

ENHANCEMENT OF ETHERNET RING PROTECTION SWITCHING PROTOCOL FOR IMPROVING PERFORMANCE OF CARRIER NETWORKS

By
A. S. M. Asadujjaman


MASTER OF SCIENCE
IN
INFORMATION AND COMMUNICATION TECHNOLOGY





INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY


The thesis titled “**ENHANCEMENT OF ETHERNET RING PROTECTION SWITCHING PROTOCOL FOR IMPROVING PERFORMANCE OF CARRIER NETWORKS**” submitted by A. S. M. Asadujjaman, Roll No. 0411312020P, and Session April, 2011, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Information and Communication Technology on May 30, 2017.

BOARD OF EXAMINERS

- 
1. **Dr. Mohammad Shah Alam** **Chairman**
(Supervisor)
Associate Professor
IICT, BUET, Dhaka

- 
2. **Dr. Md. Saiful Islam** **Member**
(Ex-Officio)
Professor and Director
IICT, BUET, Dhaka

- 
3. **Mohammad Imam Hasan Bin Asad** **Member**
Assistant Professor
IICT, BUET, Dhaka

- 
4. **Dr. Salekul Islam** **Member**
(External)
Associate Professor & Head
Dept. of CSE, United International University (UIU), Dhaka

Declaration

It is hereby declared that this thesis and any part of it has not been submitted elsewhere for the award of any degree or diploma.

Signature of the Candidate

A. S. M. Asadujjaman

0411312020P

IICT, BUET

Dedication

THIS THESIS IS DEDICATED
TO
MY PARENTS

Contents

Declaration	ii
Dedication	iii
List of Figures	vi
List of Tables	ix
List of Abbreviations	xii
List of symbols	xiii
Acknowledgment	xiv
Abstract	xvi
1 Introduction	1
1.1 Motivation	2
1.1.1 Channel Adaptability	3
1.1.2 Scalable Ethernet Architecture for Carrier Networks	3
1.2 Objectives and Scope of the Thesis	4
1.3 Thesis Contributions	4
1.4 Organization of Thesis	5
2 An Overview of Carrier Network Transmission	6
2.1 Introduction	6
2.2 Packet Switching	7
2.2.1 Advantages of Packet Switching	7
2.2.2 Carrier Ethernet	7
2.2.3 Ethernet Ring Protection Switching (ERPS)	8

2.2.4	Multi Protocol Label Switching (MPLS)	13
2.2.5	Protection technologies in MPLS	14
2.2.6	Underpinning protocols in Packet Switched Networks	15
2.3	Circuit Switching	21
2.3.1	Synchronous Digital Hierarchy (SDH)	21
2.3.2	Plesiochronous Digital Hierarchy (PDH)	21
2.3.3	Protocol signaling in Carrier Networks	22
2.3.4	Carrier Access Network Architecture	25
2.4	Wireless Transmission and Adaptive Modulation in Carrier Networks	26
2.4.1	Microwave Transmission	26
2.4.2	Adaptive Modulation	26
2.5	Software Defined Networking (SDN)	27
2.6	Issues In Carrier Network Transmission	28
2.6.1	Channel Adaptability	28
2.6.2	TDM: SONET, SDH & PDH	29
2.6.3	IP/MPLS	29
2.6.4	Carrier Ethernet	31
2.6.5	Software Defined Networking (SDN)	32
2.6.6	Research works on Scalability of Ethernet	32
2.7	Summary	33
3	Provisioning Channel Adaptability in Ethernet by Improving ERPS	35
3.1	Introduction	35
3.2	Protection switching optimization	36
3.2.1	ERPS Multi-instance ring design	36
3.2.2	ERPS Control Process Logic Enhancement	38
3.2.3	Calculation of threshold	39
3.2.4	Throughput Analysis	40
3.3	Results & Discussion	42
3.3.1	Effect of instance switching	43
3.3.2	Throughput comparison	43
3.3.3	Packet loss comparison	45
3.3.4	Comparison of Simulation Result with Analytical Model	46

3.3.5	Comparison of throughput between optimized and un-optimized protocol with analytical model	47
3.4	Summary	57
4	FCSEA: A FLOODLESS CARRIER-GRADE SCALABLE ETHERNET ARCHITECTURE	58
4.1	Introduction	58
4.2	The Design of FCSEA	59
4.2.1	Floodless ARP proxy	60
4.2.2	Protection switching consideration	68
4.3	Evaluation	73
4.3.1	Cache hit rate	73
4.3.2	Cache size	73
4.3.3	Response Time	73
4.3.4	Reflective ARP Delay	79
4.4	Summary	81
5	CONCLUSION AND FUTURE WORKS	82
5.1	Conclusion	82
5.2	Future Work	83
	Appendix A	91
	Appendix B	95

List of Figures

1.1	Simplicity of Ethernet configuration	1
2.1	Ethretnet Ring Protection Switching (ERPS) protocol operation	9
2.2	MPLS (Multi Protocol Label Switching) Local Protection	15
2.3	MPLS LSP (Label Switched Path) Protection Switching	15
2.4	DHCP (Dynamic Host Configuration Protocol) Operation	16
2.5	DHCP Discover Message	17
2.6	DHCP Offer Message	18
2.7	DHCP Request Message	19
2.8	DHCP Acknowledgement Message	20
2.9	LTE (Long Term Evolution) Network Architecture	22
2.10	Protocol stack at various LTE network elements	23
2.11	eNodeB to S-GW bearer (S1) Setup Procedure	24
2.12	eNodeB to eNodeB (X2) Interface Setup Procedure	24
2.13	Part of a telecommunication carrier access network	26
2.14	Varying capacity due to weather condition in Adaptive Modulation (AM)	27
2.15	Reference diagram for Software Defined Network (SDN)	28
2.16	Data rates in Synchronous Optical Networking (SONET)	30
2.17	Data rates in Synchronous Digital Hierarchy (SDH)	30
3.1	Instance creation per ring node – before (left) and after (right) pro- tection switching due to signal degradation at link between n1 and n6	38
3.2	Illustration for throughput analysis in a ring topology	40
3.3	Simulation scenario in OMNET++	42
3.4	Effect of instance switching on throughput	43

3.5	Performance comparison (Throughput)	44
3.6	Performance comparison (Packet loss)	45
3.7	Comparison of simulation results with analytical model for proposed modified protocol and existing ITU-T G.8032 protocol	46
3.8	Throughput comparison (degraded link no: 1, RPL link no: 6, full capacity: 1Gbps)	48
3.9	Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 51Mbps)	49
3.10	Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 44Mbps)	50
3.11	Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 37Mbps)	51
3.12	Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 29Mbps)	52
3.13	Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 51Mbps)	53
3.14	Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 44Mbps)	54
3.15	Throughput between (degraded link no: 3, RPL link no: 6, full ca- pacity: 37Mbps)	55
3.16	Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 29Mbps)	56
4.1	An example topology to illustrate the design of FCSEA	59
4.2	Proactive learning modes in FCSEA	60
4.3	Proactive ARP cache entry learning from DHCP	62
4.4	Learning ARP cache table from gratuitous ARP	63
4.5	Connectivity test of a new node (BSC) with a core network NE (Media Gateway) after network integration during PAT	63
4.6	FCSEA switch learning eNodeB (host) from protocol signaling of LTE S1 interface setup	65
4.7	Utilizing LTE signaling procedure to register an LTE eNodeB in the ARP cache of FCSEA	66

4.8	Classification of A2A (Address to Address) Mappings	67
4.9	Internal and external ports and A2A mappings	67
4.10	Prioritization of A2A (Address to Address) Mappings	68
4.11	ERPS ring topology in normal (left) and failure (right)	69
4.12	Traffic flow of Reflective ARP Request by FCSEA.	70
4.13	Reflective ARP Request (RAR) algorithm.	71
4.14	FCSEA ARP request and response processing logic (includes floodless ARP proxy, priority based cache retention and protection switching logic).	72
4.15	Telecommunication carrier network architecture.	72
4.16	Cache hit rate comparison: source hosts = 30, target hosts = 1000 . .	74
4.17	Cache hit rate comparison: source hosts = 30, target hosts = 10000 .	75
4.18	Cache hit rate comparison: source hosts = 500, target hosts = 100 . .	76
4.19	Cache hit rate comparison: source hosts = 500, target hosts = 50000	77
4.20	Cache size evaluation with 50.5K hosts (500 hosts sending ARP re- quest for 50K hosts).	78
4.21	Response time of FCSEA to reply to ARP requests compared to SDN based proxy.	78
4.22	Performance evaluation of Reflective ARP mechanism.	80

List of Tables

3.1	BANDWIDTH DEGRADATION WITH TIME USED IN OMNET++ SIMULATION	44
3.2	SIMULATION PARAMETERS	47
4.1	ARP CACHE OF PROPOSED PROXY U1	67
4.2	ROUND-TRIP DELAYS FOUND IN TELECOM BACKBONE NET- WORK	79

List of Abbreviations

AM	Adaptive Modulation
ARP	Address Resolution Protocol
CDPI	Control-Data-Plane Interface
CHADDR	Client Hardware Address
CIADDR	Client IP Address
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Naming System
ERPS	Ethernet Ring Protection Switching
FRR	Fast Re-route
GIADDR	Gateway IP Address
GUTI	Globally Unique Temporary ID
IGP	Interior Gateway Protocol
LER	Label Edge Router
LSR	Label Switched Router
LTE	Long Term Evolution
MAC	Medium Access Control
MPLS	Multiprotocol Label Switching
MW	Microwave
MME	Mobility Management Entity
NAS	Non-access stratum
NBI	Northbound Interface
PDH	Plesiochronous Digital Hierarchy
PGW	Packet Gateway
PHP	Penultimate Hop Popping
PHP	Penultimate Hop Popping
QoS	Quality of Service

QCI	QoS Class Identifier
RAR	Reflective ARP Request
S-GW	Serving Gateway
SDH	Synchronous Digital Hierarchy
SDN	Software Defined Networking
SHA	Source Hardware Address
SIADDR	Server IP Address
SONET	Synchronous Optical Network
SPA	Source Protocol Address
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VPN	Virtual Private Network
YIADDR	Your IP Address

List of symbols

T	Throughput (in Mbps)
L_i	Traffic (in Mbps) in link number i
C	Capacity of link in Mbps
C_{d1d2}	Capacity of affected link under adaptive modulation
C_{deg}	Traffic (in Mbps) in the affected link
x	Free capacity in aggregation link in the unaffected path

Acknowledgment

I would like to express my profound gratitude and high regards to my supervisor, Dr. Mohammad Shah Alam, whose wisdom, expertise, continuous guidance and generous support has made it possible for me to work on the topic that was of great interest to me. I would like to thank all my teachers from Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology (BUET). The knowledge I have acquired from the classes were of utmost importance for this thesis.

I am indebted to Mr. Mohammad Omar Khyam and Mr. Md. Asikuzzaman from University of New South Wales, Canberra, Australia for finding out time to reply my e-mails and for providing invaluable support during my research. I would like to thank the authority of Banglalink Digital Communications Limited, for providing me permission and support to continue my study and research works. I would specially like to thank Mr. Md. Tofazzal Hossain (Head of Transport Network), Mr. A. F. M. Shakhawat Ullah (Head of Data Network) and Mr. Md. Moinul Islam (Transport Network Deputy Manager) for their support and guidance during the progress of my research.

I like to express my deepest gratitude to my parents for their sacrifices and inspiration during the progress of this work.

Abstract

The ability to provide peer to peer connectivity, fast switching speed, simplicity and cost effectiveness make Ethernet desirable for LTE and future telecommunication carrier networks. With carrier grade switchover capability of Ethernet Ring Protection Switching (ERPS) Ethernet technology is posing as a viable solution in carrier networks. Network operators are replacing SONET/SDH (Synchronous Optical Network/Synchronous Digital Hierarchy) by Ethernet to achieve a more bandwidth efficient and easier-to-maintain network. In this research, work has been done on two areas to improve Ethernet as a carrier transmission technology. Firstly, adaptive modulation (AM) is an evolution in technology which may effectively double the capacity of microwave links in favorable weather condition. Carrier network transmission technologies should be optimized to be able to efficiently respond to channel capacity changes to harness the capability of AM and to address the high bandwidth requirement of next generation networks. To achieve this, an improvement of the ERPS protocol to add channel bandwidth adaptability to Carrier Ethernet has been proposed. Secondly, a single Ethernet network, required for peer to peer connectivity in LTE, poses a severe scalability limitation on access transmission network. This thesis proposes Floodless Carrier-grade Scalable Ethernet Architecture (FCSEA) to meet protection switching and scalability requirements of Ethernet in carrier networks. The proposed scheme improves ERPS protocol to support adaptive modulation in ERPS protected microwave access network for improved throughput, reliability and energy efficiency. It uses a priority based cache retention principle, exploits protocol signaling procedure of cellular network standards and intercepts dynamic configuration (DHCP) messages to ensure an entirely floodless environment for end hosts in a single Ethernet network. We present a mathematical model to verify the effectiveness of our proposed modification. Comparison of simulation result with the developed mathematical model is provided. Simulation results and

mathematical analysis show that the proposed enhancement achieves significantly higher throughput compared to existing ITU-T G.8032 protocol in varying channel conditions. Simulation results and analysis with real data is presented to verify effectiveness of FCSEA and to show its suitability over recent SDN based solutions for scalability.

Chapter 1

Introduction

Ethernet is the most widely used and dominant technology for Local Area Network (LAN). The fast switching speed, economy of scale, simplicity and flexibility [1] of Ethernet make it also desirable in other networks areas such as carrier networks. For instance, Ethernet switching is preferred to routing for high performance of LTE X2-interface [2], [3]. Recently added features such as QoS, scalability and OAM (operation, administration and management) through standardizations such as IEEE 802.1Qay, IEEE802.1Q, IEEE802.1ad IEEE 802.1ag has transformed Ethernet into a carrier-grade technology [4]. With the subsequent development of ITU-T G.8032 recommendation which adds sub-50ms failure recovery capability to Ethernet reusing generic standardized functions [5], Ethernet has finally become a viable replacement for its rival technologies such as SONET/SDH (Synchronous Optical Network/Synchronous Digital Hierarchy) for carrier network transmission. However, a simple Ethernet network is not scalable to serve a large number of nodes as required in telecommunication carrier networks. This is because of broadcast traffic caused by protocols like ARP and DHCP running in an Ethernet network for service and resource discovery.

Site1 (Source)	Site2	Site3	Site4 (Aggregation)
Slot-19 VC4:1:11 CC	Slot-4 VC4:1:11 CC	Slot-18 VC4:1:4 CC	Slot-4 VC4:1:11 CC
Slot-4 VC4:1:11	Slot-18 VC4:1:4	Slot-3 VC4:1:11	Slot-18 VC4:1:11

Service Configuration in SDH (Microwave)

Gateway (Aggregation): 10.52.0.1/28
Site1 IP (Source): 10.52.0.4/28

Service Configuration in Ethernet

Figure 1.1: Simplicity of Ethernet configuration

Currently, traffic protection technologies in carrier Ethernet networks such as Ethernet Ring Protection Switching (ERPS), employ VLANs to keep broadcast within acceptable limit. While this approach requires complex VLAN planning and configuration for every switch, it serves the purpose as a replacement of point to point TDM technologies (e.g. SDH). But, in next generation networks, such as LTE, where nodes require connectivity with their peers, VLAN based segmentation will prohibit nodes under one VLAN to communicate with their neighbors in a different VLAN.

As cellular technology advances beyond 4G, deployment of access network nodes will be denser and more links with higher capacity will be required to connect these nodes. Ease of installation and high capacity of microwave links make them an excellent choice to support this growth. Adaptive Modulation (AM) is an emerging technology in access network to address high capacity requirement of mobile evolution by ensuring reliable transmission during unfavorable channel condition while providing high capacity when channel condition permits. AM has not been considered in the design of ERPS [6]–[9]. As adaptive modulation can cause link capacity to vary by more than 100% depending on channel conditions by changing modulation scheme [10], it needs special consideration to be supported in ERPS to ensure optimum throughput. To address the issue of scalability in Ethernet, Floodless Carrier Grade Scalable Ethernet Architecture (FCSEA) has been proposed in this work. FCSEA can operate in carrier networks to replace complex VLAN based scalability solutions while providing unrestricted peer to peer connectivity. No academic research work has been done to support AM in ERPS although a few proprietary implementations of solution for bandwidth degradation due to AM for ERPS have been reported [6], [11]. These proprietary implementations need special proprietary subsystem for rerouting traffic instances and lack energy efficient instance rerouting capability. Moreover, no previous work has been done on performance analysis of ring topology under link degradation condition of AM to the best of our knowledge.

1.1 Motivation

Transmission technologies used in telecommunication carrier networks are required to evolve to be more intelligent, efficient and scalable to meet the requirement of

next generation mobile wireless network standards.

1.1.1 Channel Adaptability

Transmission networks often deploy microwave (MW) links as a transmission medium. A recent advancement in MW technology is adaptive modulation (AM). When AM is enabled in a microwave link, the link capacity may vary depending on weather condition. In unfavorable weather condition, AM will decrease the MW channel capacity. This will result in packet loss of traffic passing through the affected link. In order to make transmission networks more intelligent and efficient, we introduce the concept of channel adaptability in transmission network. Channel adaptability will enable transmission networks to respond more intelligently to channel capacity changes by rerouting traffic through alternative path instead of causing packet loss.

Transmission networks often deploy microwave (MW) links as a transmission medium. A recent advancement in MW technology is adaptive modulation (AM). When AM is enabled in a microwave link, the link capacity may vary depending on weather condition. In unfavorable weather condition, AM will decrease the MW channel capacity. This will result in packet loss of traffic passing through the affected link. In order to make transmission networks more intelligent and efficient, we introduce the concept of channel adaptability in transmission network. Channel adaptability will enable transmission networks to respond more intelligently to channel capacity changes by rerouting traffic through alternative path instead of causing packet loss.

1.1.2 Scalable Ethernet Architecture for Carrier Networks

Ethernet's fast switching speed, simplicity and cost-effectiveness makes it a viable technology for future carrier transmission network. However, a single Ethernet network is not scalable to a large number of nodes. This is mainly due to broadcast caused in a layer-2 Ethernet domain by underpinning protocols (such as ARP, DHCP). To avoid the limitation of single Ethernet domain, Virtual LAN (VLAN) is deployed in large Ethernet networks. The introduction of VLAN isolates network nodes in groups which can't communicate with each other within the layer-2 network boundaries. This is not desirable for wireless network standards such as LTE

where peer-to-peer connectivity is required (i.e. X2-interface). Moreover, VLANs introduce complexity in the network by requiring planning, deployment and maintenance of a large number of VLANs. Our proposed scalability solution will improve scalability of Ethernet by suppressing broadcast traffic. Thus the fast switching speed of Ethernet can be harnessed to achieve optimum peer-to-peer connectivity and the extreme complexity of VLANs can be avoided.

1.2 Objectives and Scope of the Thesis

The objective of this thesis is to develop a flexible and scalable protocol for telecommunication carrier networks while being compatible with existing Ethernet technology. Compatibility with existing Ethernet technology allows harnessing the simplicity, cost-effectiveness and economies-of-scale of Ethernet while enabling phased implementation. Flexibility is achieved by improvement of Ethernet Ring Protection Switching (ERPS) protocol to provision support for Adaptive Modulation (AM) (channel adaptability) into Ethernet. In order to achieve high scalability, a proxy mechanism equipped with proactive learning is proposed. The specific objectives of this thesis are-

1. To develop a new algorithm for provisioning channel adaptability into Ethernet by improving Ethernet Ring Protection Switching (ERPS) protocol (ITU-T G.8032).
2. To develop a simulation model for the existing ITU-T G.8032 protocol and the proposed new protocol.
3. To improve scalability of Ethernet as required by telecommunication carrier networks.

1.3 Thesis Contributions

In this dissertation we propose an enhancement to ITU-T G.8032 protocol and a multiple instance ERPS ring design principle enabling efficient selection of ERPS instances to switch from an affected link to ensure overall optimal capacity utilization. In order to achieve high scalability, this work uses and evolves EtherProxy

[12], a device for large enterprise networks that uses caching to suppress broadcast traffic for high scalability. The distinct contributions of this research are as below,

1. Channel Adaptability

- Improvement of ERPS protocol for instance rerouting based on channel condition.
- Energy efficient instance rerouting.
- Throughput analysis of a ring network under link bandwidth degradation due to AM changes.

2. Scalable Ethernet Architecture for Carrier Networks

- A proactive learning policy to prevent initial flooding of entire network.
- A prioritized cache retention mechanism.
- Carrier grade protection switching support.
- Simulation and analysis of EtherProxy, SDN based ARP proxy and current work to evaluate broadcast suppression capability and query response time.

1.4 Organization of Thesis

The remainder of this thesis paper is organized as follows. In Chapter 2, an overview of carrier transmission network technologies are discussed. Protocol signaling procedure in cellular network is also introduced here because this is important to understand how cellular network operations will interact with transport network. Cellular access network architecture and advanced wireless network technologies in use are also discussed in this chapter. Finally, this chapter ends with discussion of limitations of each of the transmission technologies discussed beforehand and literature review. In Chapter 3, optimization of Ethernet Ring Protection Switching (ERPS) protocol to achieve channel adaptability in carrier networks by adding support for adaptive modulation (AM) is presented with simulation results and mathematical analysis. Chapter 4 discusses the design of scalability solution using proxy mechanism and evaluation. This dissertation is concluded with future research direction in Chapter 5.

Chapter 2

An Overview of Carrier Network Transmission

2.1 Introduction

Transmission technologies in carrier networks have evolved to meet the requirement of radio technologies starting from 2G to LTE and LTE-Advanced. Carrier network transmission technologies can be classified in two types in terms of the multiplexing technique they use.

- Packet switching networks
- Circuit switching networks

Presently packet switched multiplexing technologies are replacing circuit switched networks in telecommunication carriers to support the ever increasing need for more efficient, flexible and cost effective solutions. Underlying physical infrastructure also plays a vital role on how the protocols must perform to meet the performance requirements. For example, while providing abundant capacity, fiber optic links are susceptible to physical damage leading to network outage if the network is not properly designed. Microwave links on the other hand have an inherent capacity limitation due to spectrum issues while enjoying high reliability against physical failure. Cellular transmission network technologies can use different types of underlying physical media including but not limited to,

- Optical fiber
- Microwave radio
- Copper wire

In this chapter different transmission technologies that are or have been used in transmission networks of telecommunication carriers has been explored.

2.2 Packet Switching

Packet switching is a transmission technology where traffic is composed of fixed or configurable sized chunks called packets. Each packet has a header that identifies the traffic where the packet belongs. Packet switched networks achieve higher channel utilization by sharing the same physical media for different services.

2.2.1 Advantages of Packet Switching

Packet switching is replacing circuit switched networks due to its advantages of robustness, efficiency and cost effectiveness. Advantages of packet switching networks include-

1. Increased channel capacity: different services can share a common physical link in a packet switched network as resources are not required to be dedicated for a particular service. This results in increased efficiency of packet switched networks and as a result larger capacity can be accommodated [13].
2. Cost effectiveness: The high efficiency and mass-production of packet switched technologies have resulted in packet switched network technologies becoming more cost effective than their rival technologies.
3. Robustness: Packet switched networks can be made resilient against network failures by protection switching mechanisms being innovated for these networks. Flexibility of packet switching technology makes it inherently easier to build robust networks.

Presently two types of packet switching technologies are dominant in carrier networks-

- Carrier Ethernet
- Multi-protocol Label Switching (MPLS)

2.2.2 Carrier Ethernet

Ethernet has a long track record of re-inventing itself to adapt to new industry demand. Starting from copper wire networks of 10Mbps capacity Ethernet network

capacity have increased tenfold with every new capacity upgrade. From copper wires Ethernet networks have evolved into optical and wireless networks. Ethernet standards with different capacity are [14]-

- 10 Base-T Ethernet (10 Mbps)
- Fast Ethernet (100 Mbps)
- Gigabit Ethernet (1 Gbps)
- 10 Gigabit Ethernet (10 Gbps)
- 40 Gigabit Ethernet (40 Gbps)
- 100 Gigabit Ethernet (100 Gbps)

Carrier Ethernet is the result of Ethernet re-inventing itself for telecommunication carrier networks. It has been possible by development of several standards such as IEEE 802.1ad [15], IEEE 802.1ah [16], IEEE 802.1ag [17], ITU-T G.8032 [9] etc. to enhance the reliability, scalability and manageability of Ethernet technology to meet the requirement of carriers. Clock signaling protocols such as ITU-T G.8262 [18] and IEEE 1588 [19] have come forward to meet the timing requirement of carrier nodes.

2.2.3 Ethernet Ring Protection Switching (ERPS)

In order to ensure ability of a network to heal itself from a link failure, links are arranged in a ring topology. Ring topology has the advantage of being able to maintain connectivity between all the nodes in the ring despite failure of any single link. In absence of any link failure a ring topology forms a loop which must not be allowed for effective transmission.

