-time traffic.

7.2.3 SD-WAN Technologies

It can be seen in this thesis that only the fixed network technology was considered based on IP/MPLS technology. The MPLS technology in the proposed scheme is fixed, but this can be improved for the future generation of networks by evolving into more of network virtualization such as SDN-WAN. The proposed framework could be improved to incorporate some of other technologies, which will be possible by using the software-defined nature of the WAN. This will enable to have centralized controller where there will be separation of forwarding control from data plane. Incorporating 5G with MPLS will improve the transmission rate, which will as well improve the traffic efficiency of the proposed fuzzy and neuro-fuzzy based packet scheduling schemes.

7.2.4 Traffic Efficiency in the 5G

In the proposed framework, only fairness of bandwidth allocation was considered as the means of bandwidth distribution and saving bandwidth in the core have not been considered. This can be improved by devising schemes for saving bandwidth by using a mixture of technologies and considering the medium of communication and proposing techniques that can improve traffic efficiency and reduce delay in end-to-end traffic engineering.

REFERENCES

[1] A. Alia and K. Ghassan, "Analysis of MPLS and IP Networks Performance to Improve the Qos using Opnet Simulator" *Journal of Emerging Trends in Computing and Information Sciences,* vol. 8, pp. 1-9, 2017.

[2] O. M. Heckmann, *The competitive Internet service provider: network architecture, interconnection, traffic engineering and network design*: John Wiley & Sons, 2007.

[3] Q. Zheng and G. Mohan, "Protection approaches for dynamic traffic in IP/MPLS-over-WDM networks," *IEEE Communications Magazine,* vol. 41, pp. S24-S29, 2003.

[4] O. Akinsipe, F. Goodarzi, and M. Li, "Comparison of IP, MPLS and MPLS RSVP-TE Networks using OPNET," *International Journal of Computer Applications,* vol. 58, 2012.

[5] R. S. Naoum and M. Maswady, "Performance Evaluation for VOIP over IP and MPLS," *World of Computer Science and Information Technology Journal (WCSIT),* vol. 2, pp. 110-114, 2012.

[6] GeeksforGeeks. *Computer Network | Packet Switching and Delays*. Available: https://www.geeksforgeeks.org/computer-network-packet-switching-delays/ [Accessed: Jul 2018]

[7] Trendzbee. *Cisco VNI Global Internet Video*. Available: http://trendzbee.com/ip-video-traffic-boom-2016-2021/ [Accessed: Oct 2018]

[8] Cisco. *Cisco Visual Networking*. Available: https://www.digitaltveurope.com/2016/06/08/cisco-video-to-account-for-79-of-web-traffic-by-2020/ [Accessed: Oct 2018]

[9] Statista. *Global IP data traffic from 2016 to 2021*. Available: https://www.statista.com/statistics/499431/global-ip-data-traffic-forecast/ [Accessed: Oct 2018]

[10] Cisco. *Cisco Wide Area Application Services Configuration Guide*. Available: https://www.cisco.com/c/en/us/td/docs/app_ntwk_services/waas/waas/v623/configuration/guide/cnfg/monitor.html [Accessed: Dec 2018]

[11] PacketIQ. *Intelligent Network & Application Performance Management*. Available: http://www.packetiq.com/tools/PacketIQ-Bandwidth-Statistical-Analyzer.aspx [Accessed: Dec 2018]

[12] Webnms. *Bandwidth Utilisation Report*. Available: https://www.webnms.com/telecom/examples/PerformanceManagement/reports/bandwidth_utilization/README.html [Accessed: Oct 2018]

[13] L. Chitanana, "Bandwidth management in universities in Zimbabwe: Towards a responsible user base through effective policy implementation," *International Journal of Education and Development using Information and Communication Technology,* vol. 8, p. 62, 2012.

[14] M. Kassim, M. Ismail, K. Jumari, and M. I. Yusof, "A survey: bandwidth management in an IP based network," *World Academy of Science, Engineering and Technology,* vol. 62, pp. 356-363, 2012.