Ethernet Ring Protection Switching (ERPS) is a ring protection technology developed by ITU-T under G.8032 recommendation to provide SONET/SDH-like sub-50ms carrier grade protection switching in Ethernet networks. The objective of ERPS is to ensure a loop free topology in idle state and to maintain a connected topology in link failure. This is achieved by forming a logical topology with one link blocked per physical ring. In idle state a link is blocked to prevent loop. This link is

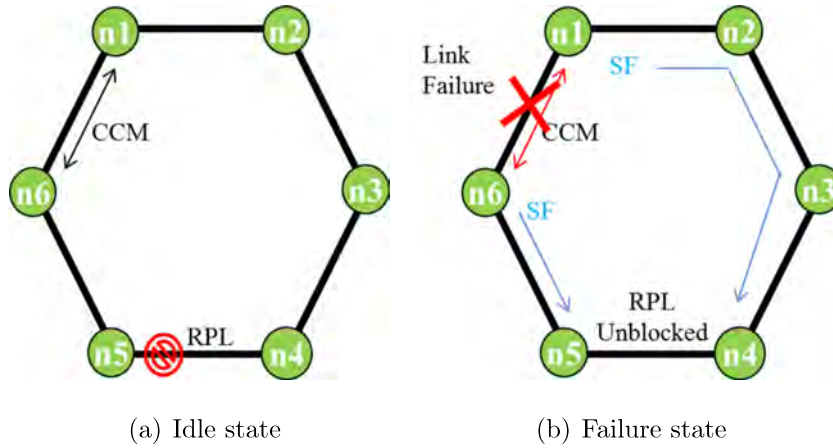


Figure 2.1: Ethernet Ring Protection Switching (ERPS) protocol operation

called ring protection link (RPL). In protection state, on detection of a link failure, failed link is blocked and RPL is unblocked. This ensures a connected topology despite link failure.

2.2.3.1 Architecture

A network protected by ERPS is viewed as a collection of main rings and sub-rings. Every main ring and sub-ring is created by nodes connected by links. A main ring is responsible for ensuring protection of each of its links. A sub-ring, on the other hand, is not responsible for ensuring protection of the common link that it shares with the main ring. Nodes that interconnect main rings with sub-rings are called interconnection nodes. Each node except interconnection nodes has two ring ports. In addition to two ring ports interconnection nodes have ports for connecting sub-rings. A ring is a physical loop of nodes where first port of first node is linked to last port of last node to close the loop. A sub-ring is a single link or a chain of links whose links at both ends connect to interconnection nodes of a main ring. Every main ring and sub-ring has a special link called ring protection link (RPL). The RPL is the ERPS mechanism to prevent formation of loop in the network. Thus an ERPS network is logically seen as a hierarchical arrangement of main rings and sub-rings with every main ring and sub-ring having an RPL.

2.2.3.2 Protection switching operation

The operation of ERPS is illustrated in Figure 2.1. In idle state every link is unblocked except ring protection link (RPL). Every ring port continuously monitors the link it is connected to by using continuity check messages (CCM). When a link failure occurs, the node detecting the failure at one of its ring ports blocks the port and transmits R-APS message on both ring ports. This message informs other ring nodes of the link failure and nodes adjacent to RPL unblocks the blocked ports. Every node including RPL adjacent nodes flush their filtering database (FDB) to re-configure the network for the new logical topology.

2.2.3.3 Ring configuration

Every ring node is configured with a unique node ID. R-APS messages use ring ID as the MAC address. Every ERPS instance has an R-APS channel uniquely identifiable by its VLAN ID (VID), ring ID pair. Thus instances can be made unique by assigning unique VIDs or ring IDs or both. VIDs can be re-used for instances having no common nodes. That is, every physical ring must have different VIDs for different instance on that physical ring but these instances may use same ring ID.

2.2.3.4 Multiple ERPS instances

ERPS operation is performed by ERPS control process and it is possible to create multiple ERPS control processes on a single node each configured with a unique set of service VLANs. Thus multiple logical instances are created on the same physical ring or sub-ring. Consequently, each ERPS instance operates independently and has its own RPL. Each ERPS instance uses a different VID for transmitting R-APS messages.

2.2.3.5 ERPS Version 1 (ITU-T G.8032v1)

Version 1 of ITU-T G.8032/Y.1344 was approved on June 22, 2008 by ITU-T Study Group 15. It defines the automatic protection switching (APS) protocol and protection switching procedure for Ethernet ring topologies at ETH layer. G.8032v1 includes details related to Ethernet ring protection characteristics, architectures and

the ring APS protocol.

Fundamentals of Ethernet ring protection switching architecture established by ERPS version 1 are-

1. loop avoidance principle
2. use of learning, forwarding and address table procedures defined in the Ethernet flow forwarding function (ETH_FF)

This version of ERPS defines following characteristics of ring protection-

1. Methods and conditions of link monitoring
2. Consideration for Ethernet traffic and bandwidth
3. Performance of protection switching

Supported commands/conditions in this version are-

A) Signal Failure (SF)

Signal failure detection on a link results in initiation of protection switching.

B) Wait to restore (WTR)

WTR timer is used to initiate blocking of RPL after SF has cleared. This timer causes delay in RPL blocking to avoid any intermittent failure to cause frequent switching.

C) No request (NR)

No request is transmitted when there is no outstanding failure (i.e. all failure has been recovered).

Following commands are mentioned but not implemented in the version-

- Lockout of protection
- Force switch
- Manual switch
- Replace the RPL
- Exercise signal
- Do not revert

Following topics are discussed in ITU-T G.8032v1 to define ring protection architecture-

- Non-revertive and revertive switching mechanism
- Triggering protection switching
- Conditions for signal failure declaration
- Protection switching models
- Blocking of traffic channel
- Blocking of R-APS (Ring-Automatic Protection Switching) channel
- Flushing filtering database (FDB)

Protection control protocol

Ring protection is based on loop avoidance. It works on the basis at least one port being blocked in a ring.

The protection control protocol allows unblocking of a port in a ring once it is blocked, only if at least one other port is blocked in the ring. The protection control protocol includes-

- Operating principles
- Protection switching behavior
- R-APS message format

This version (G.8032v1) doesn't include recommendation for-

- Blocking of traffic by both ends of RPL
- Detail of multi-ring/ladder network topology

2.2.3.6 ERPS Version 2 (ITU-T G.8032v2)

Multi-ring/ladder topology is studied in this edition.

New concepts in v2:

- ERP instance

- Interconnection node
- Major ring
- R-APS virtual channel
- RPL owner node
- RPL neighbor node
- Sub-ring link
- Wait to block timer (WTB)

2.2.3.7 ERPS Version 3 (ITU-T G.8032v3)

End-to-end service resilience is defined in G.8032v3. Here it is specified that on the active/protection path two and only two additional Ethernet ring nodes should be included that are aware of IUT-T G.8031 linear protection mechanism. It introduces the concept of access sub-ring.

2.2.4 Multi Protocol Label Switching (MPLS)

Multi-protocol Label Switching (MPLS) is a connection-oriented [20] transmission technology where routing is performed based on predefined label switched paths (LSP) and ingress/egress labels on each packet. The label switched path is a set of configuration on the routers along the path that packets travel. It defines the path that a packet belonging to the LSP will take. Label switched paths (LSP) can be configured manually or can be set up by dynamically by protocols such LDP (label distribution protocol). MPLS can encapsulate packets of various protocols such as Ethernet, E1/T1, ATM etc.

2.2.4.1 Router Hierarchy

In MPLS packets are prefixed with an MPLS header which contains one or more labels. In an MPLS network routers are distinguished by a hierarchical classification-

1. Label switching router (LSR)

These are the routers in the middle of the MPLS network. LSRs forward packets entirely based on the MPLS labels and don't look into the IP header.

2. Label edge router (LER)

These routers are located at the edge of MPLS network and label incoming IP packets unless PHP (penultimate hop popping) is used.

3. Provider (P) and Provider Edge (PE) router

In MPLS VPN networks Provider Edge (PE) routers are the ingress/egress routers whereas Provider (P) routers are the core routers that only forward based on labels.

2.2.4.2 Label distribution and routing

A) Label Distribution Protocol (LDP)

Label Distribution Protocol (LDP) distributes MPLS labels in LSRs and LERs.

B) Label-switched paths

Label switched paths (LSP) specify the path for MPLS traffic.

C) Routing

LERs and LSRs route packets in an MPLS network. When a packet arrives –

- The LERs pushes/pops a label and forwards the MPLS packet
- The LSRs forward the MPLS packet based on the top label
- Multiple labels are used in MPLS Virtual Private Network (VPN)

2.2.5 Protection technologies in MPLS

Protection can be in local or global domain in MPLS.

A) Fast Reroute (FRR)

This ensures fast protection of nodes or links [21] by means of locally configured backup tunnels. FRR is a type of protection where backup LSP is pre-calculated [22]. Figure 2.2 the blue tunnel through routers 3, 5 and 4 is configured as the backup tunnel for the vulnerable link between routers 3 and 4. Enough capacity must be provisioned in the backup tunnel to ensure the protection.

B) Label Switched Path (LSP) Protection Switching

LSP protection switching provides end to end protection for a primary tunnel by means of a backup tunnel in the secondary path.

2.2.6 Underpinning protocols in Packet Switched Networks

2.2.6.1 Dynamic Host Configuration Protocol (DHCP)

DHCP (Dynamic Host Configuration Protocol) is a protocol, based on client-server model, to provide Internet parameters to hosts. These parameters include:

- IP address
- Subnet mask
- Default gateway
- DNS (Domain Name System) servers etc.

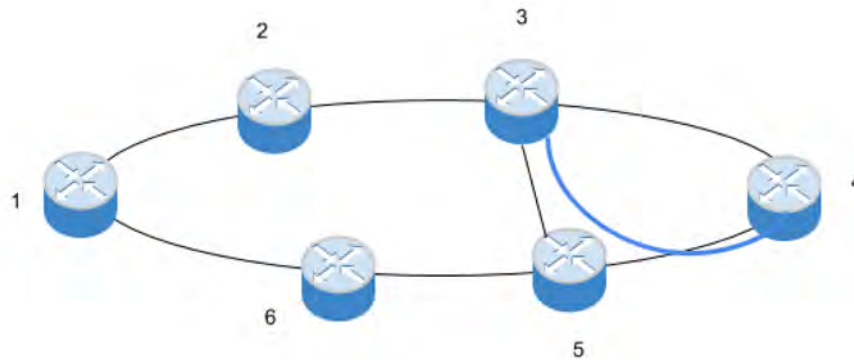


Figure 2.2: MPLS (Multi Protocol Label Switching) Local Protection

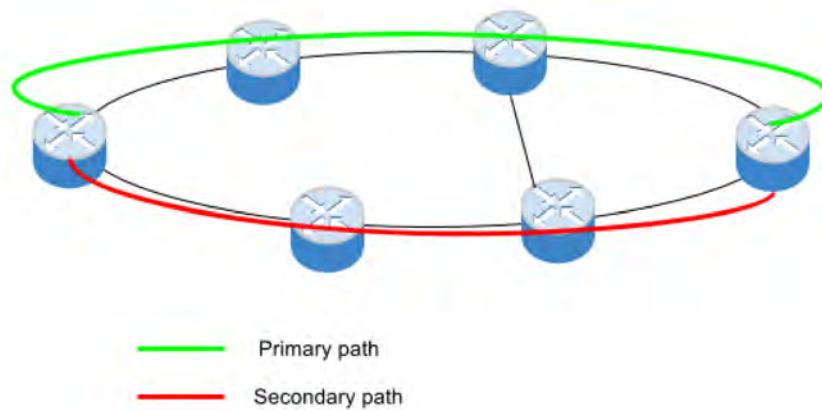


Figure 2.3: MPLS LSP (Label Switched Path) Protection Switching

2.2.6.2 Operation

DHCP operates on a connection-less model between client and server. It uses UDP as the transport layer protocol. UDP port 67 is the server port and UDP port 68 is used by the client.

DHCP operation is divided into four stages (Figure 2.4) all of which use broadcast-

- Service discovery
- IP configuration offer
- Request to server
- Acknowledgement from server

A) Service Discovery (DHCP Discover)

The hosts (DHCP client) broadcasts packets on the L2 network using the destination address 255.255.255.255.

An example request is shown in Figure 2.5.

B) IP Configuration Offer (DHCP Offer)

After receiving the discover packet server allocates an IP address for the host and sends it in the DHCP offer packet. An example packet is given in Figure 2.6.

C) Request to Server (DHCP Request)

This is the hosts response to the DHCP offer message from the server. Figure 2.7 shows an example packet.

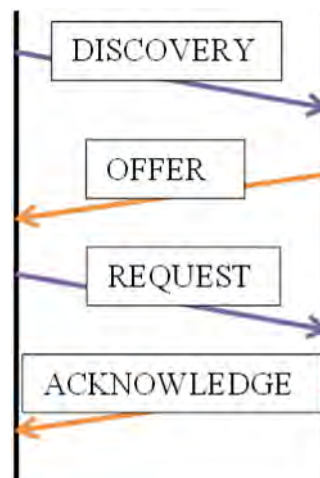


Figure 2.4: DHCP (Dynamic Host Configuration Protocol) Operation

D) Acknowledgement from Server (DHCP Acknowledgement)

This is the final stage of dynamic IP configuration process and the host is expected to configure its interface with the negotiated parameters. Figure 2.8 depicts an example of DHCP acknowledgement packet.

Due to its versatility, DHCP can be modified for use in distribution of other types of identifiers than IP address such as GUTI (Globally Unique Temporary ID) in LTE [23].

2.2.6.3 Address Resolution Protocol (ARP)

The Address Resolution Protocol (ARP) is a protocol used to map logical address (IP) to physical address (MAC). This is a mandatory protocol as physical addresses are required for any communication in layer-2. Acquired IP-to-MAC mapping is then saved in memory for future use [24]. For example: when a host wants to send a packet to the Internet, it must send the packet to the gateway. The host is configured with the IP address of the gateway but doesn't have the gateway hardware address

IP: source=0.0.0.0, destination=255.255.255.255 UDP: source port=68, destination port=67			
Octet 0	Octet 1	Octet 2	Octet 3
OP	HTYPE	HLEN	HOPS
0x01	0x01	0x06	0x00
XID			
0x12345678			
SECS		FLAGS	
0x0000		0x8000	
CIADDR (Client IP address)			
0x00000000			
YIADDR (Your IP address)			
0x00000000			
SIADDR (Server IP address)			
0x00000000			
GIADDR (Gateway IP address)			
0x00000000			
CHADDR (Client hardware address)			
0x000ABCDE			
0xF0120000			
0x00000000			
0x00000000			
DHCP options			
53: 1 (DHCP Discover)			
50: 192.168.1.100 requested			
55 (Parameter Request List):			

Figure 2.5: DHCP Discover Message

which must be obtained by ARP before any further communication can proceed.

2.2.6.4 Packet structure

ARP packets contain the following fields-

1. Hardware type (HTYPE)

This field specifies the network protocol type. Example: Ethernet is 1.

2. Protocol type (PTYPE)

This field specifies the internetwork protocol for which the ARP request is intended. For IPv4, this has the value 0x0800.

3. Hardware length (HLEN)

Octet 0	Octet 1	Octet 2	Octet 3
OP	HTYPE	HLEN	HOPS
0x02	0x01	0x06	0x00
XID			
0x12345678			
SECS		FLAGS	
0x0000		0x8000	
CIADDR (Client IP address)			
0x00000000			
YIADDR (Your IP address)			
0xC0A8000A (192.168.0.10)			
SIADDR (Server IP address)			
0xC0A80001 (192.168.0.1)			
GIADDR (Gateway IP address)			
0x00000000			
CHADDR (Client hardware address)			
0x000ABCDE			
0xF0120000			
0x00000000			
0x00000000			
DHCP options			
53 2 (DHCP Offer)			
1 (subnet mask): 255.255.255.0			
3 (Router): 192.168.0.1			
51 (IP address lease time): 86400s (1 day)			
54 (DHCP server): 192.168.0.1			
6 (DNS servers):			

Figure 2.6: DHCP Offer Message

Length (in octets) of a hardware address. (Example: Ethernet addresses size is 6.)

4. Protocol length (PLEN)

Length (in octets) of addresses used in the upper layer protocol. (The upper layer protocol specified in PTYPE.) (Example: IPv4 address size is 4.)

5. Operation

Specifies the operation that the sender is performing: 1 for request, 2 for reply.

6. Sender hardware address (SHA)

Media address of the sender.

7. Sender protocol address (SPA)

Internetwork address of the sender.

IP: source=192.168.0.1, destination=255.255.255.255			
UDP: source port=67, destination port=68			
Octet 0	Octet 1	Octet 2	Octet 3
OP	HTYPE	HLEN	HOPS
0x02	0x01	0x06	0x00
XID			
0x12345678			
SECS		FLAGS	
0x0000		0x8000	
CIADDR (Client IP address)			
0x00000000			
YIADDR (Your IP address)			
0xC0A8000A (192.168.0.10)			
SIADDR (Server IP address)			
0xC0A80001 (192.168.0.1)			
GIADDR (Gateway IP address)			
0x00000000			
CHADDR (Client hardware address)			
0x000ABCDE			
0xF0120000			
0x00000000			
0x00000000			
DHCP options			
53: 2 (DHCP Offer)			
1 (subnet mask): 255.255.255.0			
3 (Router): 192.168.0.1			
51 (IP address lease time): 86400s (1 day)			
54 (DHCP server): 192.168.0.1			
6 (DNS servers):			

Figure 2.7: DHCP Request Message

8. Target hardware address (THA)

Media address of the intended receiver. In an ARP request this field is ignored. In an ARP reply this field is used to indicate the address of the host that originated the ARP request.

9. Target protocol address (TPA)

Internetwork address of the intended receiver.

ARP protocol parameter values have been standardized and are maintained by the Internet Assigned Numbers Authority (IANA).

2.2.6.5 Gratuitous ARP (GARP)

Host can send ARP request/reply with the intention of IP address conflict detection.

This is called gratuitous ARP. Gratuitous ARP is of two types-

IP: source=192.168.0.1; destination=255.255.255.255 UDP: source port=67; destination port=68			
Octet 0	Octet 1	Octet 2	Octet 3
OP	HTYPE	HLEN	HOPS
0x02	0x01	0x06	0x00
XID			
0x12345678			
SECS		FLAGS	
0x0000		0x8000	
CIADDR (Client IP address)			
0x00000000			
YIADDR (Your IP address)			
0xC0A8000A			
SIADDR (Server IP address)			
0xC0A80001			
GIADDR (Gateway IP address switched by relay)			
0x00000000			
CHADDR (Client hardware address)			
0x000ABCDE			
0xF0120000			
0x00000000			
0x00000000			
DHCP options			
53 (DHCP ACK) or 6 (DHCP NAK)			
1 (subnet mask): 255.255.255.0			
3 (Router): 192.168.0.1			
51 (IP address lease time): 86400s (1 day)			
54 (DHCP server): 192.168.0.1			
6 (DNS servers):			
<ul style="list-style-type: none"> • 8.8.8.8, • 4.4.4.4 			

Figure 2.8: DHCP Acknowledgement Message

1. Gratuitous ARP request

This is done by the host without any real target machine

2. Gratuitous ARP reply

This is an ARP reply without being requested.

ARP is not a secured protocol and several research works have proposed modifications to secure it [25]–[27].

2.3 Circuit Switching

Circuit switching is a dedicated connection model where nodes experience the network as if they are connected to a physical electrical circuit.

2.3.1 Synchronous Digital Hierarchy (SDH)

Synchronous Optical Networking (SONET) and Synchronous Digital Hierarchy (SDH) are standardized protocols that transfer multiple digital bit streams synchronously over optical fiber using lasers or highly coherent light from light-emitting diodes (LEDs). At low transmission rates data can also be transferred via an electrical interface. The method was developed to replace the plesiochronous digital hierarchy (PDH) system for transporting large amounts of telephone calls and data traffic over the same fiber without synchronization problems.

2.3.2 Plesiochronous Digital Hierarchy (PDH)

The plesiochronous digital hierarchy (PDH) is a technology used in telecommunications networks to transport large quantities of data over digital transport equipment such as fibre optic and microwave radio systems. The term plesiochronous is derived from Greek *plēsios*, meaning near, and *chronos*, time, and refers to the fact that PDH networks run in a state where different parts of the network are nearly, but not quite perfectly, synchronized.

PDH networks are now mostly replaced by SDH.

2.3.3 Protocol signaling in Carrier Networks

Understanding of protocol signaling procedure is required to optimize the network while ensuring successful operation of various nodes that are attached to the transmission network. The protocol signal procedure of LTE has been discussed as the reference cellular technology.

2.3.3.1 Long Term Evolution (LTE)

In this section the procedures executed by LTE user equipment (UE) and the various LTE network elements in order to provide the services requested by the UE has been briefly touched upon.

A) Network Architecture

LTE network architecture is depicted in Figure 2.9.

The functions of the various network elements are:

1. Evolved Node B (eNodeB)

In LTE radio, evolved NodeB is the base station [28]. Radio Resource Management functions, IP header compression, encryption of user data streams,

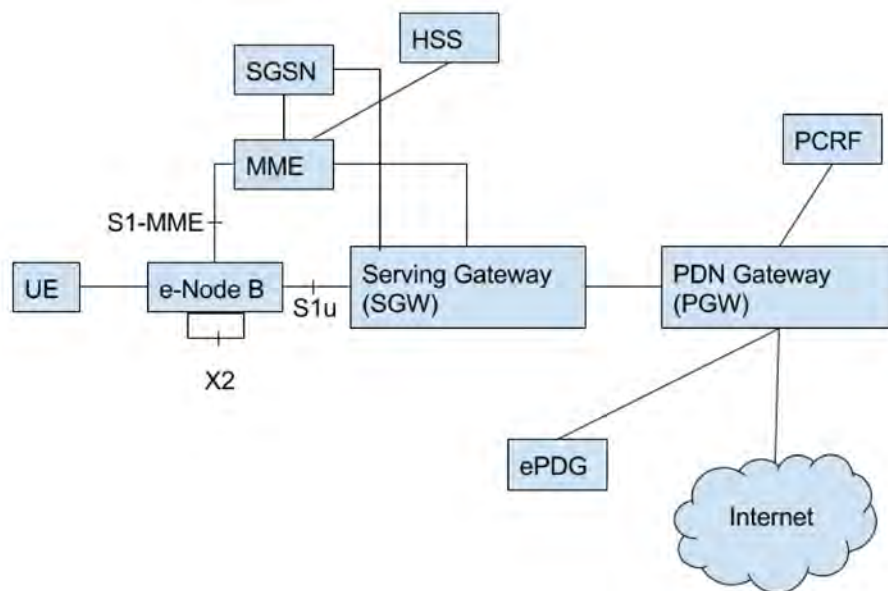


Figure 2.9: LTE (Long Term Evolution) Network Architecture

selection of an MME, routing of user plane data to S-GW, scheduling and transmission of paging message [29].

2. Mobility Management Entity (MME)

Non-access stratum (NAS) signaling (EPS Mobility Management and Connection Management) and security, AS security, tracking area list management, PDN GW and S-GW selection, handovers (intra- and inter-LTE), authentication, bearer management.

3. Serving Gateway (S-GW)

The local mobility anchor point for inter-eNodeB handover; downlink packet buffering and initiation of network-triggered service requests, lawful interception, accounting on user and QCI (QoS Class Identifier) granularity, UL/DL (Uplink/Downlink) charging per UE.

4. PDN Gateway (P-GW)

UE IP address allocation, packet filtering and PDN (Packet Data Network) connectivity, UL and DL service-level charging, gating and rate enforcement.

B) eNodeB to S-GW bearer (S1) Setup

S1 bearer is responsible for transporting packets between an eNodeB and an S-GW. Setup procedure of S1 bearer is depicted in Figure 2.11.

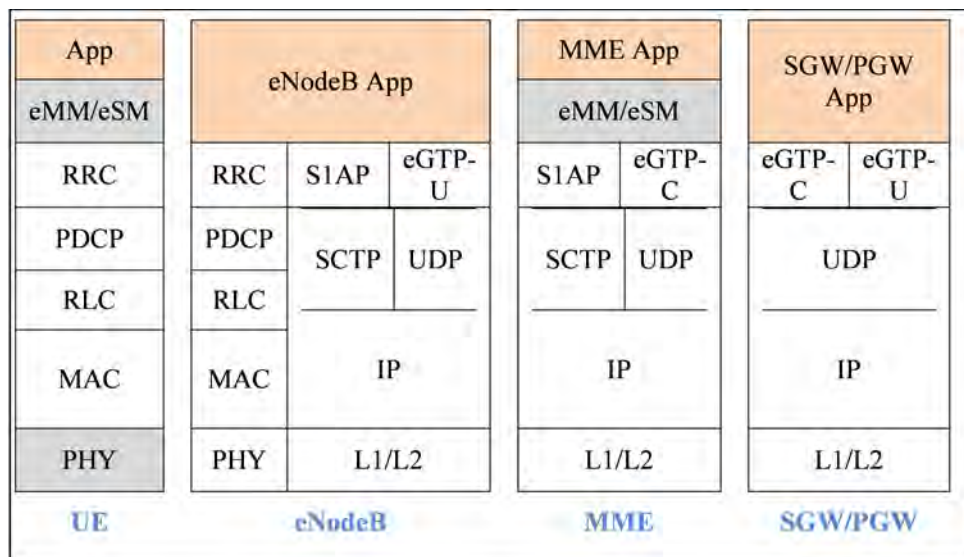


Figure 2.10: Protocol stack at various LTE network elements

C) eNodeB to eNodeB (X2) Interface Setup

LTE X2 interface has two functions: load or inference related function and handover related information. X2 interface setup procedure is depicted in Figure 2.12.

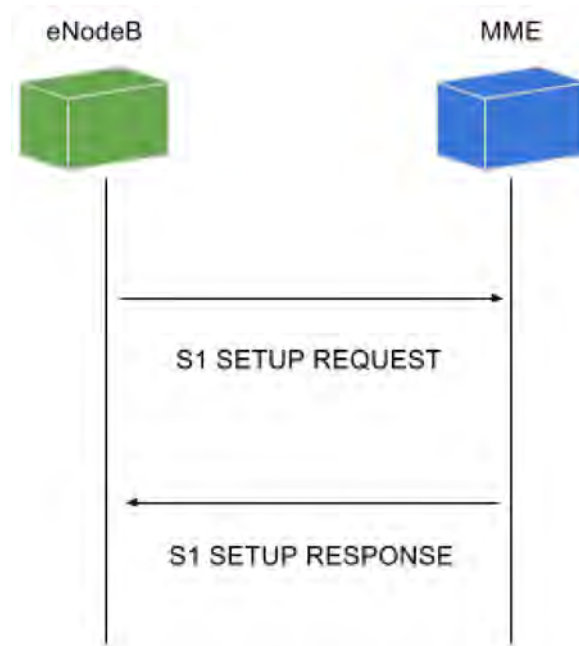


Figure 2.11: eNodeB to S-GW bearer (S1) Setup Procedure

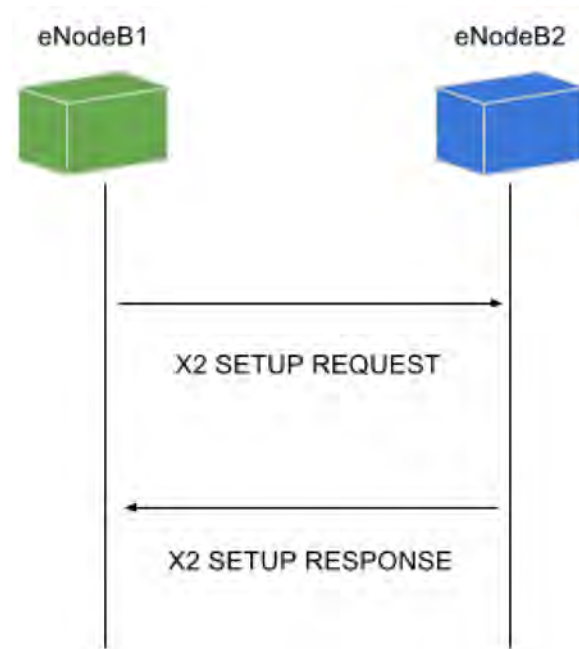


Figure 2.12: eNodeB to eNodeB (X2) Interface Setup Procedure

2.3.4 Carrier Access Network Architecture

GSM, UMTS, LTE cellular Access Network (AN) nodes such as BTS, node-B and eNodeB connect to each other via transmission links. In order to ensure protection in an outage, ring topology is usually deployed while some last-mile spur link or chain may exist where ring formation is not feasible. Some (if not all) of these links are usually microwave links as they are easier to deploy than laying optical fiber and sometimes fiber deployment is not feasible at all due to geographical condition such as hills. Traffic that passes a transmission link can be introduced from an adjacent AN node (i.e. from air interface of the AN node) or can come from a distant node through other links. The concept of the traffic introduced in the network from air interface will be termed as 'local traffic' throughout this document.

The unique architecture of cellular access network requires particular attention in protection network protocol design to ensure better performance. In GSM, BTS traffic flows through intermediate links to BSC using the Abis interface. Thus traffic from all BTS nodes in a ring will travel toward a common location. Similarly in UMTS, NodeB traffic will flow towards RNC using IuB interface. Similar scenario is applicable for LTE. This unique pattern of traffic being destined towards a common node is explored in this paper.

Figure 2.13 shows a part of access network of a telecommunication operator. Name of the operator is omitted for anonymity. In this figure Node1, Node2, Node3, Node4, Node5 and Node6 are all AN node locations that are connected to each other with microwave links to form a ring topology. Node1 is a fiber connected node and traffic from all other nodes will travel to this node to reach next layer of transmission (i.e. Aggregation) towards its destination. Here the traffic that is introduced in the network through the air interface of a node (i.e. Node1) is termed 'local traffic' of that node. If other nodes outside the ring join the ring at a ring node then sum of traffic of those nodes is also added to local traffic. In Figure 2.13 Node8 and Node10 join the ring at Node6.

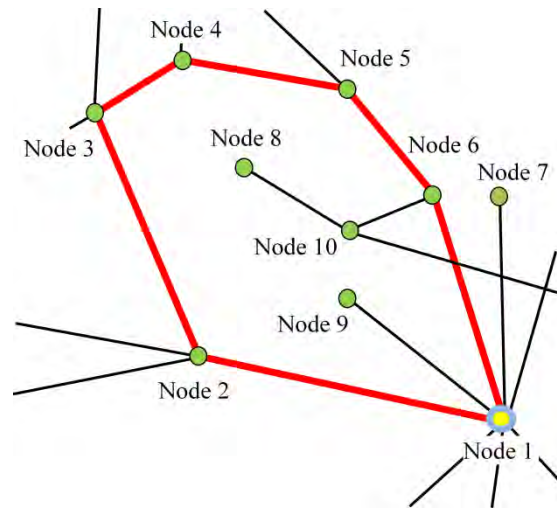


Figure 2.13: Part of a telecommunication carrier access network

2.4 Wireless Transmission and Adaptive Modulation in Carrier Networks

2.4.1 Microwave Transmission

The links connecting the access network (AN) nodes are responsible for carrying the traffic arriving from nodes. Microwave communication systems are often deployed due to their small footprint requirement and simpler deployment process compared to alternative options such as optical fiber. Moreover, optical fibers may not be even a feasible option due to embargo on land digging by regulatory bodies. Furthermore, microwave links are immune to disasters and effective in security, such as the terrorism countermeasures that have recently been gaining importance. Use of microwave communication is expanding widely all over the world [30].

2.4.2 Adaptive Modulation

Due to increased bandwidth demand of next generation cellular technologies such as LTE, microwave links are required to support high capacity. Although technologies such as cross polarization interference cancellation (XPIC) and recent E-band radios can achieve improved capacity, high capacity in microwave links is still a challenging target to achieve due to susceptibility of microwave signal to weather condition such as rain. Adaptive modulation is one powerful technology to address this problem. Depending on weather condition adaptive modulation can provide high capacity

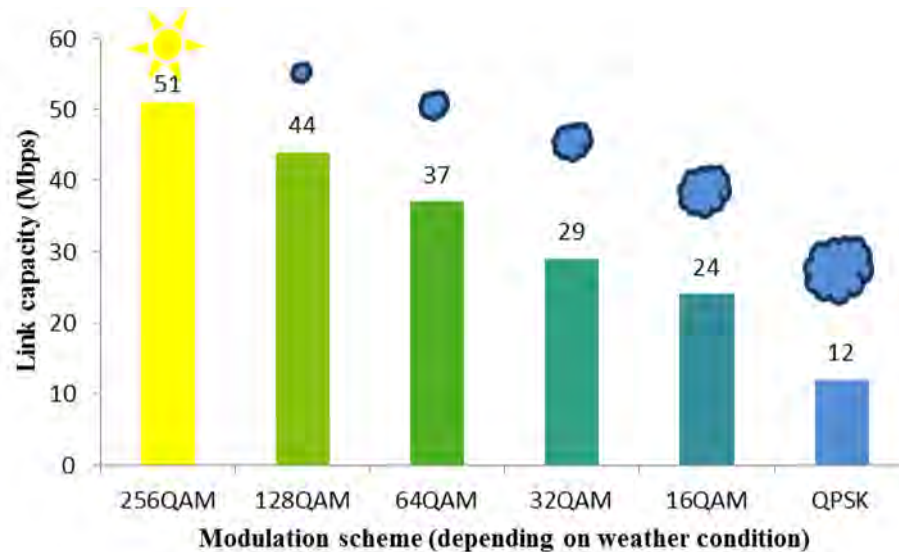


Figure 2.14: Varying capacity due to weather condition in Adaptive Modulation (AM)

during favorable weather whereas still maintaining a reliable connectivity for critical services during extreme weather period by switching to stronger modulation scheme at the price of reduced capacity. This is depicted in Figure 2.14.

2.5 Software Defined Networking (SDN)

In SDN, control plane is decoupled from data plane [31]. This allows greater control of the network through programming as the control plane is programmable while enabling faster deployment of new services. A software defined network (SDN) comprises of three main layers:

- Application Layer
- Control Layer
- Data Layer

The application layer consists of applications that communicate with SDN controller(s) to program network behavior as per requirement via the NBI (Northbound Interface). SDN controllers then translate the network requirement from SDN applications transfer into the switching devices at data layer via control-data-plane

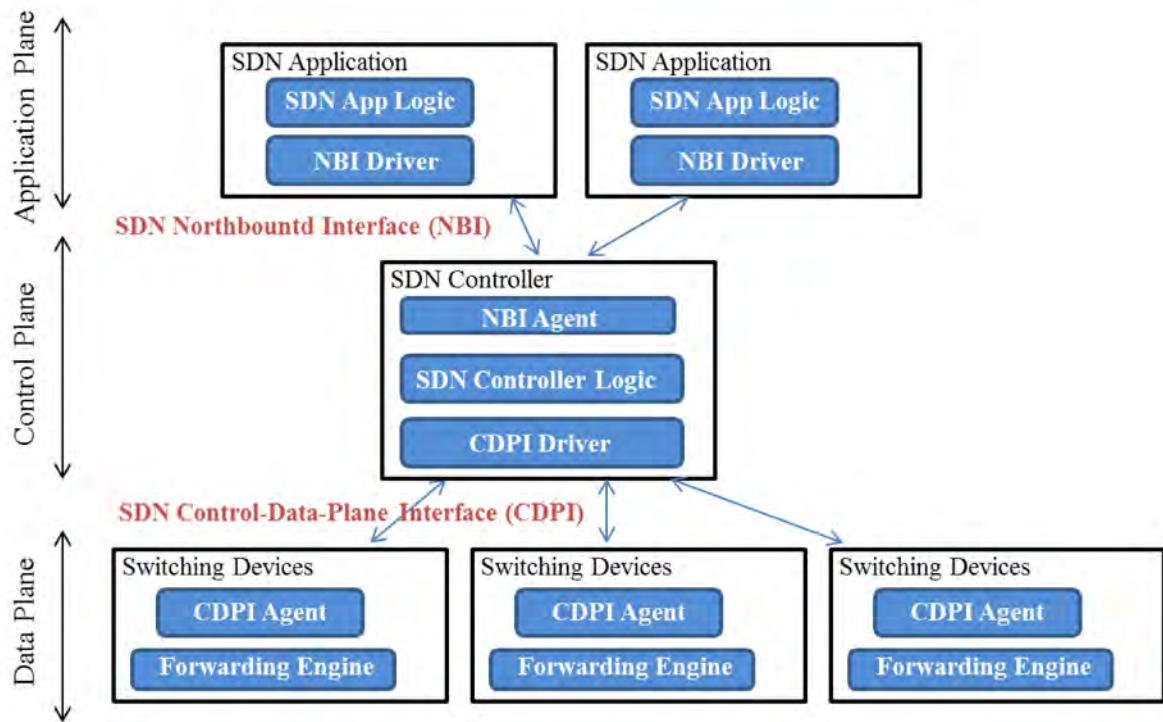


Figure 2.15: Reference diagram for Software Defined Network (SDN)

interface (CDPI). Finally, switching devices at the data layer forwards traffic according to instructions received from the SDN controller. Figure 2.15 shows a reference diagram of SDN.

SDN controllers also has the additional role of providing abstract network topology information which is consumed by SDN applications as an input to their internal logic.

2.6 Issues In Carrier Network Transmission

2.6.1 Channel Adaptability

With the development of adaptive modulation (AM) technology (discussed in chapter 2) channel bandwidth in transmission network can vary with time. While AM brings substantial advantage to transmission network, network protocols are required to be designed to handle changes in link capacity caused by it (AM).

In TDM technologies such as SONET/SDH/PDH, AM-downshift (bandwidth degradation) will cause one or more unit of traffic to be dropped depending on the degraded bandwidth. Since, TDM based transmission technologies work in fixed

bandwidth units, a small drop in bandwidth may lead to a large amount of traffic being dropped. For example, an entire STM-1 (155Mbps) may be dropped because of a bandwidth degradation of 30Mbps. For small capacity links (e.g. 180Mbps) it may be even impossible to use adaptive modulation in SDH as a bandwidth drop to 150Mbps will make total outage as STM-1 can't be accommodated in that link. In packet switched networks, traffic can be dropped on a per packet basis based on priority of the traffic by means of Quality of Service (QoS). However, this will cause traffic loss due to low priority packets being dropped.

Another approach can be - considering AM degradation as a link failure and re-routing all traffic from the affected link. This approach will require provisioning adequate capacity in the backup path which is inefficient and defeats the purpose of AM.

2.6.2 TDM: SONET, SDH & PDH

The main disadvantage of TDM is the lack of flexibility to allow efficient use of available bandwidth for higher throughput. The drawbacks of TDM technology include-

1. Low granularity (only fixed sized capacity is possible)
2. Lack of ability to re-use capacity among users
3. Complex configuration requirement of each service
4. Difficult for operation and maintenance (rerouting a link requires massive work to reroute all traffic that are traveling through that link)
5. Adaptive modulation may lead to large amount of bandwidth loss or complete outage due to fixed capacity model

Figure 2.16 and Figure 2.17 illustrate possible data rates in SONET and SDH respectively.

2.6.3 IP/MPLS

IP and MPLS are complex higher layer protocols which require significant resource overhead. The drawbacks of these higher layer protocols include:

1. Manual configuration of LSP is difficult. Sub-optimal for multi-point and multicast services due to MPLS being a point-to-point technology [32].
2. LDP adds complexity, requires IGP (i.e. OSPF, ISIS)
3. Need complex routing protocol and thus impose resource overhead
4. Need more processing time (compared to L2)
5. Need expertise and configuration of each equipment

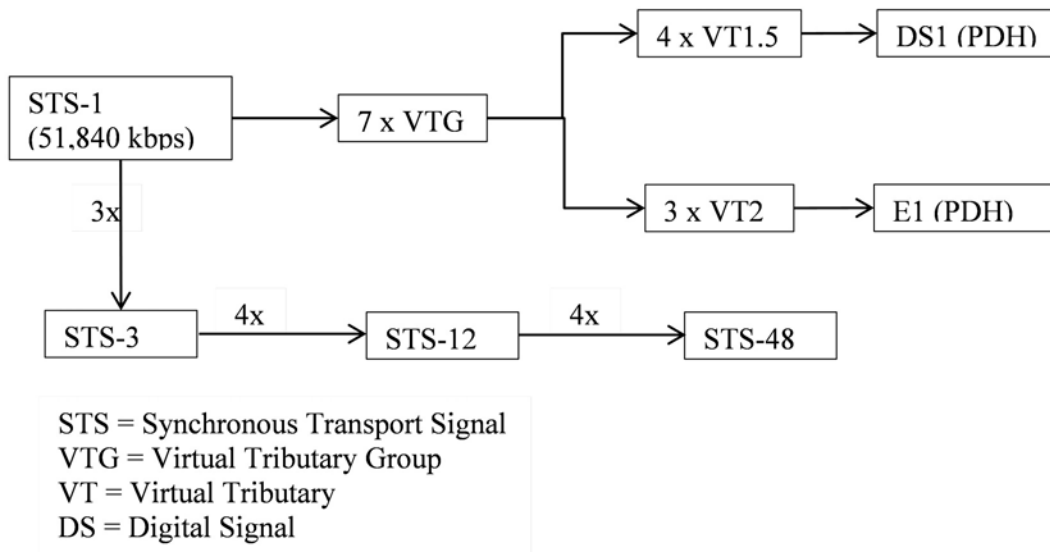


Figure 2.16: Data rates in Synchronous Optical Networking (SONET)

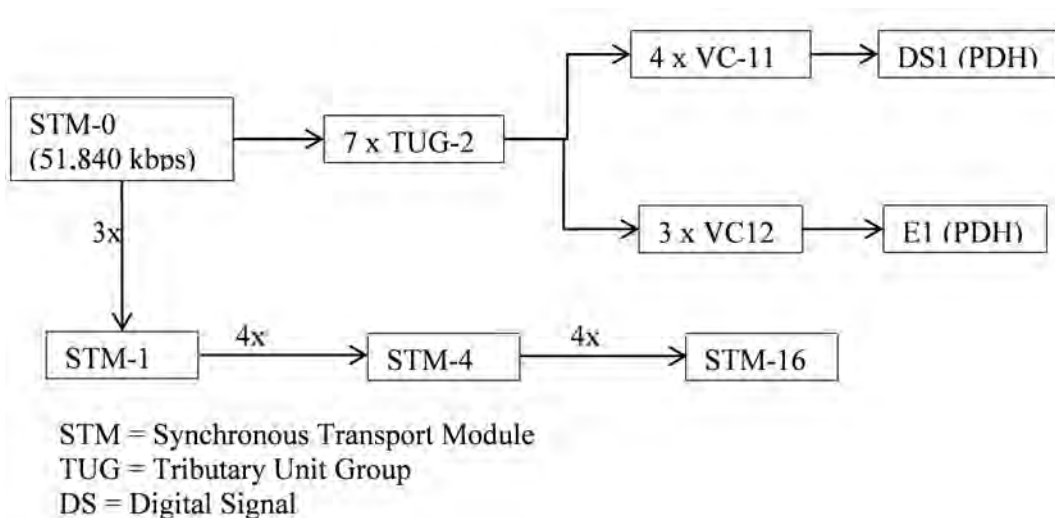


Figure 2.17: Data rates in Synchronous Digital Hierarchy (SDH)

6. IGP convergence is slow
7. MPLS FRR Protection can provide protection for only a single node or link

2.6.4 Carrier Ethernet

2.6.4.1 Protection Switching

In order to ensure uninterrupted communication in a telecommunication network, network level redundancy is deployed. To protect a network from a link failure, traffic is switched to a redundant link on failure of the primary link. This switchover should be within 50ms from the failure. Spanning Tree Protocol (STP)/Rapid Spanning Tree Protocol (RSTP) used in Ethernet networks to support redundant links can't guarantee the fast response time required by carrier networks.

With the advent of ITU-T G.8032 Ethernet Ring Protection Switching (ERPS) protocol, Ethernet is now capable of carrier grade switchover response time.

ERPSv1 (standardized in 2008) didn't support multi-ring/ladder network which is now supported by ERPSv2 (standardized in 2010). Later in 2012 ERPSv3 introduced protection recommendation for end-to-end service with support of ITU-T G.8031 Linear Protection Switching.

2.6.4.2 Scalability

A single Ethernet network is a layer-2 broadcast domain where a broadcast packet sent by one host in the network will travel every link and reach every host. All the hosts in the network have to process the broadcast packet. This limits the number of nodes that can be integrated in a single Ethernet network because a large number of hosts will generate an amount of broadcast that can quickly take a network down.

Thus Ethernet technology in itself is not scalable to a large number of nodes required by networks such as large enterprise, data centers and telecommunication carriers.

2.6.4.3 Limitation of VLAN based networks

Present Ethernet networks deploy VLANs to keep broadcast within limit. This approach has several limitations-

1. VLANs create different layer-2 domains and connectivity between two different layer-2 domains require additional layer-3 device (i.e. router)
2. Routers or layer-3 switches require handling more layer 3 subnets and IP address utilization becomes less [33]
3. Number of VLANs are limited (only 4096) and thus the number of nodes that are possible to be connected in a network using VLAN based broadcast control is limited.
4. VLANs must be configured in every layer-2 devices (switches).

The second limitation is overcome by IEEE 802.1ad provider bridging technique (commonly known as QinQ). Here, a double tagged frame can support a number of $4096 \times 4096 = 16777216$ VLANs. However, this aggravates the third problem. That is, a large number of VLANs now have to be configured and maintained.

2.6.5 Software Defined Networking (SDN)

While SDN can provide more control of the network through programmability, it is not yet a mature solution and many of its components including the north-bound interface, high-level programming language, applications and forwarding devices are still not available [34]. Although SDN controllers are not required to be physically centralized, there is still a gap between SDN and traditional network architecture in terms of how fast the control plane can communicate with the data plane. A hybrid approach which yields time critical protocol operations (e.g. failure recovery, reply to ARP request etc.) to switching devices may be more viable rather than relying solely on the SDN controller.

2.6.6 Research works on Scalability of Ethernet

To achieve a floodless layer-2 network, SEATTLE [35] uses a link-state routing to maintain switch level topology. Although SEATTLE provides end host backward compatibility, it is not backward compatible with existing Ethernet switches. Portland [36] uses a fabric manager as a central controller to suppress ARP broadcast in a multi rooted tree topology. The main drawback of this approach is topology dependency and scalability problem of a single controller. The work in [37] proposes

a Distributed Registration based Address Resolution Protocol (DRARP) and an End user enabled Mac-in-MAC (EMiM). This approach requires modification of existing hardware in end hosts. NetLord [38] uses a push based ARP proxy model but broadcast still occurs on new virtual machine (VM) boot. Moreover its memory requirement is proportional to number of servers and therefore may lead to scalability issue as number of servers grow. Torii-HLMAC [39] attempts to improve PortLand as a distributed alternative but it is also dependent on a specific topology, fat tree. Moreover Torii-HLMAC uses a new MAC address, Hierarchical Local MAC (HLMAC), and tree based forwarding scheme which makes it incompatible with traditional Ethernet switches. SAL (Smart Address Learning) [40] uses a network topology where a network devices called *edge bridges* are used to connect between network segments and the network core. SAL then only learns address mapping for tenants that reside inside a network segment. FSDM [41] and SEASDN [42] propose SDN based ARP proxy. The limitation of SDN based approach [41]–[43] is that, it suffers from slower failure recovery compared to protection mechanism available for existing Ethernet technology such as ITU-T G.8032. Moreover, these approaches are subject to scalability problem of a single controller and communication with the central controller, located across backbone network, involves a high latency. [44] introduces *Packet Templates*, a proposed feature in SDN which allows autonomous generation of packets by the SDN switches. This ability is then used to improve ARP packet processing in SDN networks.

2.7 Summary

In this chapter, different transmission technologies used in carrier networks has been introduced. Carrier networks started with TDM based transmission technologies like SDH/SONET which are now being replaced with packet switching technologies. Protocol signaling of LTE has been presented as a reference for optimizations described in subsequent chapters. Adaptive Modulation, a recent advancement in wireless transmission technology, also has been introduced in this chapter.

This chapter also presents issues in different transmission technologies. TDM based technologies suffer from lack of flexibility. IP/MPLS are less efficient in terms of switching speed due to their complex nature and also their protection switching

models suffer from high delay and requiring a large number of manual configurations.

In order to a higher efficiency transmission protocols should be designed to handle channel capacity changes for which none of the present protocols are capable.

The main drawback of Ethernet is its lack of scalability for large number of nodes.

Chapter 3

Provisioning Channel Adaptability in Ethernet by Improving ERPS

3.1 Introduction

Ethernet is a promising technology for telecommunication network applications due to its cost-effectiveness and simplicity. Recently added features such as QoS, scalability and OAM (operation, administration and management) through standardizations such as IEEE 802.1Qay, IEEE802.1Q, IEEE802.1ad IEEE 802.1ag has transformed Ethernet into a carrier-grade technology. With the subsequent development of ITU-T G.8032 recommendation which adds sub-50ms failure recovery capability to Ethernet reusing generic standardized functions, Ethernet has finally become a lucrative replacement for its rival technologies such as SONET/SDH for carrier network transmission. The prosperity of ERPS comes from its tailored design for specific technology and it should be upgraded as technology advances to uphold this success. To improve ERPS, the papers [45], [46] proposed an efficient filtering database (FDB) flushing procedure. The works in [47], [48] focus on optimal ring hierarchy selection, RPL placement and capacity dimensioning in multi-ring topology. The research in [49] presents resource planning for the ERPS considering multiple instances. As cellular technology advances beyond 4G, deployment of access network nodes will be denser and more links with higher capacity will be required to connect these nodes. Ease of installation and high capacity of microwave links make them an excellent choice to support this growth.