[15] J. Korhonen, *Introduction to 4G mobile communications*. Boston: Artech House, 2014.
[16] F. Holness and C. Phillips, "Congestion control mechanism for traffic engineering within MPLS networks," in *Next Generation Networks. Networks and Services for the Information Society*, ed: Springer, 2000, pp. 254-263.
[17] I. G. Ghetie, *Fixed-mobile wireless networks convergence: technologies, solutions, services*: Cambridge University Press New York, 2008.
[18] V. Cisco, "The zettabyte era: trends and analysis. Updated (07/06/2017)," ed: ed, 2017.
[19] Grandviewresearch. (2015). *Multi Protocol Labelled Switching (MPLS) Internet Protocol (IP) Virtual private network (VPN) Services Market Analysis* Available: https://www.grandviewresearch.com/industry-analysis/multi-protocol-labelled-switching-internet-protocol-virtual-private-network-market [Accessed: Nov 2018]
[20] M. S. M. N. VIJAYARANI, "Wired vs Wireless Using Advanced Network," *International Journal of Engineering Research and General Science,* vol. 3, 2015.
[21] K. C. Laudon and J. P. Laudon, *Management information system*: Pearson Education India, 2016.
[22] Wikipedia. *Network Topology*. Available: https://simple.wikipedia.org/wiki/Network_topology [Accessed: Nov 2018]
[23] Cisco. (2017). *Wi-Fi Network*. Available: https://www.cisco.com/c/en/us/solutions/small-business/resource-center/networking/wireless-network.html [Accessed: Nov 2018]
[24] Forwardslashtechnology. *Wireless Networking Services for Businesses*. Available: http://forwardslashtechnology.com/wireless-networking.html [Accessed: Nov 2018]
[25] Homelanxtreme. *HTML Site Map*. Available: http://www.homelanxtreme.com/wired-vs-wireless.htm [Accessed: Oct 2018]
[26] W.-F. Alliance.
[27] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wübben, J. Bartelt*, et al.*, "Towards a flexible functional split for cloud-RAN networks," in *2014 European Conference on Networks and Communications (EuCNC)*, 2014, pp. 1-5.
[28] A. H. Jafari, D. López-Pérez, H. Song, H. Claussen, L. Ho, and J. Zhang, "Small cell backhaul: challenges and prospective solutions," *EURASIP Journal on Wireless Communications and Networking,* vol. 2015, p. 206, 2015.
[29] D. K. Hunter, A. Mcguire, and G. Parsons, "Mobile backhaul for small cells [Guest Editorial]," *IEEE Communications Magazine,* vol. 51, pp. 60-61, 2013.
[30] J. Hansryd and J. Edstam, "Microwave capacity evolution," *Ericsson review,* vol. 1, pp. 22-27, 2011.
[31] M. Coldrey, H. Koorapaty, J.-E. Berg, Z. Ghebretensae, J. Hansryd, A. Derneryd*, et al.*, "Small-cell wireless backhauling: A non-line-of-sight