Adaptive Modulation (AM) is an emerging technology in access network to address high capacity requirement of mobile evolution by ensuring reliable transmission during unfavorable channel condition while providing high capacity when channel

condition permits. AM has not been considered in the design of ERPS. As adaptive modulation can cause link capacity to vary by more than 100% depending on channel conditions by changing modulation scheme, it needs special consideration to be supported in ERPS to ensure optimum throughput. No academic research work has been done to support AM in ERPS to the best of our knowledge. A few proprietary implementations of solution for bandwidth degradation due to AM for ERPS have been reported but they need special proprietary subsystem for rerouting traffic instances. Also these technologies lack energy efficient instance rerouting. Also no previous work is done on performance analysis of ring topology under link degradation condition of AM to the best of our knowledge.

In this dissertation an enhancement to ITU-T G.8032 protocol and a multiple instance ERPS ring design principle enabling efficient selection of ERPS instances to switch from an affected link to ensure overall optimal capacity utilization has been proposed. Our proposed design ensures energy efficient instance rerouting. Throughput analysis of a ring network under link bandwidth degradation due to AM changes has been presented. It is shown that the number of instances diverted from the affected link significantly impacts network throughput. We also propose a guideline to simplify calculation of optimum threshold values that are required by the proposed enhanced protocol.

3.2 Protection switching optimization

To optimize throughput and enable energy efficient traffic rerouting in reduced link bandwidth condition, caused by adaptive modulation changes, we propose a multi-instance ring network design in section 3.2.1. ERPS control process logic enhancement is proposed in section 3.2.2. Threshold calculation of values required by the proposed algorithm is given in section 3.2.3. Finally, throughput analysis in ring topology subject to bandwidth degradation is presented in section 3.2.4.

3.2.1 ERPS Multi-instance ring design

As discussed in the previous section, multiple ERPS instances can be configured in an ERPS network where each instance is responsible for protection of a subset of services. Traffic protection during bandwidth degradation can be performed by

triggering protection switching of one or more ERPS instances and thus avoiding the affected link. This will ensure utilization of free capacity of alternative path and prevent packet loss which would otherwise occur in the bandwidth degraded link.

We design the ring network by creating a separate ERPS instance corresponding to each ring node. That is, we configure multiple instances in a physical ring to protect traffic in that physical ring. One instance is created for each node in the ring. To further explain, in Figure 3.1, a ring of six nodes is shown. We create six ERPS instances corresponding to these six nodes. Two instances corresponding to nodes n5 and n6 are shown in Figure 3.1. Instances corresponding to remaining nodes are not shown for simplicity. In Figure 3.1, the solid (red) line shows traffic flow for the instance belonging to node n6 and the dashed (blue) line represents traffic flow for the instance belonging to node n5. Signal degradation in link n1-n6 may cause packet loss due to reduced bandwidth. To avoid this, our proposed algorithm will initiate protection switching for ERPS instance n5.

In a ring or sub-ring with more than one instance, every link carries traffic belonging to multiple instances unless the link is the common RPL of all of the instances. In our proposed algorithm, when a link that is carrying multiple instances is affected by signal degradation, energy efficient choice of instances is made for being steered away from the affected link. In Figure 3.1 instance n5 is selected from instances n5 and n6 to be steered away from the affected link n1-n6. This saves energy by avoiding rerouting of instance n6 which would require more transmission power and processing nodes.

Every instance is configured with two threshold values of bandwidth, one for each ring port, at every ring node. That is the threshold configuration varies for every instance in a node. Threshold values for the same instance will be different on different links. Ring ports at both end of a link are configured with the same threshold bandwidth for a ring instance. To illustrate this, in Figure 3.1, ERPS instance n6 will be configured with a threshold value for each ring port of node n6. Instance n6 will be configured with a different threshold value at node n5. Calculation of this threshold is shown in section 3.2.3.

3.2.2 ERPS Control Process Logic Enhancement

According to existing G.8032 standard protection switching is triggered on detection of signal failure (SF). We propose to augment the existing ERP standard to respond to signal degradation (SD) or modulation changes. ERPS control process is notified of the SD event by microwave transceiver. This notification includes new link bandwidth. Alternatively microwave transceiver can be periodically polled by ERPS control process to detect any change in modulation scheme. The exact process of detection of modulation scheme is out of scope of this paper and we assume that ERP control process knows when link bandwidth changes due to adaptive modulation change. Algorithm 1 is then used by the ERPS control process to decide protection switching action.

As given in the Algorithm 1, ERPS control process is modified to consider the configured threshold for the ring port to decide protection switching behavior on adaptive modulation change (i.e. upshift, downshift). When bandwidth degradation (AM downshift) occurs, protection switching is initiated if bandwidth falls below configured threshold. Similarly when bandwidth is restored (AM upshift) to the configured threshold, protection recovery is initiated.

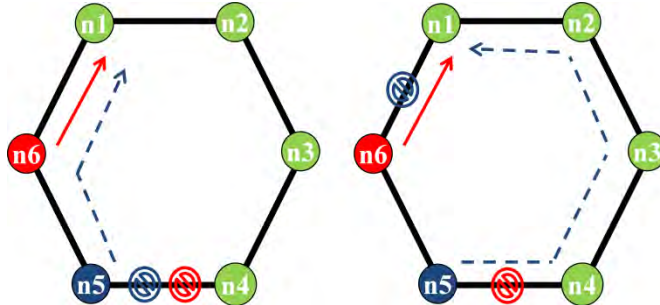


Figure 3.1: Instance creation per ring node – before (left) and after (right) protection switching due to signal degradation at link between n1 and n6

Algorithm 1 Protection Switching on Signal Degradation

- 1: p = ring port where SD is detected
 - 2: b = configured threshold at port p
 - 3: c = new link capacity after AM bandwidth change
 - 4: **if** $c < b$ **and** ring port p is forwarding **then**
 - 5: Block ring port p
 - 6: Initiate protection switching by transmitting R-APS (SF)
 - 7: **else if** $c \geq b$ **and** ring port p is blocked **then**
 - 8: Initiate restoration by transmitting R-APS (NR)
 - 9: Unblock ring port p followed by RPL block
 - 10: **end if**
-

3.2.3 Calculation of threshold

Protection switching behavior due to adaptive modulation change will depend on the configured threshold that is used by the proposed algorithm and therefore careful selection of the threshold values must be ensured during network planning.

As we will show in the next section, maximum throughput is achieved when an optimum number of instances are switched. If the number of instances for which protection switching is initiated at a certain scenario is not properly selected then throughput will not be optimum. In order to achieve maximum throughput we minimize packet loss at the affected link. The algorithm for threshold bandwidth calculation per service instance at every node is discussed in the next paragraph.

Let $L = L_1, L_2, \dots, L_N$ be the array of all links in the ring. Now we split the array L into two arrays where links at the left side of RPL and fiber connected node are stored at array $LL = L_{L1}, L_{L2}, \dots, L_{LP}$ and links at the right side of RPL and fiber connected node are stored at array $L_R = L_{R1}, L_{R2}, \dots, L_{RQ}$. Where $N = P + Q$ is the total number of links in the ring. Here fiber connected node is the common destination of all traffic as discussed in Chapter 2, section 2.2.3.1. Links are sorted according to their distance from the fiber connected node. For the i^{th} link from fiber connected node, we define cumulative traffic $C_j(L_{Li})$ as the cumulative sum of inbound traffic $T_j(L_{Li})$ of j^{th} upstream ring node from the link.

$$C_j(L_{Li}) = \sum_j T_j(L_{Li}) \quad (3.1)$$

Here inbound traffic includes local traffic and traffic from spur nodes. The term local traffic is defined in Chapter 2, section 2.2.3.1. Now for all instances corresponding to j^{th} upstream node from i -th link threshold is set at $C_j(L_{Li})$ for all links of L_L . Similarly, threshold is calculated for all links of L_R .

3.2.4 Throughput Analysis

Here, we develop an analytical model to estimate throughput in a ring topology. Although we apply this analysis to verify the effectiveness of our network design and algorithm for ERPS (Ethernet Ring Protection Switching), this analysis is also applicable for any other ring protection technology. Here it is assumed that all links have equal capacity. Also for the sake of simplicity it is considered that aggregation node $n1$ doesn't have any local traffic.

Let $r1$ and $r2$ be the RPL owner and RPL neighbor nodes respectively as in Figure 3.2. Thus link $r1$ - $r2$ is the RPL (Ring Protection Link). Node $n1$ is the aggregation node for the ring. That is, node $n1$ is the common node where all traffic from all nodes in the ring accumulates. This was discussed in Chapter 2, section 2.2.3.1. As traffic from all nodes will flow towards $n1$, total traffic at node $n1$,

$$T = \min\left(\sum_{i=r1}^{n2} L_i, C\right) + \min\left(\sum_{i=r2}^{n1} L_i, C\right) \quad (3.2)$$

Where L_i is the traffic generated by i th node and C is the link capacity. Now if bandwidth of link $d1$ - $d2$ is degraded due to adaptive modulation (AM) downshift

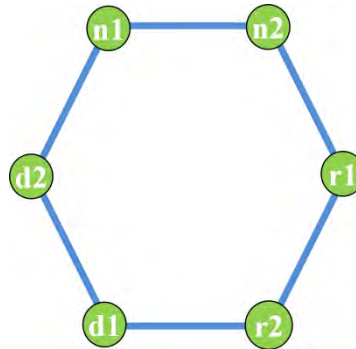


Figure 3.2: Illustration for throughput analysis in a ring topology

(modulation changed to lower efficiency scheme) and if the degraded bandwidth of link d1-d2 is C_{d1d2} then throughput,

$$T = \min\left(\sum_{i=r1}^{n2} L_i, C\right) + \min\left(\sum_{i=r2}^{d1} L_i, C_{d1d2}\right) + \min\left(\sum_{i=d2}^{n1} L_i, C - C_{deg}\right) \quad (3.3)$$

Where,

$$C_{deg} = \min\left(\sum_{i=r2}^{d1} L_i, C_{d1d2}\right) \quad (3.4)$$

Right hand side of equation (3.3) consists of three components. The first and third component represents traffic passing through unaffected links. The second component determines traffic through affected link.

Now let the traffic quantity be such that,

$$\sum_{i=r2}^{d1} L_i > C_{d1d2} \quad (3.5)$$

And, let x be the free bandwidth in the link n1-n2,

$$x = C_{n1n2} - \min\left(\sum_{i=r1}^{n2} L_i, C\right) \quad (3.6)$$

Where, C_{n1n2} is the bandwidth of the link n1-n2.

We utilize this free capacity in the optimized protocol by initiating protection switching on some ERPS instances in the affected link.

Resulting throughput will be,

$$T = \min\left(\sum_{i=r1}^{n2} L_i, C\right) + \min\left(\sum_{i=r2}^{d1} L_i - x, C_{d1d2}\right) + \min\left(\sum_{i=d2}^{n1} L_i, C - C_{deg}\right) + x \quad (3.7)$$

The throughput of the optimized protocol is found from equation (3.7) under the condition of equation (3.5) and from equation (3.3) when equation (3.5) is not applicable.

Therefore the throughput obtained by the optimized protocol,

$$x = \begin{cases} T = \min(\sum_{i=r1}^{n2} L_i, C) + \min(\sum_{i=r2}^{d1} L_i - x, C_{d1d2}) + \min(\sum_{i=d2}^{n1} L_i, C - C_{deg}) + x, & \sum_{i=r2}^{d1} L_i > C_{d1d2} \\ T = \min(\sum_{i=r1}^{n2} L_i, C) + \min(\sum_{i=r2}^{d1} L_i, C_{d1d2}) + \min(\sum_{i=d2}^{n1} L_i, C - C_{deg}), & \text{otherwise} \end{cases} \quad (3.8)$$

3.3 Results & Discussion

In this section we present simulation results to evaluate performance of existing ITU-T G.8032 and our proposed improved protocol. We created a simulation model of ITU-T G.8032 in OMNET++ simulation platform. The network given in Figure 3.3 was used to perform the simulation.

In Figure 3.3 switches are used to represent ERPS nodes with AM capable microwave links on ring ports. Traffic is generated by hosts which may represent co-located cellular access network (AN) node such as BTS, Node-B, eNodeB etc. As discussed in section 2.2.3.1, traffic will travel towards a common aggregation node to reach the common destination such as BSC, RNC etc. Here $switch_{48}$ represents the aggregation switch. The link $switch_{48} - switch_{server}$ is used as a probe to measure the throughput of the network. The link between $switch_{43}$ and $switch_{44}$ is configured as the ring protection link (RPL).

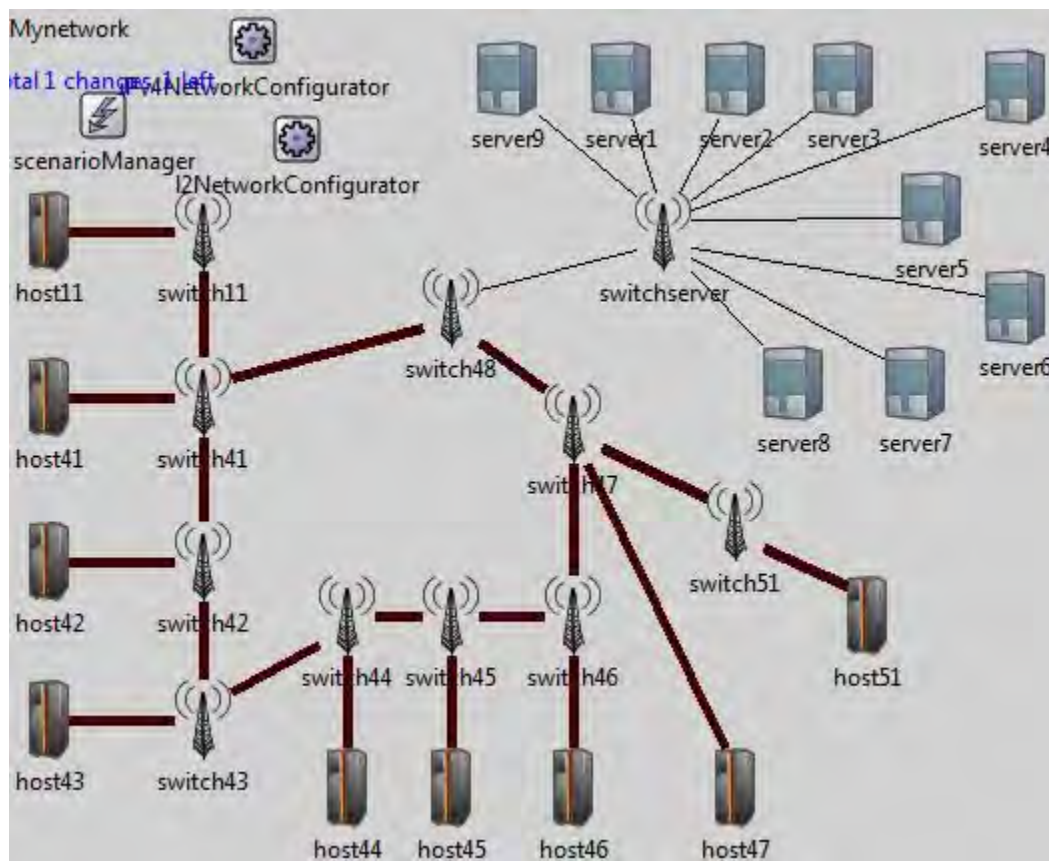


Figure 3.3: Simulation scenario in OMNET++

3.3.1 Effect of instance switching

In order to observe the effect of switching instances on overall throughput, threshold values of instances at the affected link ($switch_{41} - switch_{48}$) are varied and several simulations were conducted. As instances were switched gradually from the affected link, overall throughput at the link between $switch_{48} - switch_{server}$ was recorded. The set of results obtained in the series of simulations is plotted in Figure 3.4.

In Figure 3.4, initially switched instance count is zero and throughput is 37.1Mbps. This corresponds to behavior of unmodified ITU-T G.8032 where no action is taken in response to signal degradation (SD). Then switching one instance increases throughput to 43.8Mbps. Gradually throughput increases up to switching of three instances and peaks at 52.9Mbps. After this, switching of more instances results in decreased throughput. Therefore, we can conclude that optimal switching of instances away from the degraded link can improve overall throughput under signal degradation scenario.

3.3.2 Throughput comparison

In order to observe effect of adaptive modulation's bandwidth degradation on throughput, all the ring links are initially operated at 128QAM.

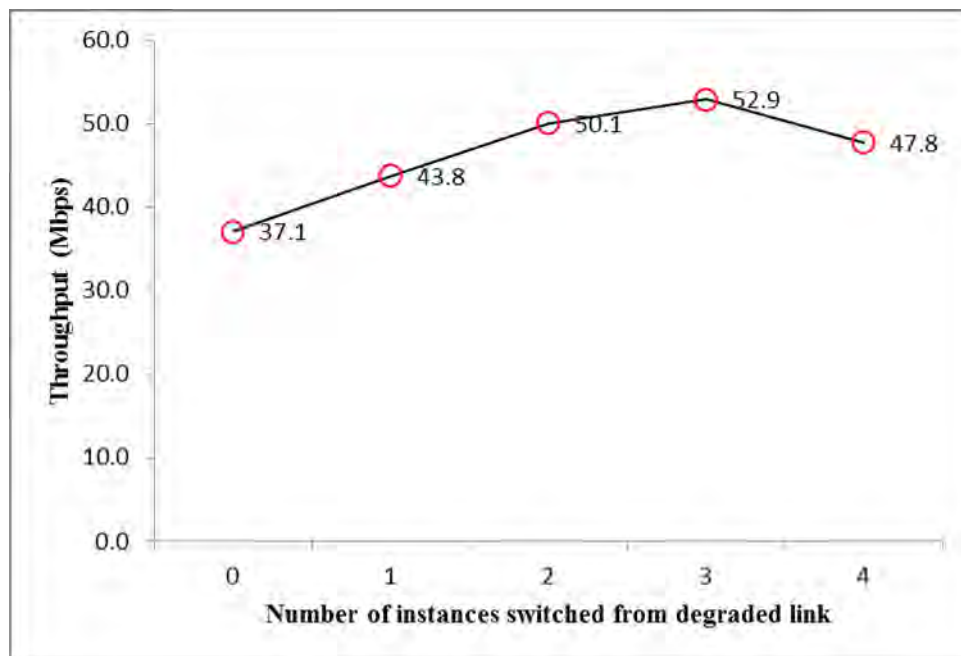


Figure 3.4: Effect of instance switching on throughput

We have considered a 7MHz microwave radio frequency (RF) channel bandwidth. This results in a data rate of 44Mbps (see Appendix A). Then modulation scheme is degraded to obtain link bandwidth as shown in TABLE 3.1.

As we can see from Figure 3.5, while bandwidth is degraded by AM jumping to lower order modulation, overall throughput is affected in existing ITU-T G.8032 protocol. The proposed protocol keep the overall throughput unchanged at maximum until severe degradation to QPSK. At QPSK link bandwidth is degraded to 12Mbps and ITU-T G.8032 throughput degrades to 37.0Mbps whereas the proposed protocol is able to maintain an approximately 65% higher throughput of 52.6 Mbps in this condition.

Table 3.1: BANDWIDTH DEGRADATION WITH TIME USED IN OMNET++ SIMULATION

Time (sec)	Degraded bandwidth (Mbps)
1	29
3	24
5	12

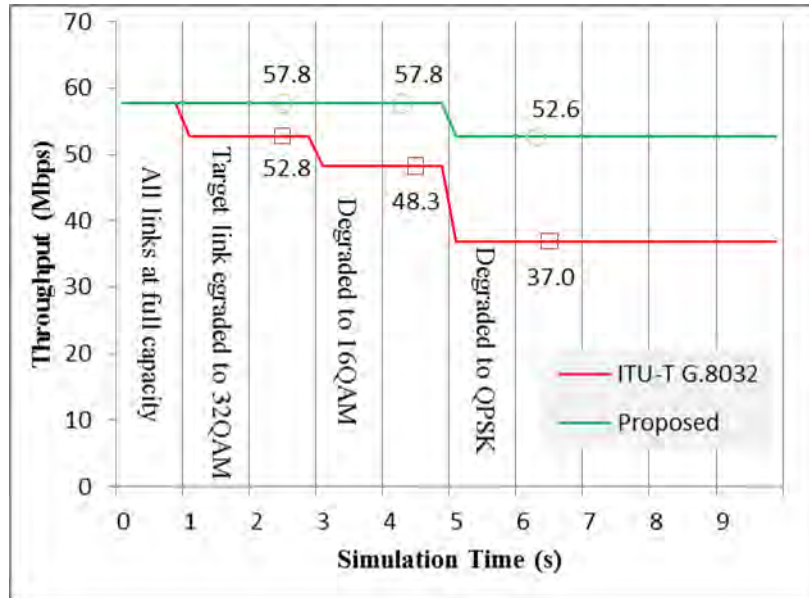


Figure 3.5: Performance comparison (Throughput)

3.3.3 Packet loss comparison

The throughput degradation shown in the throughput comparison graph (Figure 3.5) is due to packet loss in the affected microwave link (between $switch_{41}$ and $switch_{48}$). To get a clearer picture of effect of AM bandwidth degradation on packet loss, the affected link (between $switch_{41}$ and $switch_{48}$) is monitored for lost packets.

This is done by calculating the total inbound traffic at $switch_{41}$ including local traffic and traffic from links with $switch_{11}$ and $switch_{42}$ compared with transmitted traffic in link between $switch_{41}$ and $switch_{48}$. This gives the packet drop due to insufficient link bandwidth.

The result is plotted in Figure 3.6. It is to be noted that the same set of events was used (TABLE 3.1) for the packet loss experiment as the set of events used for throughput experiment. Other parameters such as packet size were also kept same for both experiments.

As we can see from Figure 3.6, packet loss is gradually increased in ITU-T G.8032 as modulation scheme switched to lower order. On the other hand our proposed scheme avoids any packet loss until 5s where modulation scheme switched to extremely low bandwidth. Even during this severe degradation, our proposed enhancement produces significantly lower packet loss.

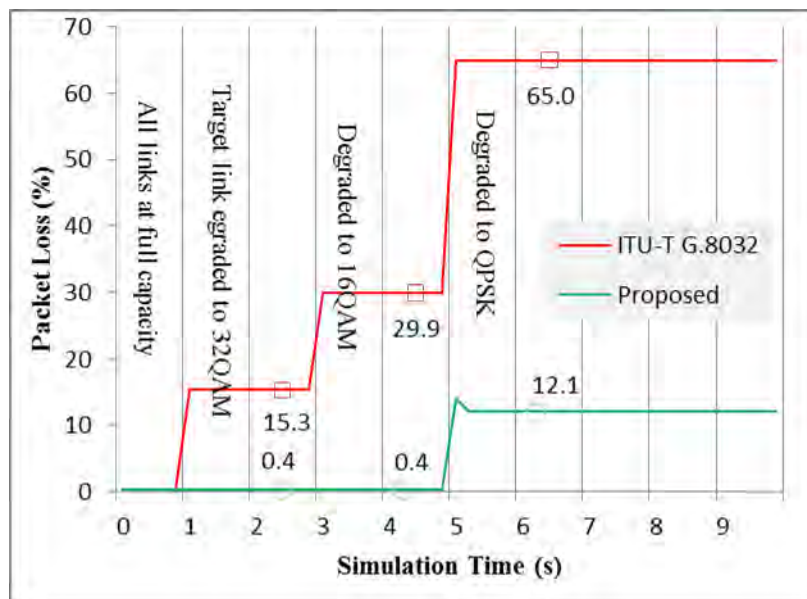
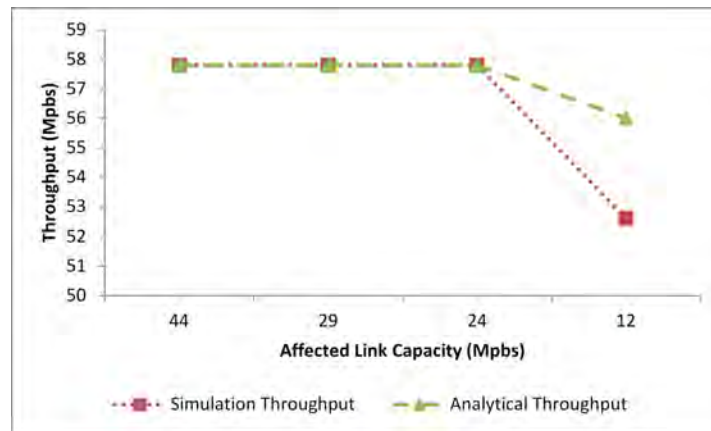


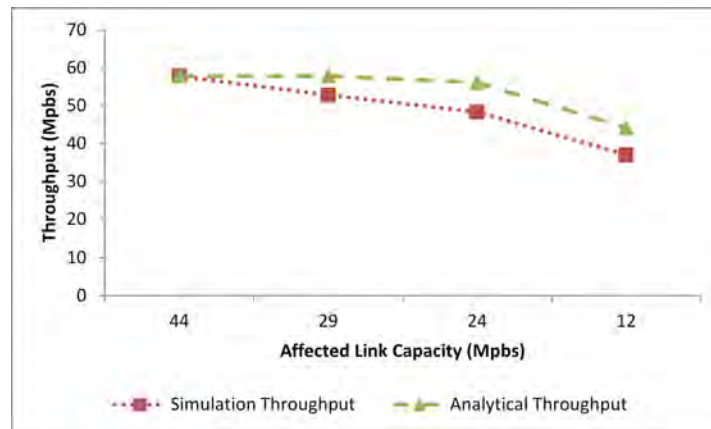
Figure 3.6: Performance comparison (Packet loss)

3.3.4 Comparison of Simulation Result with Analytical Model

We have compared the throughput suggested by mathematical analysis given in 3.2.4 with simulation throughput to verify the accuracy of the mathematical model. This is given in Figure 3.7. We can see that the simulation throughput matches the mathematical model's throughput with a small deviation which is due to packet header overhead. It is clear that our proposed modification performs better than the unmodified protocol from both simulation result and mathematical analysis.



(a) Proposed protocol



(b) Existing protocol

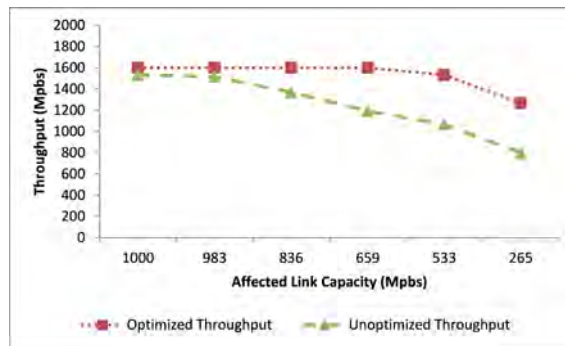
Figure 3.7: Comparison of simulation results with analytical model for proposed modified protocol and existing ITU-T G.8032 protocol

3.3.5 Comparison of throughput between optimized and un-optimized protocol with analytical model

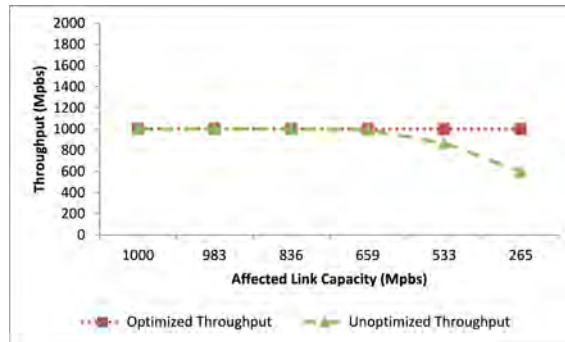
A software has been developed (source code given in Appendix B) in Python using PyXLL [50] based on the mathematical model developed in section 3.2.4 to support this work. The software was used to generate graphs to observe how the optimized protocol will perform in a wide variety of input conditions. A Microsoft Excel add-in called Daniel's Excel Toolbox [51] was used to export the graphs. The parameters used for the series of analysis with software are listed in Table 3.2. The graphs are given in Figure 3.8 through Figure 3.16.

Table 3.2: SIMULATION PARAMETERS

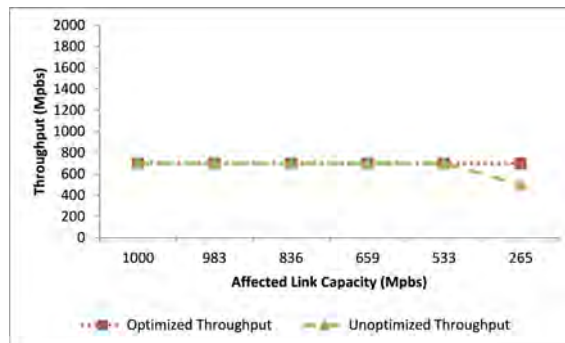
Parameter	Value/Range
Number of nodes	9
Number of links in ring	8
Link capacity	12Mbps-1Gbps
Microwave channel spacing	7MHz-56MHz
Modulation scheme	QPSK-256QAM
Node traffic	4.5Mbps-177.7Mbps



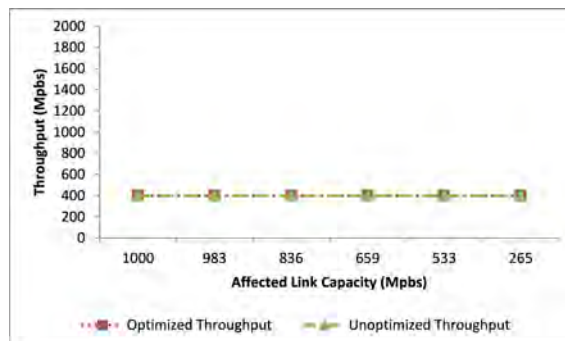
(a) Node Traffic = 177.7 Mbps



(b) Node Traffic = 111.1 Mbps

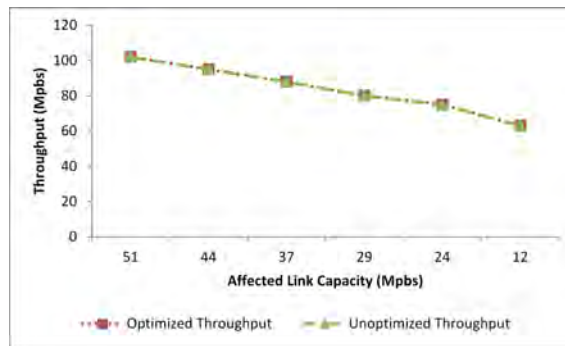


(c) Node Traffic = 77.7 Mbps

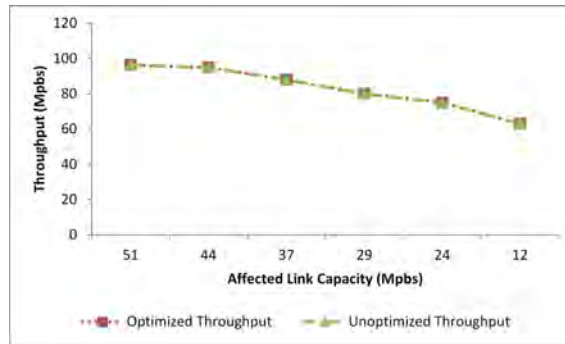


(d) Node Traffic = 44.4 Mbps

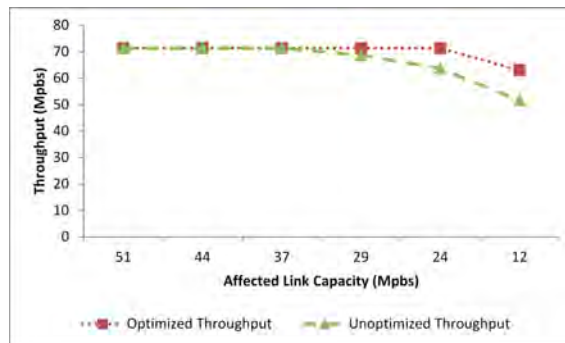
Figure 3.8: Throughput comparison (degraded link no: 1, RPL link no: 6, full capacity: 1Gbps)



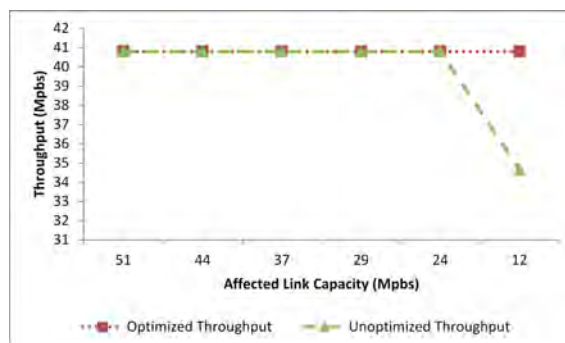
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

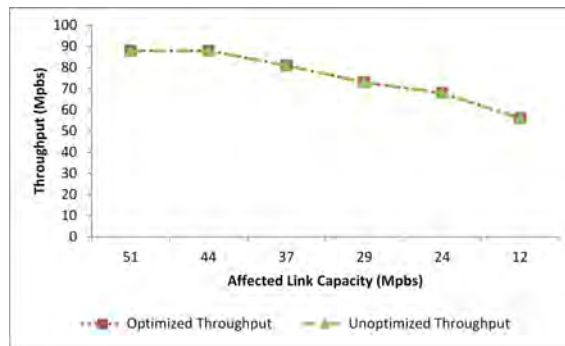


(c) Node Traffic = 7.9 Mbps

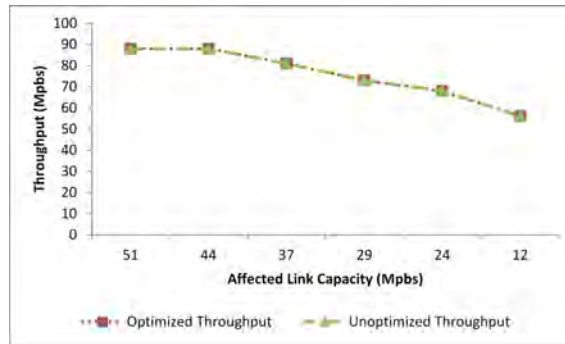


(d) Node Traffic = 4.5 Mbps

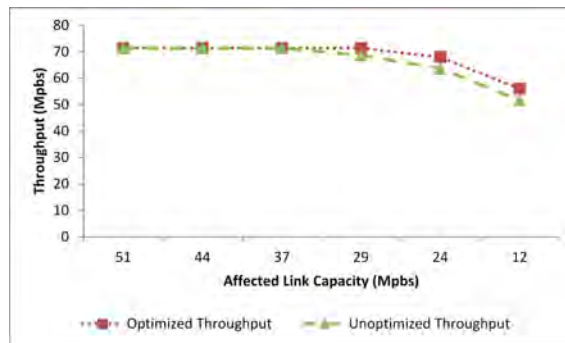
Figure 3.9: Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 51Mbps)



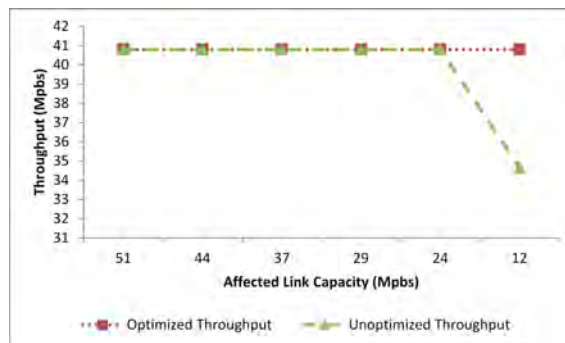
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

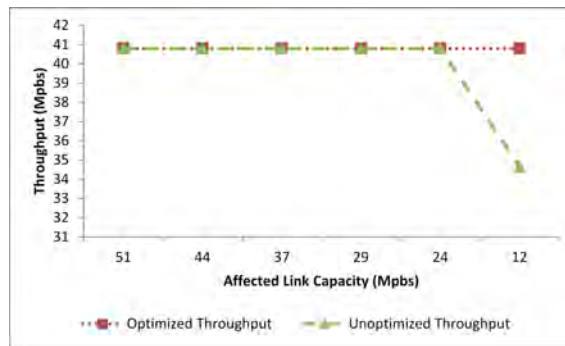


(c) Node Traffic = 7.9 Mbps

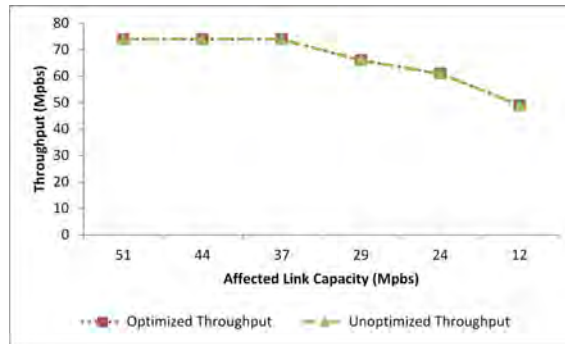


(d) Node Traffic = 4.5 Mbps

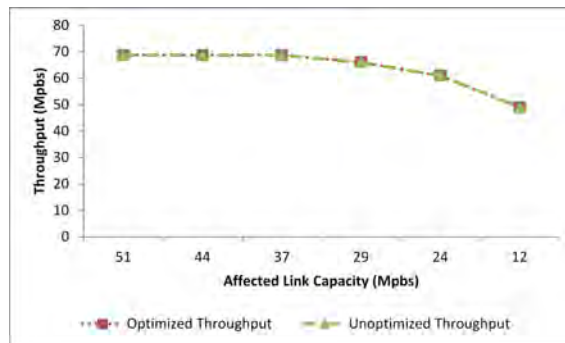
Figure 3.10: Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 44Mbps)



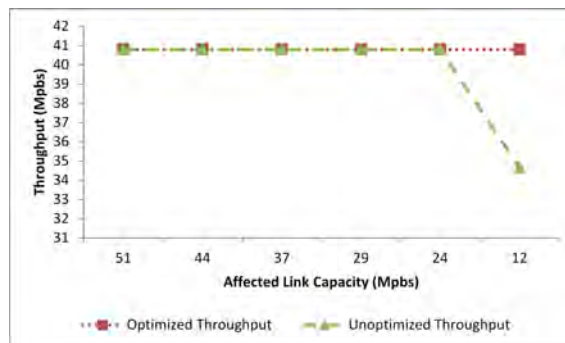
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

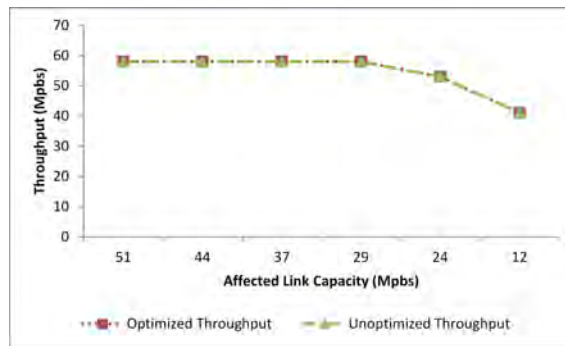


(c) Node Traffic = 7.9 Mbps

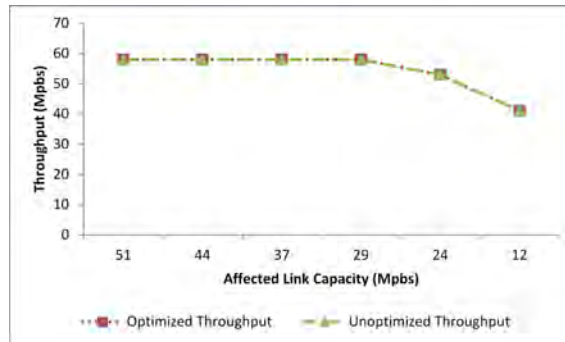


(d) Node Traffic = 4.5 Mbps

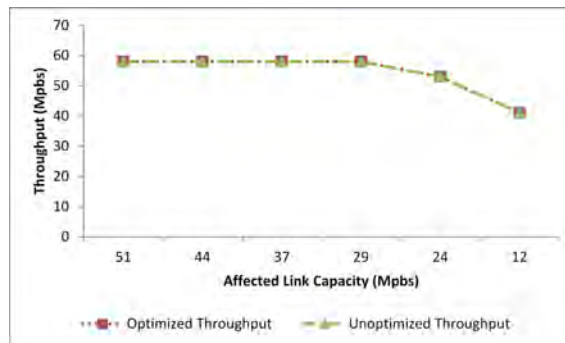
Figure 3.11: Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 37Mbps)



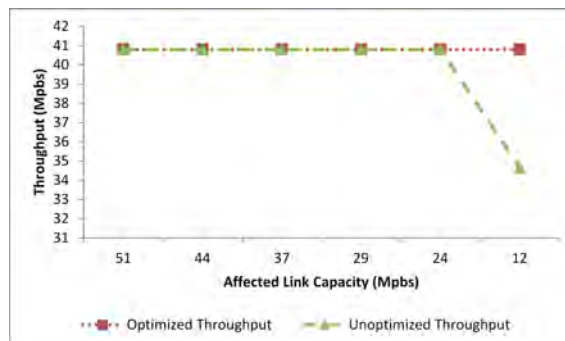
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

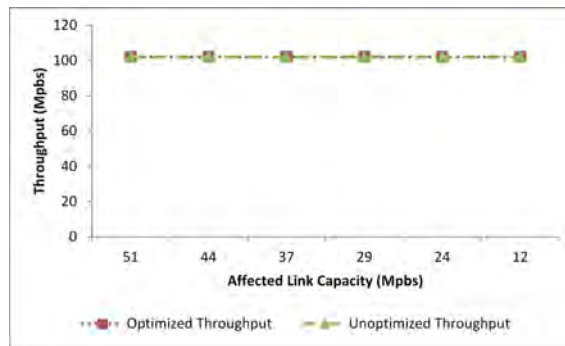


(c) Node Traffic = 7.9 Mbps

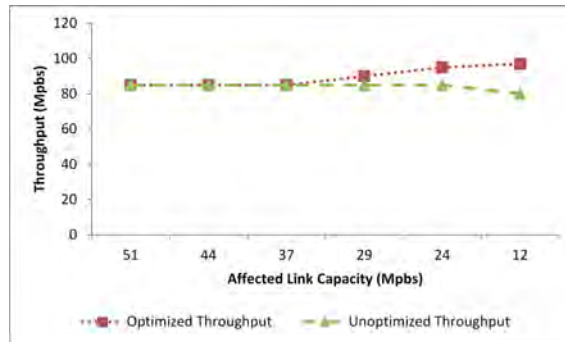


(d) Node Traffic = 4.5 Mbps

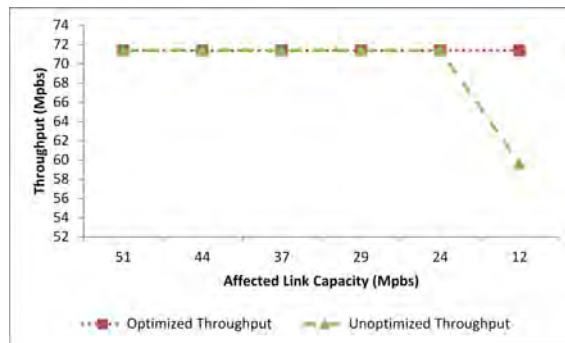
Figure 3.12: Throughput comparison (degraded link no: 1, RPL link no: 4, full capacity: 29Mbps)



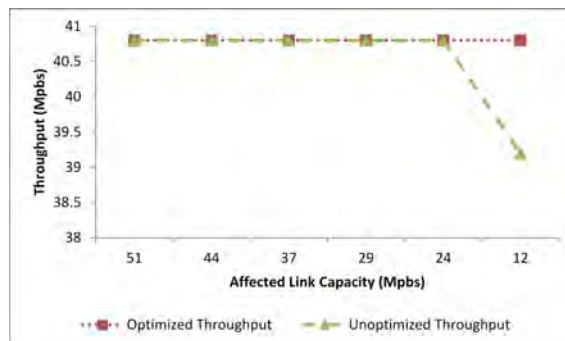
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

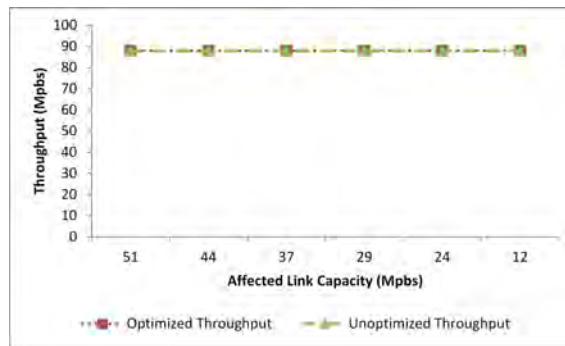


(c) Node Traffic = 7.9 Mbps

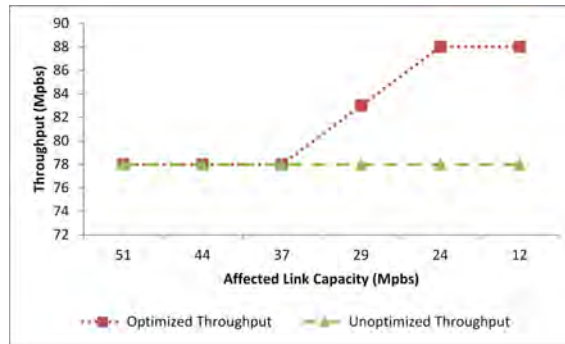


(d) Node Traffic = 4.5 Mbps

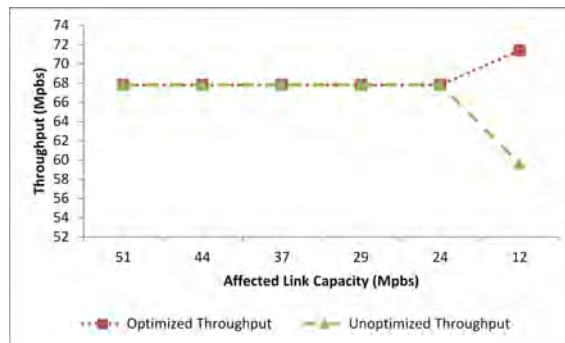
Figure 3.13: Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 51Mbps)



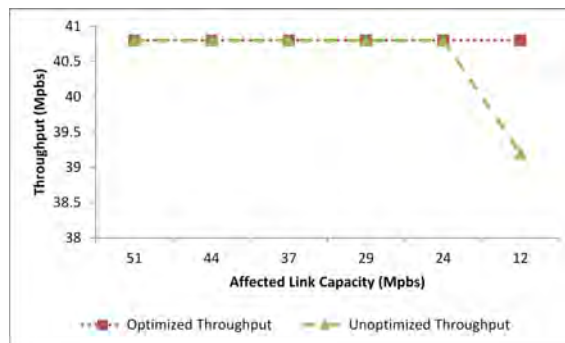
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

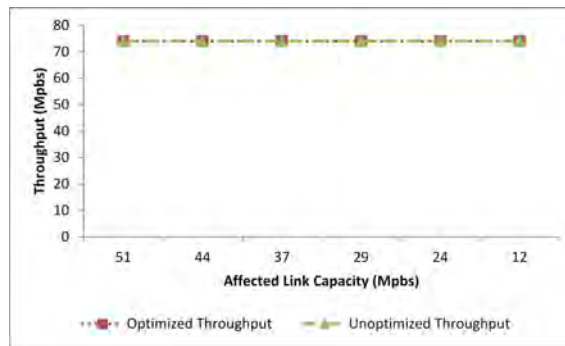


(c) Node Traffic = 7.9 Mbps

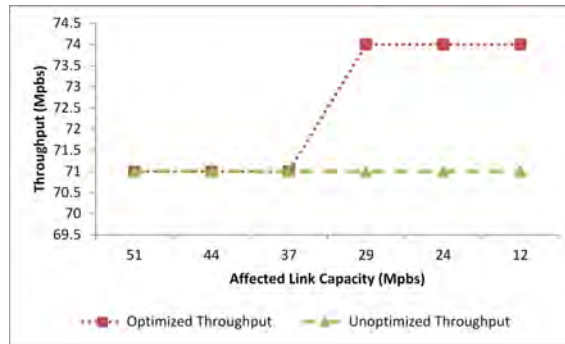


(d) Node Traffic = 4.5 Mbps

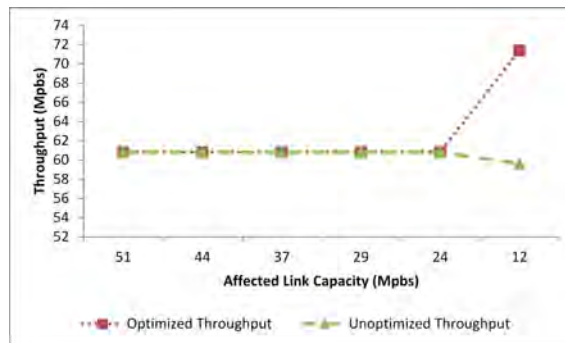
Figure 3.14: Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 44Mbps)