approach for point-to-point microwave links," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, 2012, pp. 1-5.
[32] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang*, et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE access,* vol. 1, pp. 335-349, 2013.
[33] Wikipedia. *IPv4*. Available: https://en.wikipedia.org/wiki/IPv4 [Accessed: Oct 2018]
[34] Wikipedia. *IPv6*. Available: https://en.wikipedia.org/wiki/IPv6_address [Accessed: Oct 2018]
[35] J. Postel, "RFC 791: Internet protocol," 1981.
[36] J. Postel, "User datagram protocol," 2070-1721, 1980.
[37] B. A. Forouzan, *TCP/IP protocol suite*: McGraw-Hill, Inc., 2002.
[38] C. Petrie and O. Spatscheck, "Future Internet Protocols," *IEEE Internet Computing,* vol. 16, pp. 11-13, 2012.
[39] D. Saucez, L. Iannone, O. Bonaventure, and D. Farinacci, "Designing a deployable internet: The locator/identifier separation protocol," *IEEE Internet Computing,* pp. 14-21, 2012.
[40] Towards converged 5G mobile networks: Challenges and current Available: https://itunews.itu.int/En/5228 [Accessed 2018]
[41] *Ericsson outlines trends and changes from mwc 2015* Available: http://www.rappler.com/technology/news/87626 [Accessed 2018]
[42] *Forum preparing for 5g in the real-world* Available: http://www.rcrwireless.com/20150504/opinion/reader [Accessed 2018]
[43] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution towards 5G multi-tier cellular wireless networks: An interference management perspective," *arXiv preprint arXiv:1401.5530,* 2014.
[44] A. Agarwal, G. Misra, and K. Agarwal, "The 5th generation mobile wireless networks-key concepts, network architecture and challenges," *American Journal of Electrical and Electronic Engineering,* vol. 3, pp. 22-28, 2015.
[45] (2013). *5G Infrastructure Public-Private Partnership*. Available: http://5g-ppp.eu/ [Accessed 2018]
[46] M. Fallgren and B. Timus, "Scenarios, requirements and KPIs for 5G mobile and wireless system," *METIS deliverable D,* vol. 1, p. 1, 2013.
[47] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia*, et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine,* vol. 52, pp. 26-35, 2014.
[48] F. De Rango, A. Malfitano, and S. Marano, "Cross-Layer Architecture: The WiMAX Point of View," *Quality of Service Architectures for Wireless Networks: Performance Metrics and Management: Performance Metrics and Management,* p. 57, 2010.
[49] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE communications magazine,* vol. 52, pp. 36-43, 2014.
[50] Cisco. *Multiprotocol Label Switching*. Available: https://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/5_x/nx-os/mpls/configuration/guide/mpls_cg/mp_mpls_overview.pdf [Accessed: Sept 2018]

[51] P. Knight and C. Lewis, "Layer 2 and 3 virtual private networks: taxonomy, technology, and standardization efforts," *IEEE Communications Magazine,* vol. 42, pp. 124-131, 2004.
[52] G. R. Ash, *Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks*: Morgan Kaufmann, 2006.
[53] M. Khan, "MPLS Traffic Engineering in ISP Network," *International Journal of Computer Applications,* vol. 59, 2012.
[54] Cisco. *Visual Networking Index*. Available: https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/white-paper-listing.html [Accessed Sept 2017]
[55] A. V. Lemeshko, A. Hailan, and Z. A. Sabeeh, "Mathematical model of MPLS-network supporting traffic engineering DiffServ," 2011.
[56] A. Viswanathan, R. Callon, and E. C. Rosen, "Multiprotocol label switching architecture," 2001.
[57] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao, "Overview and principles of Internet traffic engineering," 2070-1721, 2002.
[58] S. Fahmy and R. Jain, "1 Resource ReSerVation Protocol (RSVP)," *the Handbook. I trust you will find the Handbook to be an invaluable reference.,* 2000.
[59] K. Nichols, V. Jacobson, and L. Zhang, "A two-bit differentiated services architecture for the Internet," RFC 2638, July 2070-1721, 1999.
[60] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "RFC 2205-Resource ReSerVation Protocol (RSVP)–version 1 functional specification," Standards Track RFC1997.
[61] Huawei. *VPLS*. Available: http://support.huawei.com/enterprise/docinforeader!loadDocument1.action?contentId=DOC1000009655&partNo=100102 [Accessed: Oct 2018]
[62] Exponential-e. *MPLS vs VPLS*. Available: https://www.exponential-e.com/about-us/why-exponential-e/mpls-vs-vpls [Accessed Sept 2017]
[63] Shopforbandwidth. *MPLS | Multiprotocol Label Switching*. Available: https://shopforbandwidth.com/mpls/ [Accessed: Sept 2017]
[64] W. Henderickx. (2018). *EnterpriseAI: Architecting Your Network for 5G*. Available: https://www.enterpriseai.news/2018/07/16/architecting-your-network-for-5g/ [Accessed 2019]
[65] C. System. (2002). *Implementing MPLS Traffic Engineering* Available: https://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k_r5-3/mpls/configuration/guide/b-mpls-cg53x-asr9k/b-mpls-cg53x-asr9k_chapter_0110.html [Accessed: Oct 2018]
[66] A. Alidadi, M. Mahdavi, and M. Hashmi, "A new low-complexity QoS routing algorithm for MPLS traffic engineering," in *Communications (MICC), 2009 IEEE 9th Malaysia International Conference on*, 2009, pp. 205-210.
[67] T. M. Chen, "Network traffic modeling," *The handbook of computer networks,* vol. 3, 2007.
[68] A. Kumar, V. Kumar, R. Kumar, S. Srivastava, and R. K. Singh, "Performance evaluation of IP network and MPLS network using NS2 simulator," in *Futuristic Trends on Computational Analysis and*

Knowledge Management (ABLAZE), 2015 International Conference on, 2015, pp. 369-375.
[69] F. Le Faucheur, "Maximum allocation bandwidth constraints model for DiffServ-aware MPLS traffic engineering," 2005.
[70] Posterus. *Bandwidth allocation methods in MPLS-TE networks*. Available: http://www.posterus.sk/?p=16328&output=pdf [Accessed: July 2017]
[71] J. Ash, "Max Allocation with Reservation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering & Performance Comparisons," 2005.
[72] F. Le Faucheur, "Russian dolls bandwidth constraints model for DiffServ-aware MPLS traffic engineering," 2005.
[73] E. D. Osborne and A. Simha, *Traffic engineering with MPLS*: Cisco Press, 2002.
[74] Informit. *Network Management and MPLS*. Available: http://www.informit.com/articles/article.aspx?p=101751 [Accessed: June 2018]
[75] S. G. TECH. *Cisco three-layer hierarchical model*. Available: https://smarttechrom.blogspot.com/2016/02/cisco-three-layer-hierarchical-model.html [Accessed: Nov 2018]
[76] Slideshare. *Introduction to Mobile Core Network*. Available: https://www.slideshare.net/yusufd/introduction-to-mobile-core-network-17667704 [Accessed: Nov 2018]
[77] NetworkWorld. *MPLS*. Available: https://www.networkworld.com/article/2297171/sd-wan/network-security-mpls-explained.html [Accessed: June 2018]
[78] A. Research. *MPLS-TP based Packet Transport Solution* Available: http://acgcc.com/wp-content/uploads/2014/08/ECI-TELECOMS-MPLS-TP-NPT-SOLUTION.pdf [Accessed: Oct 2018]
[79] C. W. Braun and D. W. Petr, "Performance evaluation of dynamic and static bandwidth management methods for ATM networks," 1994.
[80] S. Chong, S.-q. Li, and J. Ghosh, "Predictive dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM," *IEEE Journal on Selected Areas in Communications,* vol. 13, pp. 12-23, 1995.
[81] T. Oh, K. Lee, and S. Cho, "An analytical performance model for co-management of last-level cache and bandwidth sharing," in *Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on*, 2011, pp. 150-158.
[82] R. Scopigno and A. Vesco, "A distributed bandwidth management scheme for multi-hop wireless access networks," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, 2011, pp. 534-539.
[83] S. Borst, S. Hanly, and P. Whiting, "Optimal resource allocation in HetNets," in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 5437-5441.
[84] G. Ou, "Managing Broadband Networks: A Policymaker’s Guide," *ITIF, December,* 2008.