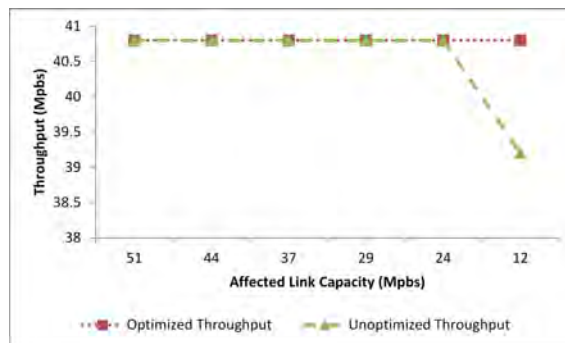
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps

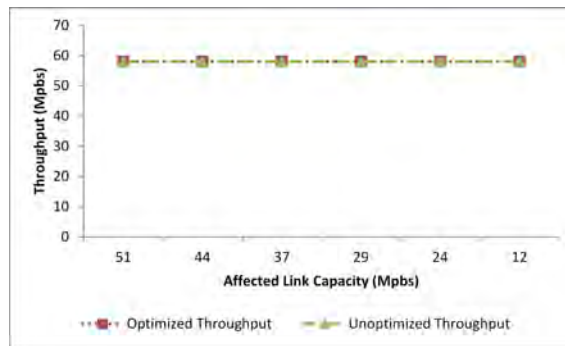


(c) Node Traffic = 7.9 Mbps

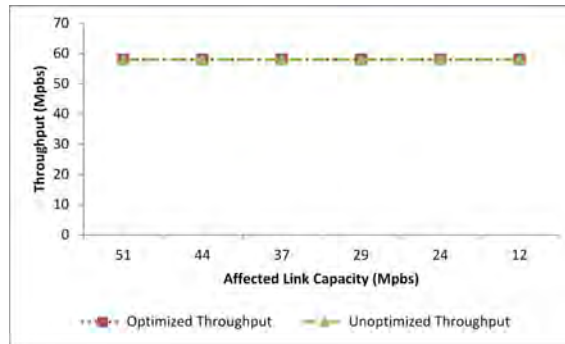


(d) Node Traffic = 4.5 Mbps

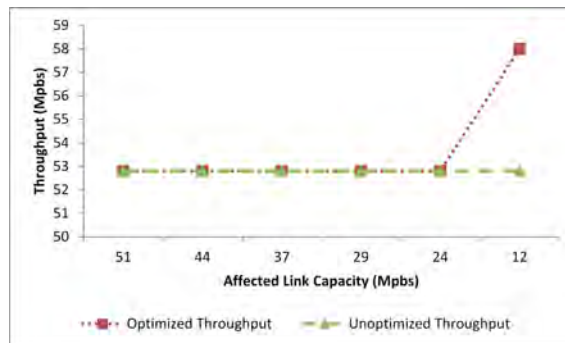
Figure 3.15: Throughput between (degraded link no: 3, RPL link no: 6, full capacity: 37Mbps)



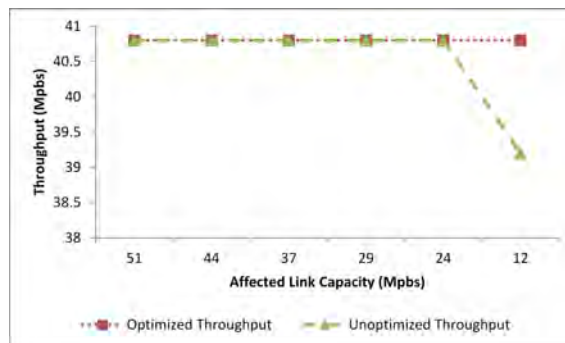
(a) Node Traffic = 18.1 Mbps



(b) Node Traffic = 11.3 Mbps



(c) Node Traffic = 7.9 Mbps



(d) Node Traffic = 4.5 Mbps

Figure 3.16: Throughput comparison (degraded link no: 3, RPL link no: 6, full capacity: 29Mbps)

3.4 Summary

In this paper, we propose an enhancement of ERPS protocol supporting adaptive modulation in an ERPS network to improve overall throughput. We present a principle for designing multiple instance ring network and show calculation of threshold values required by the proposed algorithm. In order to evaluate performance of the proposed algorithm we have designed our own simulation model in OMNET++. Using our simulation model we have shown that by switching instances from the degraded link, overall throughput can be improved.

We have also shown that our proposed algorithm always provides better performance in terms of throughput and packet loss. It is observed that the proposed protocol can achieve approximately 65% higher throughput compared to existing ITU-T G.8032 standard in severe signal degradation condition.

In this paper we have presented simulation results for a single physical ring network. Performance in physical multiple ring network should be further studied. Also notification of SD (Signal Degradation) event at microwave transceiver to ERPS should be investigated in further depth.

Chapter 4

FCSEA: A FLOODLESS CARRIER-GRADE SCALABLE ETHERNET ARCHITECTURE

4.1 Introduction

Ethernet is the most widely used and dominant technology for Local Area Network (LAN). The fast switching speed, cost effectiveness and simplicity of Ethernet makes it also desirable in other networks areas such as carrier networks. For instance, Ethernet switching is preferred to routing for high performance of LTE X2-interface [2], [3]. However, a simple Ethernet network is not scalable to serve a large number of nodes as required in telecommunication carrier networks. This is because of broadcast traffic caused by protocols like ARP (Address Resolution Protocol) and DHCP (Dynamic Host Configuration Protocol) running in an Ethernet network for service and resource discovery.

Currently, protection technologies in carrier Ethernet networks such as Ethernet Ring Protection Switching (ERPS), employ VLANs to keep broadcast within acceptable limit. While this approach requires complex VLAN planning and configuration for every switch, it serves the purpose as a replacement of point to point TDM technologies (e.g. SDH). But, in next generation networks, such as LTE, where nodes require connectivity with their peers, VLAN based segmentation will prohibit nodes under one VLAN to communicate with their neighbors in a different VLAN. FCSEA can operate in carrier networks to replace complex VLAN based scalability solutions while providing unrestricted peer to peer connectivity.

This work evolves from EtherProxy [12], a device for large enterprise networks that uses caching to suppress broadcast traffic. The distinct contributions of this paper are as below,

1. A proactive learning and prioritized cache retention policy to prevent initial flooding of entire network.
2. Carrier grade protection switching support.
3. Simulation and analysis of EtherProxy, SDN based ARP proxy and current work to evaluate broadcast suppression capability and query response time.

4.2 The Design of FCSEA

The FCSEA partitions a network into segments to protect a segment of the network from broadcast generated by other segments. Each FCSEA segment consists of a subset of hosts and one or more FCSEA enabled devices. This subset of hosts are said to be served by the FCSEA for this segment. Figure 4.1 shows an example topology where hosts H1.1 and H1.2 are served by FCSEA device U1 to form a segment (Segment 1). The topology in Figure 4.1 requires only one FCSEA enabled device per segment. This type of topology is unable to protect a network from link failures. For this reason ring topology is deployed in carrier networks to protect against link failures. In case of a ring topology, two FCSEA devices will be required as shown in Figure 4.11.

To briefly summarize, effective positioning of the protocol requires:

1. ability to intercept all traffic towards and from a subset of hosts,

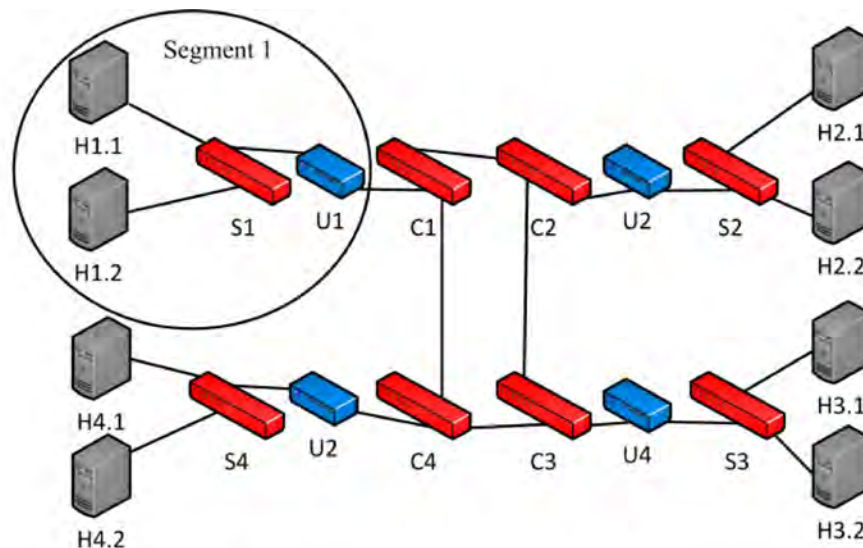


Figure 4.1: An example topology to illustrate the design of FCSEA

2. multiple devices in case of a ring or mesh topology for a subset of hosts.

4.2.1 Floodless ARP proxy

EtherProxy uses an on demand pull method to learn MAC addresses. This results in broadcast for the first time request of a target IP. Moreover, if the target host IP is not live in the network, broadcast will continue to happen. The FCSEA protocol proactively learns the logical (IP) address to hardware address (MAC) mapping before a request is made for the first time. Thus FCSEA can serve ARP requests without ever require flooding the hosts for their reply. If a target IP for an ARP request coming from outside the segment is not known, FCSEA ARP module drops the request. As ARP request to unknown target host IP is blocked, the FCSEA must have knowledge of all legitimate hosts inside the segment it is serving. This is achieved by proactive learning.

4.2.1.1 Proactive Learning of ARP Cache Entries

ARP cache entries are learned proactively in contrast with reactively broadcasting the segment on ARP request arrival. Following methods are followed to do this-

- Learning from DHCP assignment
- Learning from gratuitous ARP request/reply
- Learning from host initiated communication

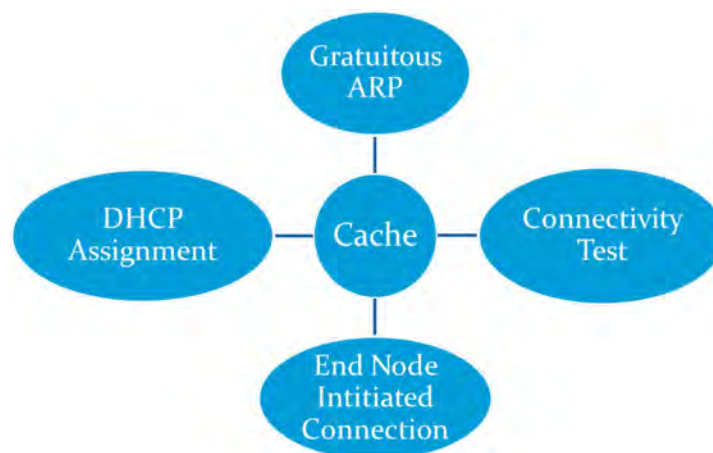


Figure 4.2: Proactive learning modes in FCSEA

A) Learning from DHCP assignment

FCSEA learns hardware addresses to IP address mapping by intercepting DHCP address assignments, inspired by [41], [42]. This is done in the ARP module. The ARP module intercepts DHCP messages and parses them to extract information. The ARP matches DHCPREQUEST message with respective DHCPACK message. DHCPREQUEST message is matched against CHADDR (Client Hardware Address) and SIADDR (Server IP Address) of DHCPACK. When a matching exchange of a DHCPREQUEST message and a subsequent DHCPACK message is found, ARP module creates an entry in the ARP cache. DHCPACK message is used as the source of information to build the ARP cache entry. Two fields are required to construct an ARP cache entry. These are host IP address and host MAC address. Host IP address is obtained from YIADDR (Your IP Address) of DHCPACK message and CHADDR field of DHCPACK message is used to obtain host MAC address.

The algorithm for populating ARP cache from DHCP packets is given in the flowchart in Figure 4.3.

B) Learning from gratuitous ARP request/reply

FCSEA listens for GARP (gratuitous ARP) request/reply to proactively learn ARP cache records. From a GARP request/reply SPA (Sender Protocol Address) and SHA (Sender Hardware Address) fields are extracted. These fields are then used to build the ARP cache entry. The host IP of the ARP cache entry is obtained from SPA and the host MAC address is found at SHA field of GARP request/reply.

The logic for learning ARP cache entries from GARP is illustrated in the diagram in Figure 4.4.

C) Learning from host initiated communication

It is possible that neither DHCP is enabled at the host nor GARP is used. In this scenario, in absence of DHCP and GARP, FCSEA still need to have knowledge of host IP and MAC address to serve ARP requests. As FCSEA drops any ARP request for which target is not known, it has to ensure enrollment of all legitimate hosts inside its serving segment even in absence of DHCP and GARP mechanisms of proactive learning. That is a third alternative method of learning ARP cache entries is required for a reliable proactive learning.

We have identified two possible ways for this third option of proactive learning.

1. Connectivity test
2. End node initiated connection to core NE
 - i. Connectivity Test

If DHCP is not used for IP address assignment then it can be assumed that the host IP address is commissioned manually. This is usually the case in carrier networks where a new node such as LTE eNodeB is commissioned manually. After commissioning, a PAT (Pre-Acceptance Test) is performed where to check the connectivity a ping test is performed. This connectivity test usually checks that the newly integrated node can reach one or more core network nodes. For example, a newly integrated IP BSC can be tested for reachability of a Media Gateway (MgW) from the IP BSC. Thus this test will send packets through the FCSEA enabled

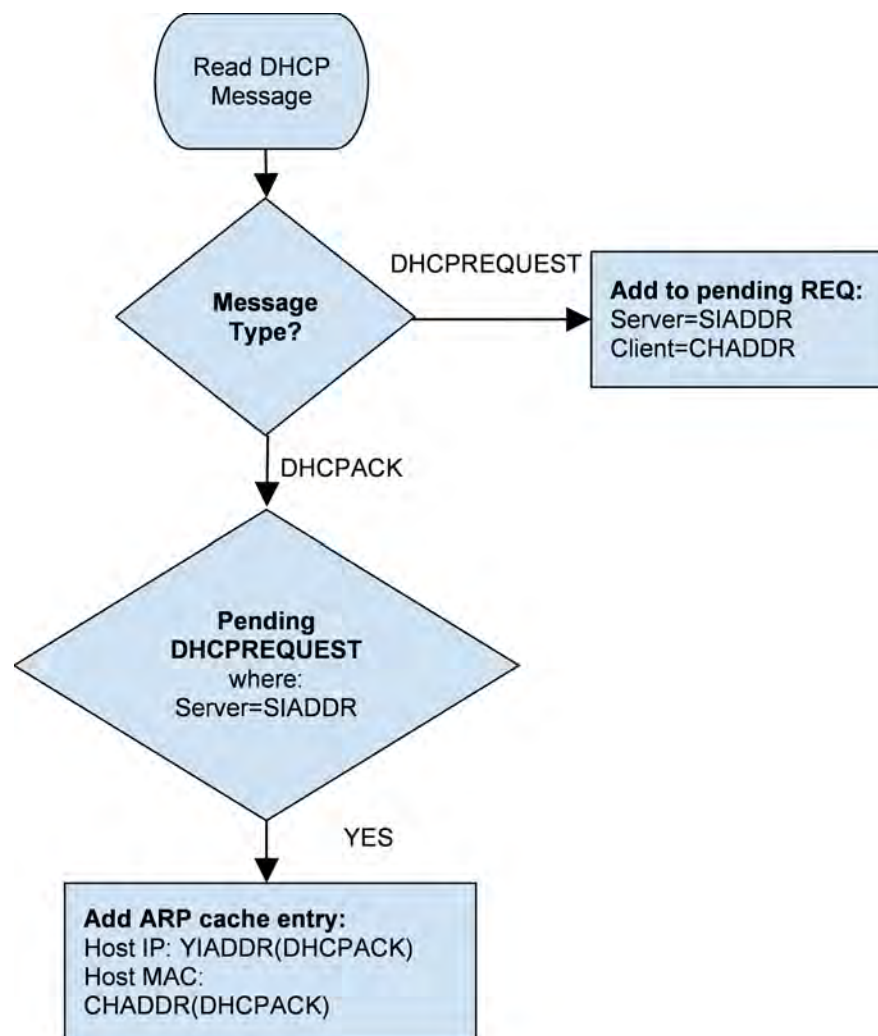


Figure 4.3: Proactive ARP cache entry learning from DHCP

device to its gateway or next hop. Before the new node can send packets, it will perform an ARP request to learn the hardware address of the gateway or next hop. FCSEA uses the ARP request generated by this ping test to learn the new host.

Learning A2A mapping from connectivity test is illustrated in Figure 4.5. In Figure 4.5, a new node (BSC) B1 is integrated in the network at switch S1. FCSEA switch U doesn't have any knowledge of B1 at this point. Let us consider that, DHCP is not enabled at B1 and it is not configured not to generate any gratuitous ARP

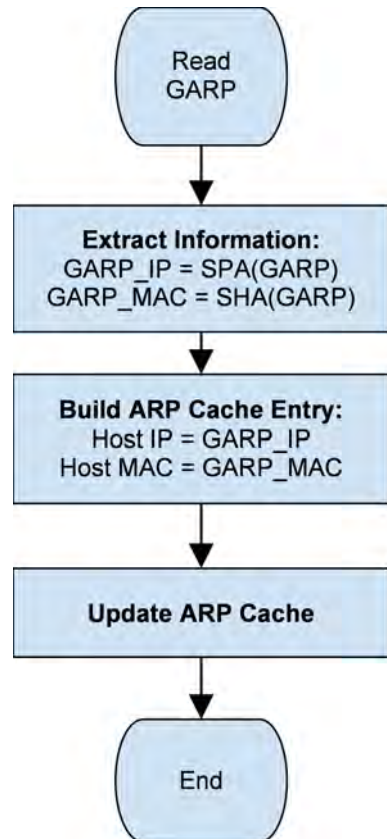


Figure 4.4: Learning ARP cache table from gratuitous ARP

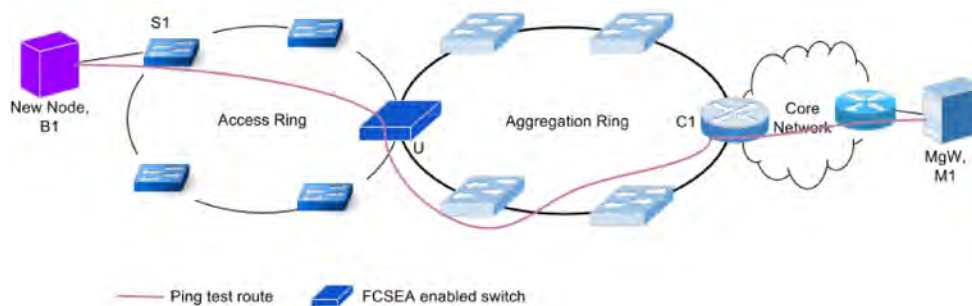


Figure 4.5: Connectivity test of a new node (BSC) with a core network NE (Media Gateway) after network integration during PAT

request/reply. B1 is configured its IP address manually. After the configuration, it issues a ping command with the IP address of M1 to test its connectivity. This ping command will lead to the following steps:

1. B1 will generate ARP request to find MAC address of its gateway C1.
2. U will receive the request from B1 with destination MAC broadcast address.
3. U will learn B1's A2A mapping from the ARP request and forward the L2 frame with C1's MAC as destination.
4. B1 will send ping request (SNMP) packet to M1 with layer-2 frame destination address of C1.
5. Ping request packet will be routed through core network to M1.
6. M1 will reply to the ping request.
7. M1's reply reach router C1.
8. C1 forwards the reply to B1 in layer-2 frame with destination MAC of B1.
9. B1 receives the reply and this is shown in the terminal to confirm the connectivity.

In step 3 above, FCSEA switch U learns the MAC to IP mapping of B1 and updates its ARP cache. Thus FCSEA learns A2A mapping in absence of DHCP and GARP.

ii. End node initiated connection to core NE

This option explores the role of a network element in the network to register it in the FCSEA ARP cache. Access Network devices need to establish bearer interface with core NEs. For example in LTE, eNodeB will establish S1 interface with MME (Mobility Management Entity) and SGW (Serving Gateway). This is done as a part of NAS (Non-Access-Stratum) attach procedure of UE (User Equipment) which is required for IP connectivity of the UE in the PDN (Packet Data Network). To establish S1 interface with MME, eNodeB will send a S1 SETUP REQUEST to MME [S1 Setup Procedure]. We utilize this signaling procedure to enroll the host (e.g. eNodeB) in ARP cache by intercepting the communication between the host (e.g. eNodeB) and the core network server (e.g. MME/SGW).

Figure 4.6 shows protocol signalling procedure to setup LTE S1 interface between eNodeB and MME where FCSEA enabled device intercepts the communication.

A network topology is illustrated in Figure 4.7 to show packet traveling path for establishment of LTE S1 interface between eNodeB and core network (MME). The FCSEA enabled device will intercept this communication (Ethernet header) to learn about the eNodeB.

In Figure 4.7 a new node (eNodeB E1) is added to the access ring (segment) and FCSEA device U1 is serving this segment. This can be new rollout or an eNodeB coming up in the network after maintenance work. This new eNodeB will now participate in the following steps:

1. E1 will broadcast connectivity information over radio interface to all UE (User Equipment) in the cell.
2. UEs can now attach to the network by eNodeB E1.
3. E1 will initiate S1-interface setup procedure.
4. ARP request will be generated as part of step 3.
5. U1 will register eNodeB E1 from the request in step 4.

4.2.1.2 Priority Cache Retention and Learning

Remembering logical address to hardware address mappings, referred to as A2A (Address to Address) mappings, is the key to FCSEA protocol operation. FCSEA

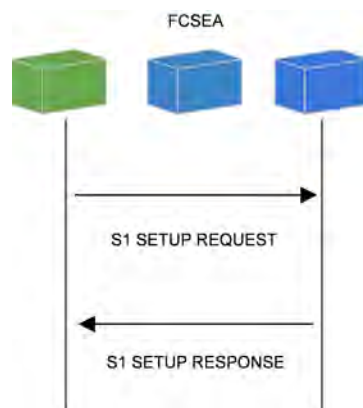


Figure 4.6: FCSEA switch learning eNodeB (host) from protocol signaling of LTE S1 interface setup

needs the logical-address-to-hardware-address mapping to reply to ARP queries, to forward broadcast message to unicast destination and to be able to drop a request for an unknown ARP target. However, as the number of nodes in the network increases, the number of entries in the ARP cache continues to increase. The resulting memory requirement to store all the entries in the ARP cache also increases. This rise in memory size may lead to a cache size that is impractical and eventually may impose scalability limitation on the number of NEs an FCSEA network can support. We propose a priority based cache retention and learning strategy to address this issue. This strategy defines what A2A mappings can be learnt, when these entries can be learnt and which of the learnt mappings can be overwritten to keep the overall memory requirement within feasible limit.

A) Classification of A2A Mappings

This work classifies A2A mappings stored as ARP cache entries into two broad types, each of these types are further classified into two sub-types. This is shown in Figure 4.8.

i. Internal Mapping

Internal mapping is mapping learned by ARP requests arriving on internal port. To clarify, in Figure 4.9, the port of U1 toward S1 is the internal port of U1.

Internal and External ports are illustrated in Figure 4.9-

To further illustrate the concept of internal and external ports or A2A mappings, ARP Cache of FCSEA enabled switch U1 from Figure 4.9 is shown in Table 4.1. In Figure 4.9 ports numbered 1 and 2 are internal ports whereas 3 and 4 numbered

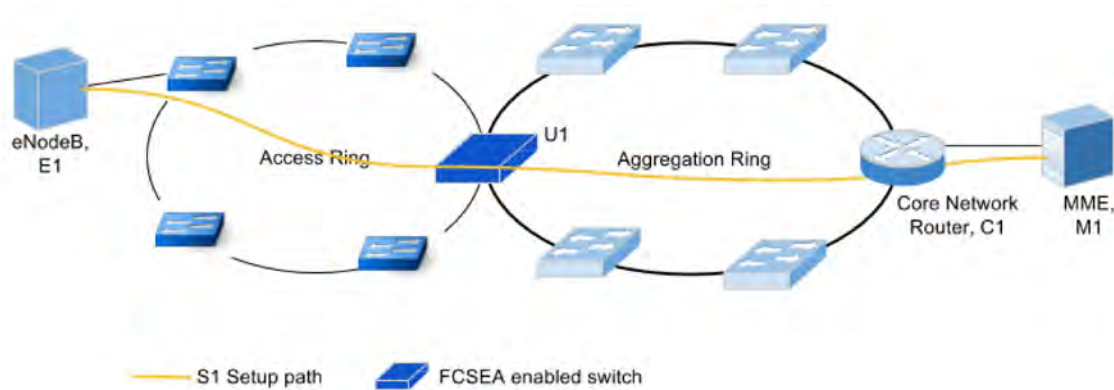


Figure 4.7: Utilizing LTE signaling procedure to register an LTE eNodeB in the ARP cache of FCSEA

Table 4.1: ARP CACHE OF PROPOSED PROXY U1

Entry ID	IP Address	MAC Address	Port	ARP Type	Entry Type
1	H1_IP	H1_MAC	1	Request	Internal
2	H6_IP	H6_MAC	4	Request	External
3	H7_IP	H7_MAC	4	Response	External

ports are external ports. Entry ID 1 is learnt from ARP request on port 1 which is an internal port therefore this entry is an “internal request” type. Similarly, Entries with ID 2 and 3 are learnt on external port number 4. Therefore both are “external” entries.

ii. External Mapping

External mappings are those learned from ARP requests and replies arriving on the external port (in Figure 4.9, the port of U1 towards C1).