[85] C. Gallon and O. Schelén, "Bandwidth management in next generation packet networks," *MSF, August,* 2005.
[86] M. P. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms," *Communications Magazine, IEEE,* vol. 42, pp. S8-15, 2004.
[87] Amazonaws. *Bandwidth Management Technique for Improving Virtual Class in Rural Area Network*. Available: https://s3.amazonaws.com/academia.edu.documents/32411186/Bandwidth-Management-Technique-for-Improving-Virtual-Class-in-Rural-Area-Network.pdf [Accessed: Jun 2017]
[88] S. Machiraju, M. Seshadri, and I. Stoica, "A scalable and robust solution for bandwidth allocation," in *Quality of Service, 2002. Tenth IEEE International Workshop on*, 2002, pp. 148-157.
[89] P. K. Lakshmanan and Q.-A. Zeng, "Dynamic adaptive resource allocation scheme for multimedia services in wireless and mobile networks," pp. 1-7.
[90] P. Van Mieghem and F. A. Kuipers, "Concepts of exact QoS routing algorithms," *IEEE/ACM Transactions on networking,* vol. 12, pp. 851-864, 2004.
[91] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: extensions to RSVP for LSP tunnels," 2070-1721, 2001.
[92] K. Kar, M. Kodialam, and T. V. Lakshman, "Minimum interference routing of bandwidth guaranteed tunnels with MPLS traffic engineering applications," *IEEE Journal on selected areas in communications,* vol. 18, pp. 2566-2579, 2000.
[93] M. Kodialam and T. Lakshman, "Minimum interference routing with applications to MPLS traffic engineering," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2000, pp. 884-893.
[94] A. Kotti, R. Hamza, and K. Bouleimen, "Bandwidth constrained routing algorithm for MPLS traffic engineering," in *Networking and Services, 2007. ICNS. Third International Conference on*, 2007, pp. 20-20.
[95] K. Gopalan, T.-c. Chiueh, and Y.-J. Lin, "Load balancing routing with bandwidth-delay guarantees," *IEEE Communications Magazine,* vol. 42, pp. 108-113, 2004.
[96] Y. Yang, L. Zhang, J. K. Muppala, and S. T. Chanson, "Bandwidth–Delay constrained routing algorithms," *Computer networks,* vol. 42, pp. 503-520, 2003.
[97] S. Kulkarni, R. Sharma, and I. Mishra, "New QoS routing algorithm for MPLS networks using delay and bandwidth constraints," *Int. J. Inf,* vol. 2, pp. 285-293, 2012.
[98] M. N. Soorki and H. Rostami, "Label switched protocol routing with guaranteed bandwidth and end to end path delay in MPLS networks," *Journal of Network and Computer Applications,* vol. 42, pp. 21-38, 2014.
[99] R. Guedrez, O. Dugeon, S. Lahoud, and G. Texier, "Label encoding algorithm for MPLS segment routing," in *2016 IEEE 15th International*