B) Prioritizing A2A Mappings

Priority cache retention and learning principle processes the A2A Mappings

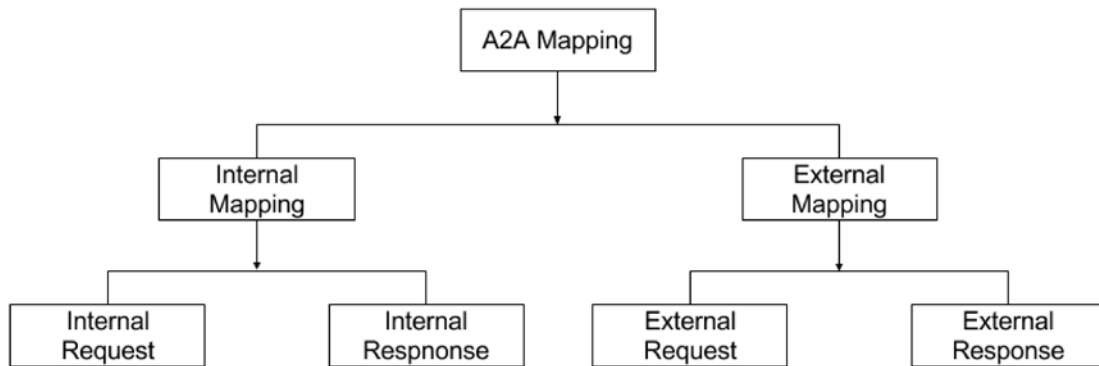


Figure 4.8: Classification of A2A (Address to Address) Mappings

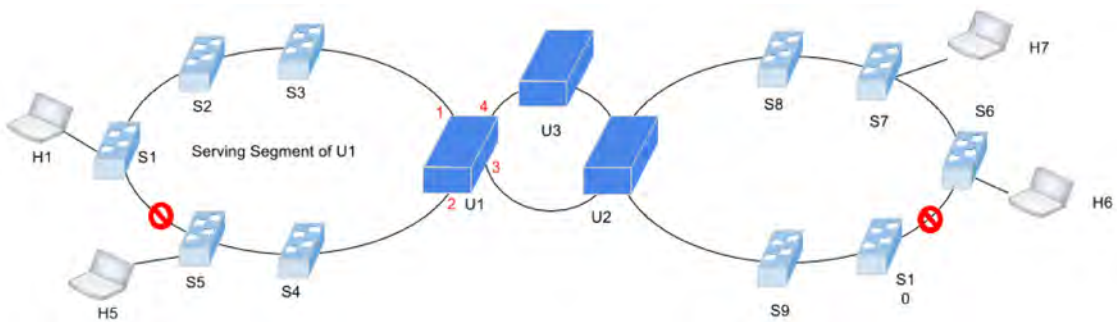


Figure 4.9: Internal and external ports and A2A mappings

based on their type. Internal mappings are assigned highest priority and never overwritten. Whereas, external ARP requests are not learnt if the cache is already full. LRU replacement policy is implemented similar but unlike EtherProxy, only on external response mappings (External RESP).

4.2.2 Protection switching consideration

In order to ensure high service availability, topology with redundant links is deployed in Ethernet carrier networks. A protection switching protocol then ensures a loop free logical topology. FCSEA devices are designed to ensure flawless protection switchover in a ring topology.

4.2.2.1 Protection switching in ERPS network

ERPS, as discussed in section 2.2.3, is a protection switching mechanism where multiple rings interconnect to span the entire network. One ring in ERPS is considered as the main ring while rest of the rings are called sub-rings. A sub-ring connects the main ring or other sub-rings at two interconnection nodes. In Fig. 2 C1-C2-C3-C4 is the main ring and C1-S1-S2-S4-S3-C2 is a sub-ring. Every switch in Fig. 2 is connected to one host, not shown in the figure for simplicity. The host connected to S1 is called H1 and so on. C1 and C2 are interconnection nodes. In the terminology of FCSEA, we can consider each sub-ring as a segment and position FCSEA devices (U1 and U2 in Fig. 2) at interconnection nodes to separate

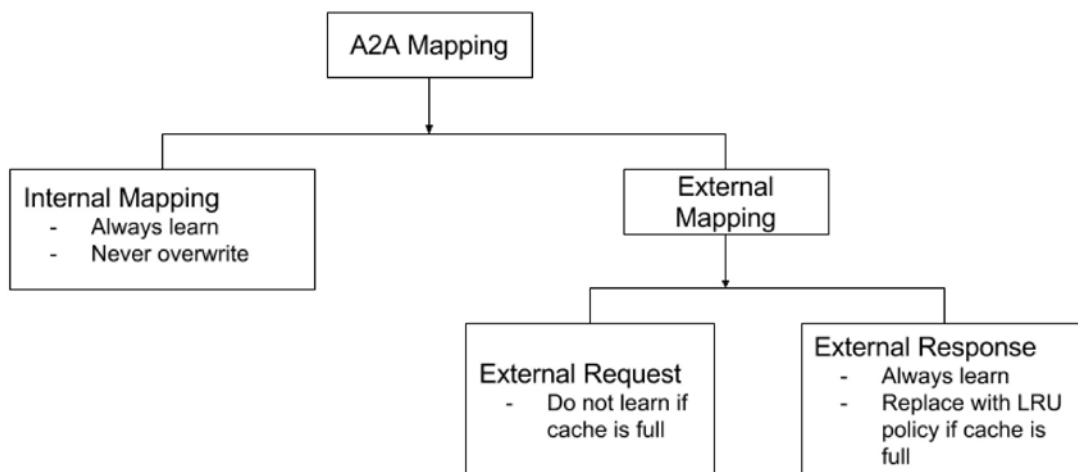


Figure 4.10: Prioritization of A2A (Address to Address) Mappings

a sub-ring from the main ring or to separate one sub-ring from another. Now to ensure a loop free topology ERPS will block one link in the segment (S2-S4) and thus create two partitions in the segment. If a link fails in the segment, the blocked link will be unblocked thus the partitions will change. The segment is now divided at the failed link instead of the blocked link causing some hosts (H2) to migrate from one partition to the other partition. We assume that U1 and U2 is replaced by EtherProxies E1 and E2 respectively. EtherProxy (E2), at this condition, replies to ARP request by an idle host (H4) querying a migrant host (H2), if corresponding entry is in a valid status. The likelihood that the A2A mapping for target host is in a valid cache entry is very high because the DROP_TIMEOUT is a large value (i.e. measured in hours). The reason behind EtherProxy E2 (imagined at the position of U2) replying to ARP request by H4 (connected to S4) is that, the EtherProxy (E2) learns about the migrant host (H2) from the external interface before link failure. This causes switches (e.g. S4 and S3) to think that the migrant host (H2 which is connected with S2) is towards the EtherProxy E2. Future traffic from H4 will be forwarded towards E2 by S2. But H2 can not be reached in this direction, leading to packet loss. FCSEA solves this problem by using an additional ARP request which we call Reflective ARP Request (RAR).

4.2.2.2 Reflective ARP Request (RAR)

In FCSEA, we reflectively issue an ARP request for the original target on the internal interface and only reply to the original request if the reflective request is not answered

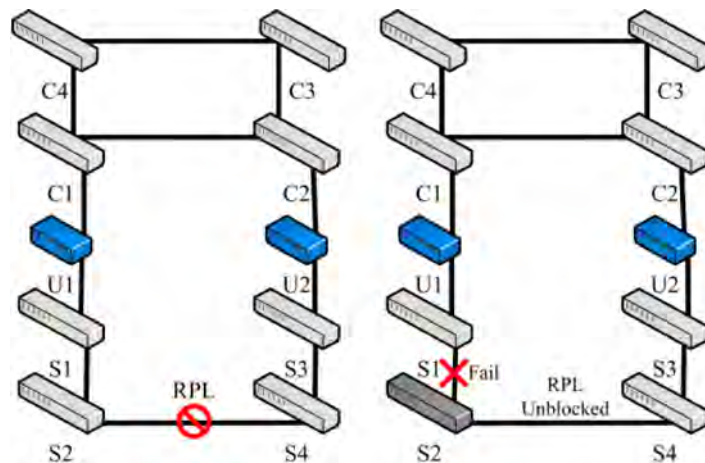


Figure 4.11: ERPS ring topology in normal (left) and failure (right)

within a timeout period which is configured as Reflective ARP Request Timeout (RARTout). If there is a reply to the reflective ARP request, then FCSEA drops the original request.

Figure 4.12 shows a segment with one FCSEA device U and five switches. Host R is requesting the hardware address of host T in an ARP request. U receives the ARP request from R at port number 1 (internal interface). Let's imagine that the ARP cache entry for host T was learned on the external interface of U and is concurrent (unused flag not set) and valid. Now U enters RAR process and issues an ARP request with its own hardware address as source address in the layer-2 frame. U replaces the destination hardware address of the ARP request which is a broadcast address with the MAC address of T (from cache). T will respond to both ARP requests (directly received from R and RAR received from U). R will get the physical address of T from ARP reply of T as they are in the same segment. After receiving response to the RAR from T, U will discard the ARP request from R knowing that T is in the same segment.

The Reflective ARP Request logic is illustrated in the flowchart in Figure 4.13.

The complete algorithm of FCSEA protocol is given in the flow chart of Figure 4.14.

4.2.2.3 Response time

As shown in Figure 4.15, let round-trip delay of each access ring, aggregation ring and backbone network link be t_{ACC} , t_{AGG} , and t_{BACK} respectively. Thus total



Figure 4.12: Traffic flow of Reflective ARP Request by FCSEA.

delay to cross access ring, $T_{ACC} = \sum t_{ACC}$. Similarly, delay in aggregation ring, $T_{AGG} = \sum t_{AGG}$ and delay in backbone network, $T_{BACK} = \sum t_{BACK}$. Two neighbors in a network may be in the same access ring or in different access rings. In the latter case aggregation network must be traveled. Therefore response time of FCSEA can be written as,

$$T_{FCSEA} = T_{ACC} + T_{AGG} + T_{ACC} = \sum T_{ACC} + T_{AGG} \quad (4.1)$$

or,

$$T_{FCSEA} = \sum \sum t_{ACC} + \sum t_{AGG} \quad (4.2)$$

For SDN based proxy FSDM, delay of backbone network is included to reach central controller. Therefore response time of FSDM can be written as,

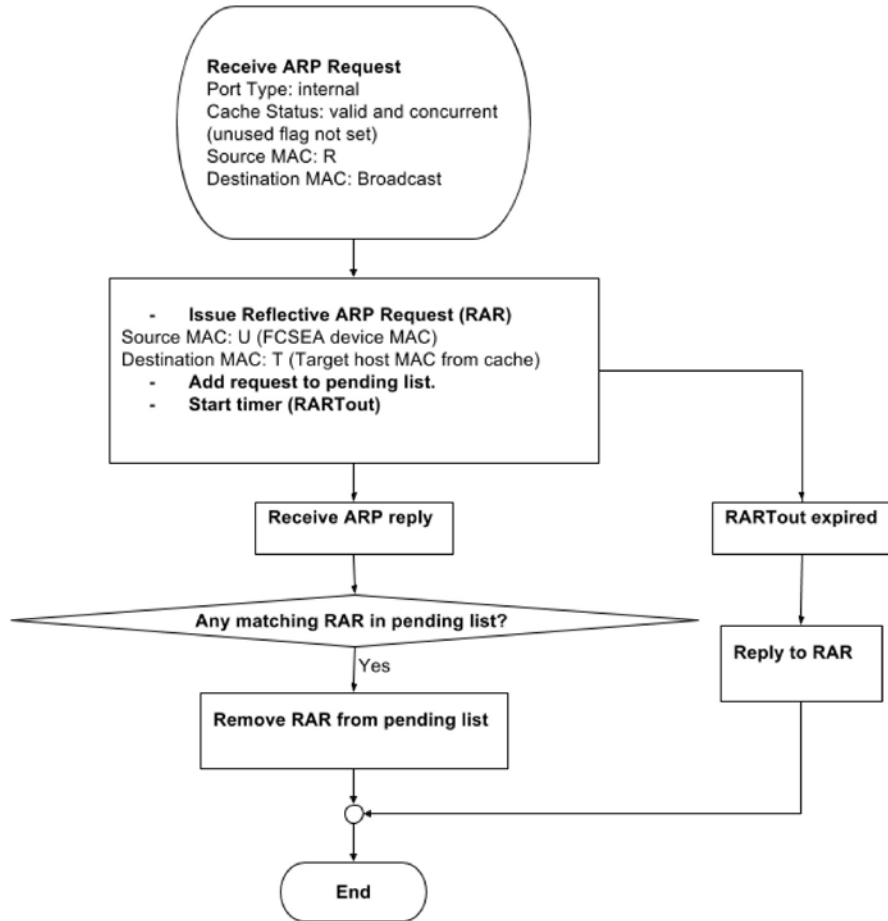


Figure 4.13: Reflective ARP Request (RAR) algorithm.

$$T_{FSDM} = T_{ACC} + T_{AGG} + T_{ACC} + \sum T_{BACK} \quad (4.3)$$

Or,

$$T_{FSDM} = \sum \sum t_{ACC} + \sum t_{AGG} + \sum \sum t_{BACK} \quad (4.4)$$

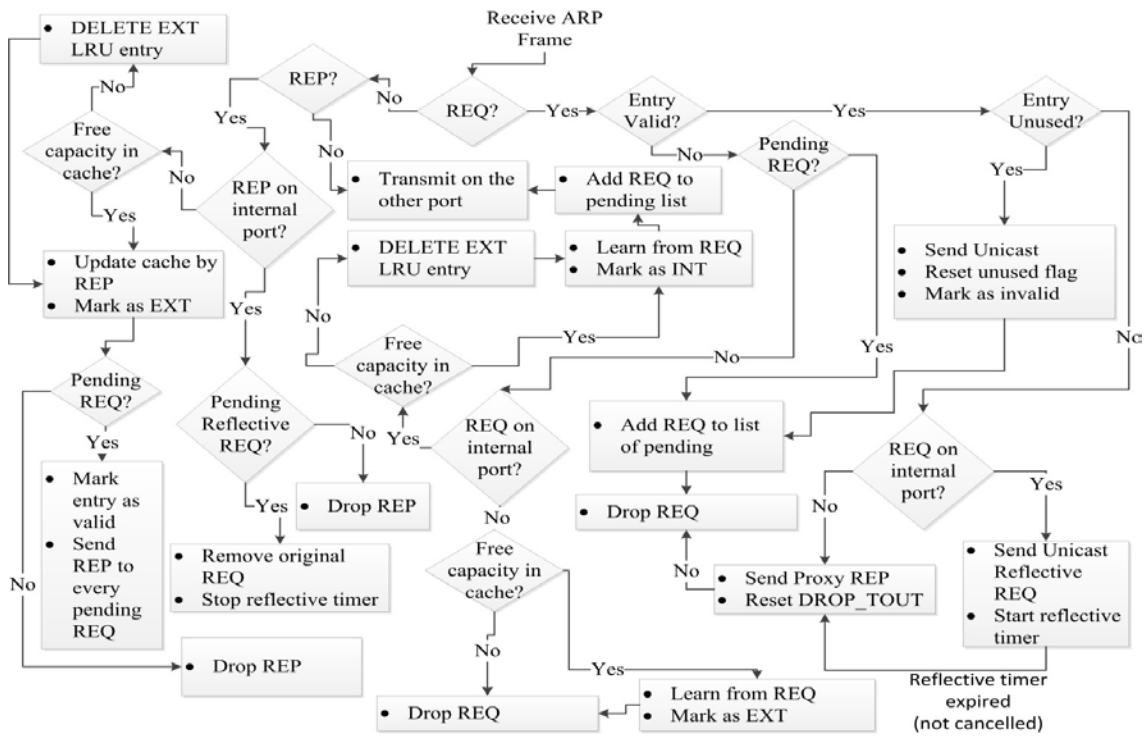


Figure 4.14: FCSEA ARP request and response processing logic (includes floodless ARP proxy, priority based cache retention and protection switching logic).

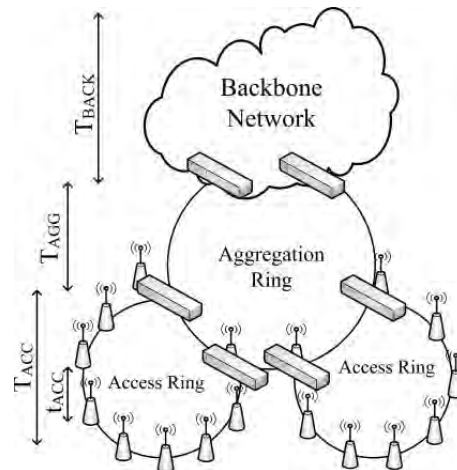


Figure 4.15: Telecommunication carrier network architecture.

4.3 Evaluation

In this section we evaluate the effectiveness of the FCSEA protocol using event based simulation. A simulation software was developed in C to evaluate performance of the proposed scheme and previous works. Simulation was performed with a request generation rate of 60 requests per minute while number of source and target hosts were varied widely (between 30 and 50,000).

4.3.1 Cache hit rate

In this experiment (Figure 4.16 to Figure 4.19) we evaluate cache hit rate over time to see effectiveness of the system to suppress broadcast traffic. We consider a successful hit if no host outside the segment from where ARP request is generated, receive the request. This way a hit is equivalent to a suppressed broadcast. The number of different hosts a requesting host can seek was varied which gave significant difference in resulting hit rate.

Results show that, hit rate of EtherProxy is decreased when-

1. Number of hosts increase
2. The number of target host per requesting host increase

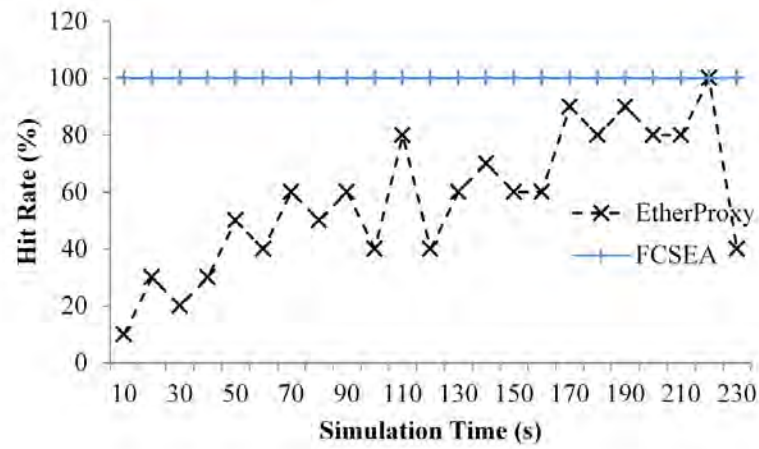
FCSEA always have 100% cache hit which suggest a floodless network with respect to hosts.

4.3.2 Cache size

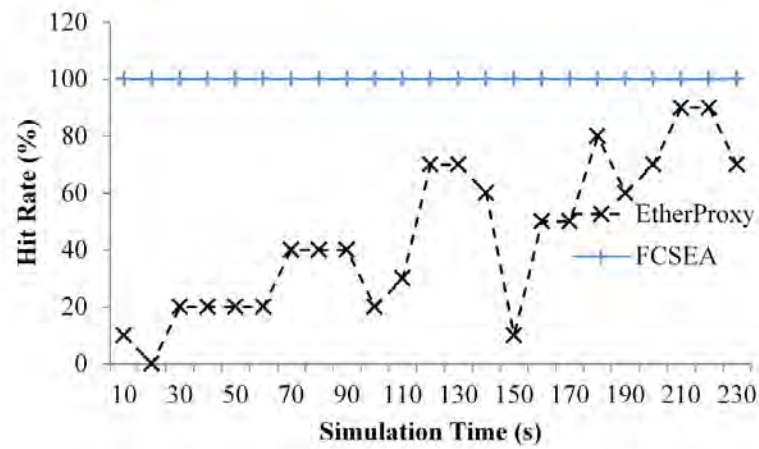
In this experiment (Figure 4.20), we observe characteristics of the mappings of different types: Internal, External (RESP) as defined in section 4.2.1.2 to confirm the priority cache retention and learning principle's limited memory requirement.

4.3.3 Response Time

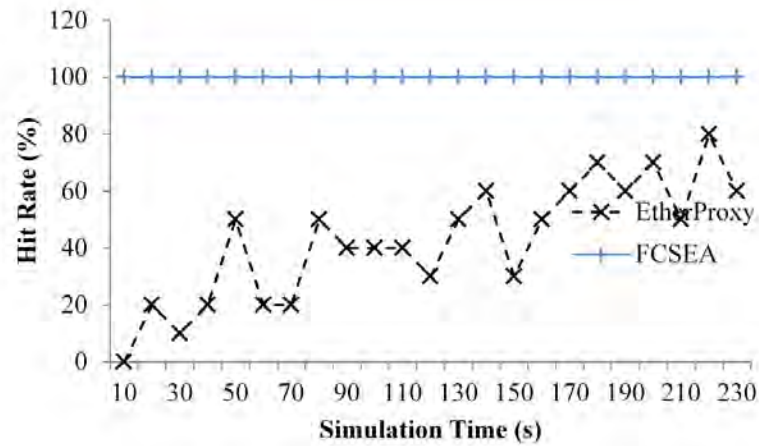
This experiment (Figure 4.21) measures the response time, that is, the delay between generation of an ARP request by a host and receiving the corresponding ARP reply. Delays shown in analysis in section 4.2.2.3 are individually measured for eight regions of a telecommunication carrier network. Backbone network delays found for different regions of the carrier network are listed in Table 4.2. These delays are



(a) 10 different hosts

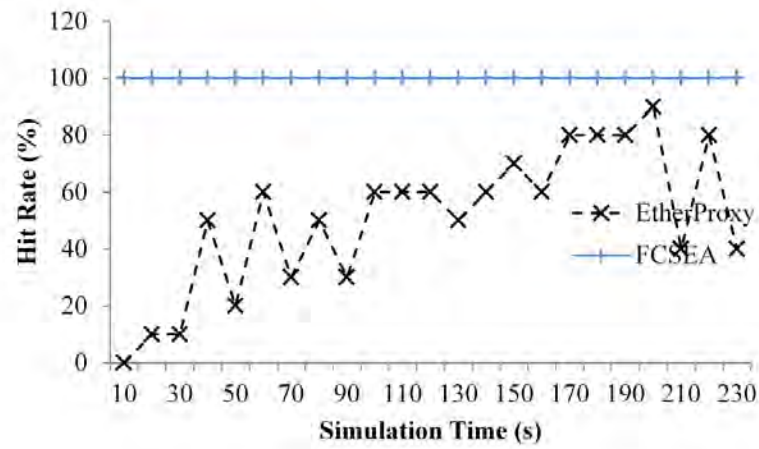


(b) 16 different hosts

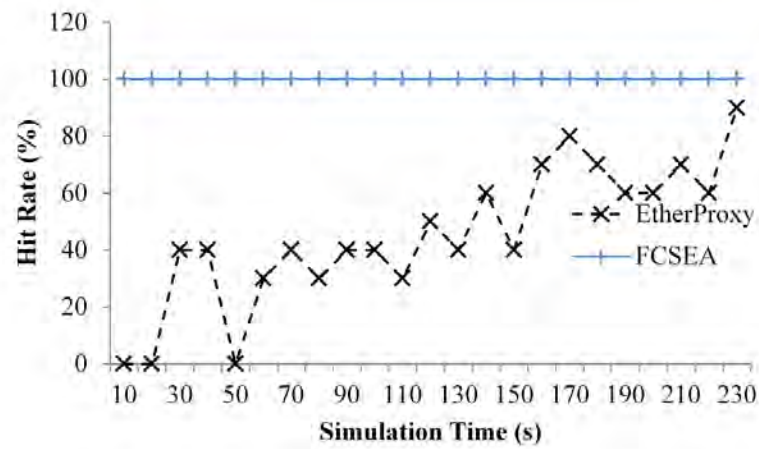


(c) 28 different hosts

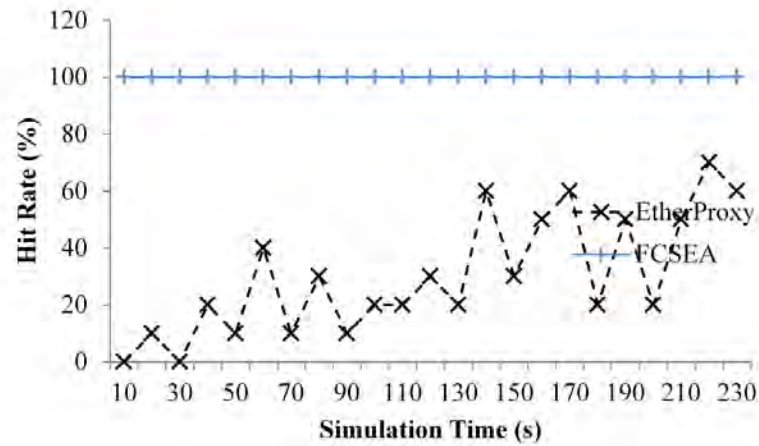
Figure 4.16: Cache hit rate comparison: source hosts = 30, target hosts = 1000