Symposium on Network Computing and Applications (NCA), 2016, pp. 113-117.
[100] T. Radivilova, L. Kirichenko, and O. Yeremenko, "Calculation of routing value in MPLS network according to traffic fractal properties," in *2017 2nd International Conference on Advanced Information and Communication Technologies (AICT)*, 2017, pp. 250-253.
[101] N. Wim Henderickx. (2018). *Architecting Your Network for 5G*. Available: https://www.enterpriseai.news/2018/07/16/architecting-your-network-for-5g/ [Accessed August, 2018]
[102] M. Masood and M. M. Fouad, "Analysis of artificial intelligence-based metaheuristic algorithm for MPLS network optimization," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, 2018, pp. 1-4.
[103] J. Ghetie, *Fixed Mobile Wireless Networks Convergence*: Cambridge University Press, 2008.
[104] I. P. Kun, "QoS In Packet Networks," ed: Springer, 2005.
[105] A. Mishra and P. Venkitasubramaniam, "Anonymity and fairness in packet scheduling: a quantitative tradeoff," *IEEE/ACM Transactions on Networking,* vol. 24, pp. 688-702, 2016.
[106] J. Gozdecki, A. Jajszczyk, and R. Stankiewicz, "Quality of service terminology in IP networks," *Communications Magazine, IEEE,* vol. 41, pp. 153-159, 2003.
[107] G. Fodor, A. Eriksson, and A. Tuoriniemi, "Providing quality of service in always best connected networks," *Communications Magazine, IEEE,* vol. 41, pp. 154-163, 2003.
[108] T. T. SasanAdibi, S. Parekh, and R. Jain, "Quality of Service Architecture for Wireless Networks," *Information Science Reference, Hershey New York, USA,* 2010.
[109] J. W. Evans and C. Filsfils, *Deploying IP and MPLS QoS for multiservice networks: Theory and practice*: Elsevier, 2010.
[110] Y. Chen, T. Farley, and N. Ye, "QoS requirements of network applications on the Internet," *Information Knowledge Systems Management,* vol. 4, pp. 55-76, 2004.
[111] Cisco. (Oct, 2017). *Cisco Learning Network*. Available: https://learningnetwork.cisco.com/thread/121078 [Accessed: Nov, 2018]
[112] TechTarget.SearchTelecom. *MPLS QoS: Implementing the best model for guaranteed service*. Available: https://searchtelecom.techtarget.com/tip/MPLS-QoS-Implementing-the-best-model-for-guaranteed-service [Accessed: Nov, 2018]
[113] SearchEnterprisewan.TechTarget. *MPLS QoS models*. Available: https://searchenterprisewan.techtarget.com/tip/MPLS-QoS-models [Accessed: Dec, 2018]
[114] S. Zorić and M. Bolić, "Fairness of scheduling algorithms for real-time traffic in DiffServ based networks," in *MELECON 2010-2010 15th IEEE Mediterranean Electrotechnical Conference*, 2010, pp. 1591-1596.
[115] T.-Y. Tsai, Y.-L. Chung, and Z. Tsai, "Introduction to packet scheduling algorithms for communication networks," in *Communications and Networking*, ed: IntechOpen, 2010.

[116] K. J. Xu, "Radio Resource Management for Satellite UMTS. Dynamic scheduling algorithm for a UMTS-compatible satellite network," University of Bradford, 2013.

[117] Z. Patel and U. Dalal, "Implementation and evaluation of dynamically weighted low complexity fair queuing (DWLC-FQ) algorithm for packet scheduling in WiMAX networks," *China Communications,* vol. 13, pp. 128-140, 2016.

[118] S. M. Musa, M. Tembely, M. N. Sadiku, and P. H. Obiomon, "Modeling and Simulation of Queuing Scheduling Disciplines on Packet Delivery for Next Generation Internet Streaming Applications," 2014.

[119] S. A. Sharafali, M. M. Al-Quzwini, and R. S. Fyath, "Performance Evaluation of MPLS TE Signal Protocols for Voice Applications with QoS Implementation," *International Journal of Networks and Communications,* vol. 5, pp. 1-9, 2015.

[120] J. C. Bennett and H. Zhang, "Hierarchical packet fair queueing algorithms," *IEEE/ACM Transactions on Networking (TON),* vol. 5, pp. 675-689, 1997.

[121] C. Chrysostomou, A. Pitsillides, L. Rossides, M. Polycarpou, and A. Sekercioglu, "Congestion control in differentiated services networks using Fuzzy-RED," *Control Engineering Practice,* vol. 11, pp. 1153-1170, 2003.

[122] J. D. Mallapur, S. Abidhusain, S. S. Vastrad, and A. C. Katageri, "Fuzzy based bandwidth management for wireless multimedia networks," in *Information Processing and Management*, ed: Springer, 2010, pp. 81-90.

[123] M.-Y. Chow, S. Altug, and H. J. Trussell, "Heuristic constraints enforcement for training of and knowledge extraction from a fuzzy/neural architecture. I. Foundation," *IEEE Transactions on Fuzzy Systems,* vol. 7, pp. 143-150, 1999.

[124] M. Ali, P. Pillai, and Y. Hu, "Load-aware radio access selection in future generation Satellite-terrestrial wireless networks," *International Journal of Wireless & Mobile Networks,* vol. 4, p. 35, 2012.

[125] S. Mitra and Y. Hayashi, "Neuro-fuzzy rule generation: survey in soft computing framework," *IEEE transactions on neural networks,* vol. 11, pp. 748-768, 2000.

[126] M. Brown, K. Bossley, D. Mills, and C. Harris, "High dimensional neurofuzzy systems: overcoming the curse of dimensionality," in *Proceedings of 1995 IEEE International Conference on Fuzzy Systems.*, 1995, pp. 2139-2146.

[127] H. S. Yazdi, S. E. Hosseini, and M. S. Yazdi, "Neuro-fuzzy based constraint programming," *Applied Mathematical Modelling,* vol. 34, pp. 3547-3559, 2010.

[128] W. T. Kasch, J. R. Ward, and J. Andrusenko, "Wireless network modeling and simulation tools for designers and developers," *IEEE Communications Magazine,* vol. 47, pp. 120-127, 2009.

[129] W. H. Tranter, T. S. Rappaport, K. L. Kosbar, and K. S. Shanmugan, *Principles of communication systems simulation with wireless applications* vol. 1: Prentice Hall New Jersey, 2004.

[130] A. S. Sethi and V. Y. Hnatyshin, *The practical OPNET user guide for computer network simulation*: Chapman and Hall/CRC, 2012.

[131] T. Issariyakul and E. Hossain, "Introduction to Network Simulator 2 (NS2)," in *Introduction to Network Simulator NS2*, ed: Springer, 2009, pp. 1-18.

[132] D. N. Bhargava, D. R. Bhargava, A. Kumawat, and B. Kumar, "Performance of TCP-Throughput on NS2 by Using Different Simulation Parameters," *International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-4 Issue-6 December-2012,* 2012.

APPENDIX B: Delay and Jitter Models

Let Tj represent the delay experienced by the jth packet going through a queue. The difference of transit time between two consecutive packets of the tagged flow can be written as:

$$J_j = T_{j+1} - T_j$$

The average end-to-end delay jitter can be in form of expected absolute value of random variable

$$J = E\left[\left|T_{i+1} - T_i\right|\right]$$

By adopting the approximate formulas for the (J) in three cases, in which the arrival rate stream is small, large and intermediate.

$$J = E\left[\left|T_{i+1} - T_i\right|\right] = \frac{1}{\square}$$ for small arrival rate stream

$$J = E\left[\left|T_{i+1} - T_i\right|\right] = \frac{1}{\square}$$ for large arrival rate stream

where $\square = \square - \square$

λ: the total arrival rate

μ: the service rate

$$J = \frac{1}{\square}\left[1 - e^{(\square-1)\square}\left(1 - \frac{\square}{\square} + e^{\square-1}\right)\right]$$ for intermediate arrival stream

where, Utilisation $\square = \frac{\square}{\square}$

Therefore, average deviation of delays from average delay is given as:

$$\square_D = \sqrt{\frac{1}{1-N}\sum_{i=1}^{N}\left(d - \square\right)_D}$$

While the variance and standard deviation can be used to describe the variability of a random variable. Mean, variance, standard deviation and standard error are given as follows:

Mean, $\bar{x} = \frac{\sum_{i=1}^{N} X_i}{N}$

Variance, $\sigma^2 = \frac{N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}{N^2}$

Standard deviation, $\sigma = \sqrt{\frac{N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}{N^2}}$

The coefficient of variation is defined as the ratio of the standard deviation of the mean value. It is given by:

Coefficient of variation (CV) = $\frac{\sigma}{\bar{x}}$

Standard error = $\frac{\sigma}{\sqrt{N}}$

For large samples N $\geq$30, Confidence interval (CI) = $\bar{x} \pm z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{N}}$

For small samples N < 30, Confidence interval (CI) = $\bar{x} \pm t_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{N}}$

where marginal error = t $\frac{\sigma}{\sqrt{N}}$

90% Confidence intervals for the mean of multimedia service for packet transmitted in MPLS network is given as:

90% CI = point estimate $\pm 1.65 \times SE$

APPENDIX C: MPLS Tables

There are three categories of MPLS information based tables with subfields, which is used to manage information related to LSP namely: Partial Forwarding Table (PFT), Label Information Base (LIB), and Explicit Routing information Base (ERB). Their tables are shown in table 1, table 2, and table 3 as obtained from simulation.

Label Information Base (LIB) Table: This table contains the following five fields:

1. Incoming Label (iLabel)
2. Incoming interface (iIface)
3. Outgoing Label (oLabel)
4. Outgoing interface (oIface)
5. LIBptr (used in the PFT and ERB tables)

Partial Forwarding (PFT) Table: this table contains the following three fields:

1. FEC
2. PHB (per-hop behavior)
3. LIBptr (pointer to the LIB table)

Explicit Routing information Base (ERB) Table: This table must contain the following three fields:

1. FEC
2. LSPID (Label Switched Path Identity)
3. LIBptr (pointer to the LIB table)

--)___PFT dump___[node: 4] (--

FEC	PHB	LIBptr	Alternative Path
5	-1	0	-1
7	-1	1	-1
6	-1	2	-1
9	-1	4	-1
10	-1	5	-1
13	-1	7	-1
14	-1	9	-1

11	-1	11	-1

--)____ERB dump____[node: 4] (--

FEC	LSPid	LIBptr
9	3600	11

____LIB dump____[node: 4]

#	iIface	iLabel	oIface	oLabel	LIBptr
0:	-1	1	5	0	-1
1:	-1	2	7	0	-1
2:	-1	3	5	2	-1
4:	-1	5	7	9	-1
5:	-1	6	5	8	-1
7:	-1	8	5	10	11
9:	-1	10	7	14	-1
11:	-1	-1	7	15	-1

--)____PFT dump____[node: 9] (--

FEC	PHB	LIBptr	Alternative Path
10	-1	1	-1
13	-1	3	-1
6	-1	4	-1
7	-1	5	-1
0	-1	6	-1
5	-1	7	-1
1	-1	8	-1
2	-1	9	-1
3	-1	10	-1

4	-1	11	-1

--)____ERB dump____[node: 9] (--

FEC	LSPid	LIBptr

___LIB dump____[node: 9]

#	iIface	iLabel	oIface	oLabel	LIBptr
1:	-1	2	10	0	-1
3:	-1	-1	10	0	-1
4:	-1	5	10	1	-1
5:	-1	6	8	1	-1
6:	-1	7	8	8	-1
7:	-1	8	10	6	-1
8:	-1	9	8	10	-1
9:	-1	10	8	11	-1
10:	-1	11	8	12	-1
11:	-1	12	8	13	-1

--)____PFT dump____[node: 10] (--

FEC	PHB	LIBptr	Alternative Path
6	-1	0	-1
9	-1	2	-1
14	-1	4	-1
5	-1	5	-1
7	-1	6	-1
0	-1	7	-1
1	-1	8	-1

2	-1	9	-1
3	-1	10	-1
4	-1	11	-1

--)____ERB dump____[node: 10] (--

FEC	LSPid	LIBptr
9	3600	13

____LIB dump____[node: 10]

#	iIface	iLabel	oIface	oLabel	LIBptr
0:	-1	1	6	0	-1
2:	-1	3	9	0	-1
4:	-1	-1	9	0	-1
5:	-1	6	6	1	-1
6:	-1	7	6	2	-1
7:	-1	8	6	6	-1
8:	-1	9	6	7	-1
9:	-1	10	6	10	-1
10:	-1	11	6	12	-1
11:	-1	12	6	14	-1
13:	8	14	9	0	-1