(a) 10 different hosts

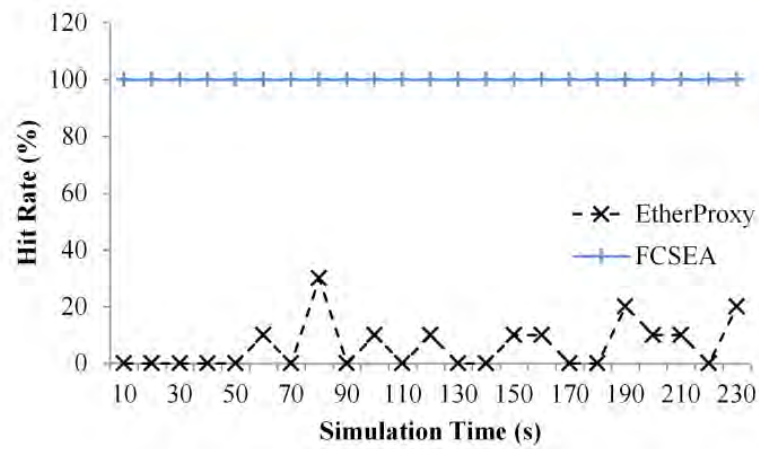


(b) 16 different hosts

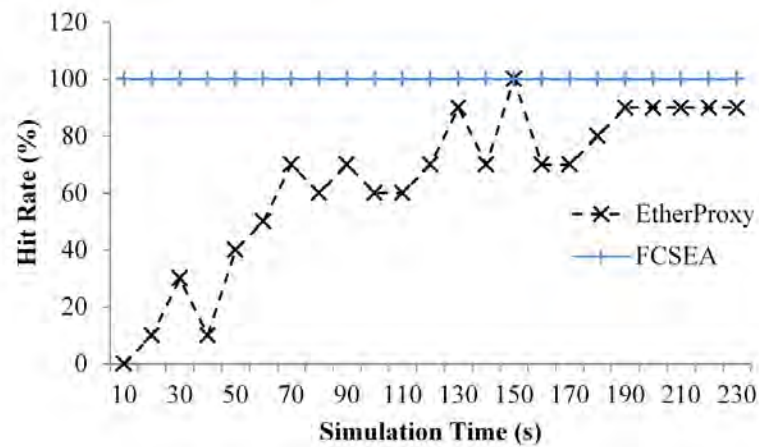


(c) 28 different hosts

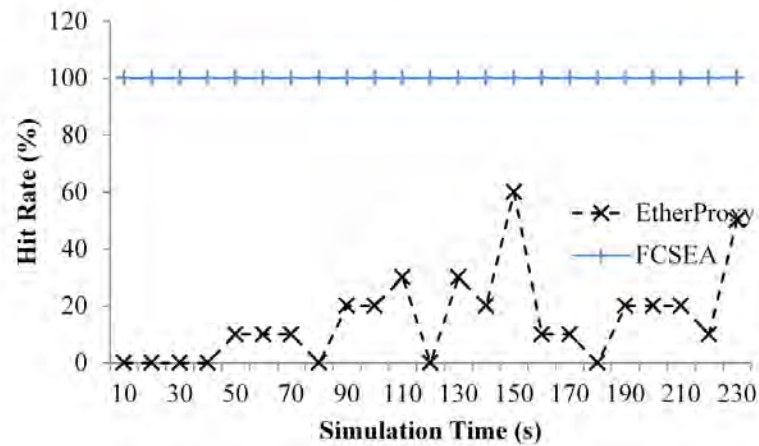
Figure 4.17: Cache hit rate comparison: source hosts = 30, target hosts = 10000



(a) 1 different hosts

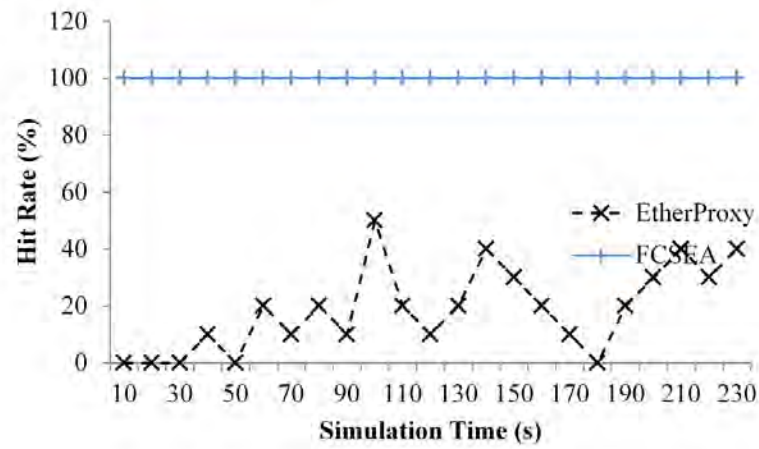


(b) 28 different hosts

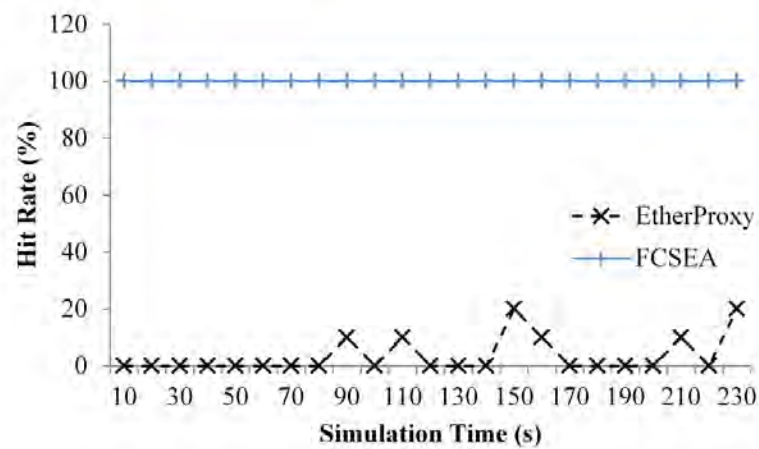


(c) 1000 different hosts

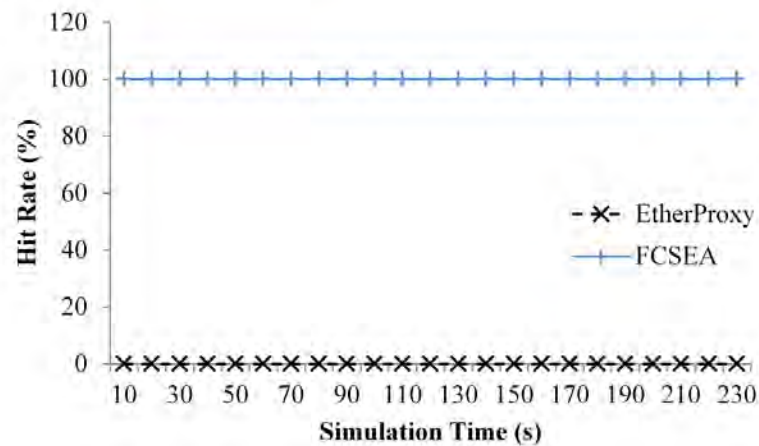
Figure 4.18: Cache hit rate comparison: source hosts = 500, target hosts = 100



(a) 1 host



(b) 28 different hosts



(c) 1000 different hosts

Figure 4.19: Cache hit rate comparison: source hosts = 500, target hosts = 50000

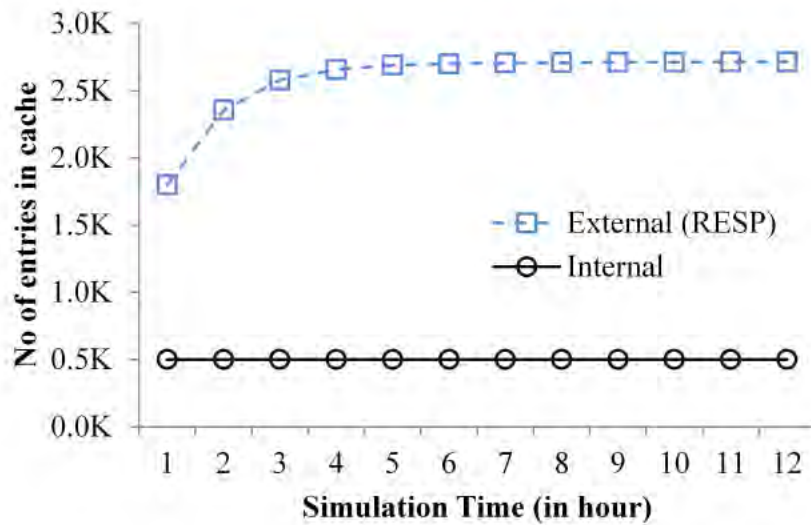


Figure 4.20: Cache size evaluation with 50.5K hosts (500 hosts sending ARP request for 50K hosts).

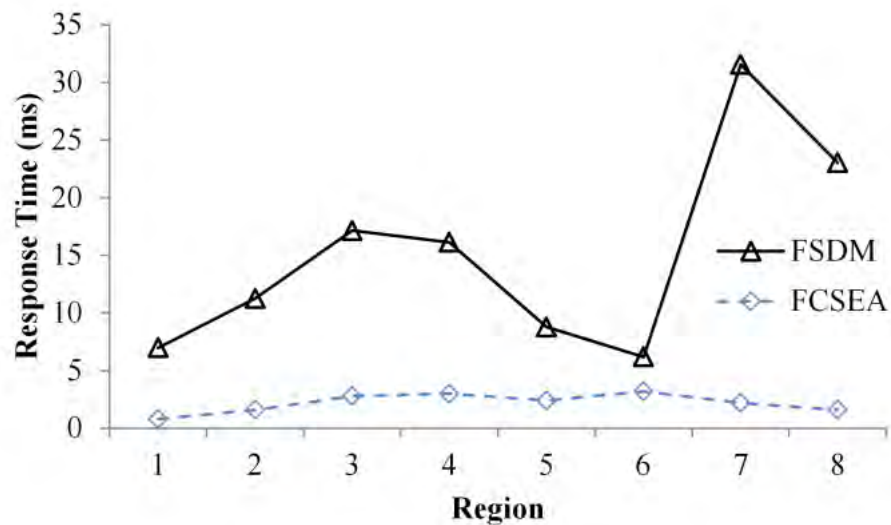


Figure 4.21: Response time of FCSEA to reply to ARP requests compared to SDN based proxy.

used as channel delays in the simulation. As we can see from Figure 4.21, FCSEA always have a lower delay compared to SDN based proxy FSDM (Floodless Service Discovery Mechanism). A sudden rise of delay in FSDM is noticed in region 7. This is because more backbone network links were involved during the simulation in this region.

Table 4.2: ROUND-TRIP DELAYS FOUND IN TELECOM BACKBONE NETWORK

Region	Delay (ms)
1	6.2
2	9.7
3	14.3
4	6.6
5	6.4
6	3
7	9.8
8	7.2

4.3.4 Reflective ARP Delay

In this experiment (Figure 4.22) we assess the delay caused by reflective ARP mechanism of FCSEA. Eight different experiments were performed by generating ARP requests between eight different randomly selected pairs of source and target hosts. Query response time using FCSEA and without using FCSEA is measured for same pair of hosts. Results show that reflective ARP doesn't introduce any significant delay overhead on FCSEA. As we can see in Figure 4.22, query response time of FCSEA (Reflective ARP) is higher than response time of no proxy in some experiments and vice versa. This is because when FCSEA device is closer to ARP source than target, reflective ARP require less time.

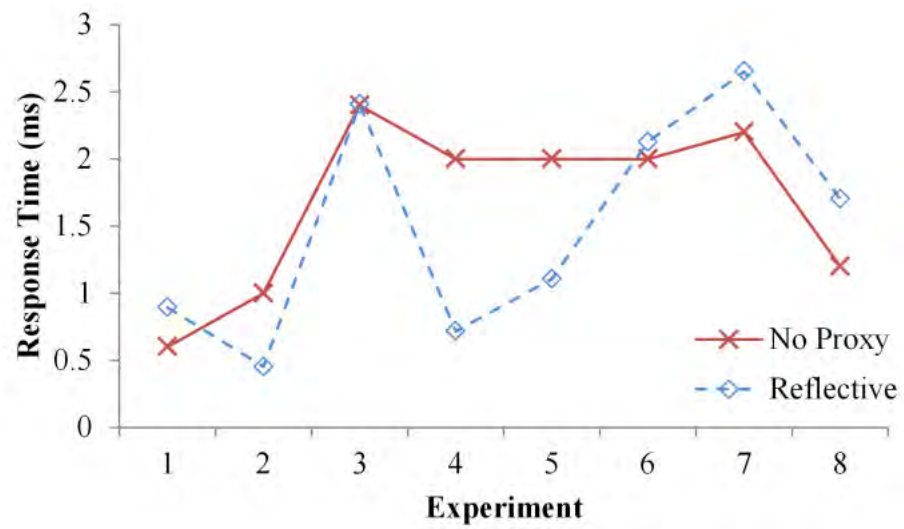


Figure 4.22: Performance evaluation of Reflective ARP mechanism.

4.4 Summary

In this paper, we propose FCSEA, a protocol to scale Ethernet to large carrier networks. Our work is based on EtherProxy, a protocol to scale Ethernet by suppressing broadcast traffic. With the aim to scale Ethernet for carrier network topologies, we improve EtherProxy in several ways. Firstly, we introduce a floodless proxy mechanism to overcome the initial-flooding limitation of EtherProxy. Secondly, we also employ a priority based caching mechanism to achieve a scalable memory requirement. Finally, we add support for redundancy protected networks for high service availability with a reflective ARP request mechanism. FCSEA is a simple protocol with limited memory requirement. Therefore it can be implemented as a software up-gradation on existing Ethernet hardware. It can also be implemented as a standalone device. Simulation results show that our proposed protocol can effectively ensure a floodless environment for end nodes. We also show that FCSEA outperforms SDN based broadcast suppression protocols due to its fast query response time.

Chapter 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

In this thesis, two significant improvements on Ethernet is proposed to make it more efficient and scalable. Firstly, we have introduced the concept of *Channel Adaptability* and proposed an optimization of the ITU-T G.8032 protocol to make Ethernet adaptable to channel capacity changes. Secondly, we have proposed an Ethernet architecture that will increase the scalability of Ethernet to meet the requirement of carrier networks.

In the first improvement, to provision channel adaptability into Ethernet, carrier network architecture has been inspected. We realize that a major part of carrier network traffic is hub-and-spoke in nature. This nature of traffic is applied to develop a multi-instance ERPS ring design principle.

The multi-instance ring design principle developed in this research supports energy-efficient instance re-routing to achieve channel adaptability with help of the modification on ERPS protocol.

The ring topology was analyzed and a mathematical model was formulated to estimate aggregate traffic in a ring network under link blockage (to prevent loop) and subject to capacity degradation (due to adaptive modulation events). This mathematical model will be equally applicable to ERPS network or any other protection switching protocol designed for ring topology.

A simulation model was developed for the ERPS protocol in OMNET++. The simulation model was used to observe performance of ring networks under link failure and bandwidth degradation. The developed simulation model has been then modified to according to the improvement of ERPS protocol proposed in this re-

search. We have then used the simulation model for existing and proposed ITU-T G.8032 protocol to compare performance in terms criteria including throughput and packet loss.

In the second improvement, to enhance the scalability existing literature in the field was reviewed. It was found that in order to increase the scalability of Ethernet some have proposed complete new layer-2 technology while others have modified Ethernet in a way that is not compatible with existing Ethernet devices. Some recent works in this fields suggests use of software defined network (SDN) technology to address the issue of scalability. This SDN based approach may incur significant delay in connection establishment. With the objective of leveraging the economies of scale of existing Ethernet technology and to ensure compatibility with already deployed traditional devices, we have based our work on a previous research EtherProxy. This research improves EtherProxy to make it deployable in carrier ring topology. It also improves EtherProxy in terms of ability to suppress broadcast. Thus our scalability solution is able to provide carrier grade services while maintaining compatibility with and leveraging economies of scale of traditional Ethernet technology.

With simulation, mathematical analysis and calculation with real data it is shown that the proposed protocols effectively increase performance of Ethernet compared to previous works in the field.

5.2 Future Work

To improve the ERPS protocol, it was assumed that microwave radio channel condition will be available to ERPS processes. The optimum way to inform ERPS processes of the adaptive modulation change events if to be further investigated.

In this work, focus was given on suppressing broadcast to protect end nodes from having to process too many requests by thousands of nodes in the network. However, broadcast will still be generated and will be passed to the network backbone. The amount of load imposed on the backbone by these broadcasts is to be further studied. Future research works are required to determine the impact of broadcast that still persists on the backbone network.

From the simulation results it is observed that performance of the network can be significantly improved not only by adapting to channel capacity but also to dynamic

utilization. This area can be explored to further improve the performance of ring transmission topologies.

Bibliography

- [1] S. Sharma and T.-c. Chiueh, “Programmable Ethernet switches and their applications”, *ArXiv preprint cs/0411019*, 2004.
- [2] “Backhauling X2”, *Cambridge Broadband Networks Limited white paper*, 2011.
- [3] M. Howard, “Using carrier Ethernet to backhaul LTE”, *Infonetics Research White Paper*, 2011.
- [4] A. Reid, P. Willis, I. Hawkins, and C. Bilton, “Carrier Ethernet”, *IEEE Communications Magazine*, vol. 46, no. 9, 2008.
- [5] J.-d. Ryoo, H. Long, Y. Yang, M. Holness, Z. Ahmad, and J. K. Rhee, “Ethernet ring protection for carrier ethernet networks”, *IEEE Communications Magazine*, vol. 46, no. 9, 2008.
- [6] “Microwave adaptive bandwidth feature: Make better use of available bandwidth”, *Cisco white paper*, 2013.
- [7] “Ethernet ring protection switching”, *ITU-T Rec. G.8032/Y.1344*, 2008.
- [8] “Ethernet ring protection switching”, *ITU-T Rec. G.8032/Y.1344*, 2012.
- [9] “Ethernet ring protection switching”, *ITU-T Rec. G.8032/Y.1344*, 2015.
- [10] “Evolving microwave mobile backhaul for next-generation networks”, *NEC white paper*, 2008.
- [11] “Evolving microwave mobile backhaul for next-generation networks”, *Cisco white paper*, 2008.
- [12] K. Elmeleegy and A. L. Cox, “Etherproxy: Scaling ethernet by suppressing broadcast traffic”, *INFOCOM*, pp. 1584–1592, 2009.
- [13] “Packet switching”, *EComputerNotes*, [Online]. Available: <http://ecomputernotes.com> (Last accessed on: 6 Dec 2016).

- [14] “Ethernet”, *Cisco support website*, [Online]. Available: <http://www.cisco.com> (Last accessed on: 7 Dec 2016).
- [15] “Provider bridges”, *IEEE Standard 802.1ad*, 2006.
- [16] “Provider backbone bridges”, *IEEE Standard 802.1ah*, 2008.
- [17] “Connectivity fault management”, *IEEE Standard 802.1ag*, 2007.
- [18] “Timing characteristics of a synchronous Ethernet equipment slave clock”, *ITU-T Rec. G.8262/Y.1362*, 2016.
- [19] J. Eidson and K. Lee, “IEEE 1588 standard for a precision clock synchronization protocol for networked measurement and control systems”, *Sensors for Industry Conference, 2nd ISA*, pp. 98–105, 2002.
- [20] “Multiprotocol label switching (MPLS)”, [Online]. Available: <http://mplsinfo.org/> (Last accessed on: 8 Dec 2016).
- [21] M. Hlozak, D. Uhrin, J. C.-W. Lin, and M. Voznak, “Effective planning and analysis of huawei and cisco routers for MPLS network design using fast reroute protection”, *Proceedings of the 2nd Czech-China Scientific Conference*, 2017.
- [22] S. Pasqualini, A. Iselt, A. Kirstädter, and A. Frot, “Mpls protection switching versus OSPF rerouting”, *Quality of Service in the Emerging Networking Panorama*, pp. 174–183, 2004.
- [23] J. Ha and N. Park, “Unified access control for 5G convergence network with DHCP”, *Ubiquitous Wireless Broadband (ICUWB), International Conference on*, pp. 1–4, 2016.
- [24] M. Khana, Z. Ullahb, Q. Mudassir, and S. U. Khan, “Address resolution protocol caching using optimized RAM-based CAM”, *International Journal of Computer Science and Information Security*, vol. 14, no. 8, p. 1014, 2016.
- [25] P. Pandey, “Prevention of ARP spoofing: A probe packet based technique”, *Advance Computing Conference (IACC), 3rd International*, pp. 147–153, 2013.
- [26] B. Issac, “Secure ARP and secure DHCP protocols to mitigate security attacks”, *ArXiv preprint arXiv:1410.4398*, 2014.

- [27] T. Alharbi, D. Durando, F. Pakzad, and M. Portmann, “Securing ARP in software defined networks”, *Local Computer Networks (LCN), 41st Conference on*, pp. 523–526, 2016.
- [28] “The evolved packet core”, [Online]. Available: <http://www.3gpp.org> (Last accessed on: 9 Dec 2016).
- [29] “Protocol signaling procedures in LTE”, [Online]. Available: <http://radisys.com/> (Last accessed on: 9 Dec 2016).
- [30] K. Masahiro, A. Yuu, and A. Takahiro, “Development of iPASOLINK series and super-multilevel modulation technology”, *NEC Technical Journal*, vol. 8, no. 2, pp. 66–69, 2014.
- [31] “SDN architecture overview”, [Online]. Available: <http://opennetworking.org/> (Last accessed on: 12 Dec 2016).
- [32] D. Cai, A. Wielosz, and S. Wei, “Evolve carrier ethernet architecture with SDN and segment routing”, *A World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE 15th International Symposium on*, pp. 1–6, 2014.
- [33] K. Kataoka, N. Agarwal, and A. V. Kamath, “Scaling a broadcast domain of ethernet: Extensible transparent filter using SDN”, *Computer Communication and Networks (ICCCN), 23rd International Conference on*, pp. 1–8, 2014.
- [34] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, “A survey on software-defined networking”, *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2015.
- [35] C. Kim, M. Caesar, and J. Rexford, “Floodless in seattle: A scalable ethernet architecture for large enterprises”, *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 3–14, 2008.
- [36] R. Niranjana Mysore, A. Pamboris, N. Farrington, N. Huang, P. Miri, S. Radhakrishnan, V. Subramanya, and A. Vahdat, “Portland: A scalable fault-tolerant layer 2 data center network fabric”, *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 39–50, 2009.

- [37] X. Sun and Z. Wang, “An efficient and scalable metro-ethernet architecture”, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 3, no. 4, pp. 25–42, 2009.
- [38] J. Mudigonda, P. Yalagandula, J. Mogul, B. Stiekes, and Y. Pouffary, “Netlord: A scalable multi-tenant network architecture for virtualized datacenters”, *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 62–73, 2011.
- [39] E. Rojas and G. Ibáñez, “Torii-hlmac: A distributed, fault-tolerant, zero configuration fat tree data center architecture with multiple tree-based addressing and forwarding”, *Global Communications Conference (GLOBECOM)*, pp. 2523–2528, 2012.
- [40] A. Shpiner, I. Keslassy, C. Arad, T. Mizrahi, and Y. Revah, “SAL: Scaling data centers using smart address learning”, *Network and Service Management (CNSM), 10th International Conference on*, pp. 248–253, 2014.
- [41] W. Jian, H. Tao, L. Jiang, and L. Yunjie, “A novel floodless service discovery mechanism designed for software-defined networking”, *China Communications*, vol. 11, no. 2, pp. 12–25, 2014.
- [42] N. Jehan and A. M. Haneef, “Scalable ethernet architecture using SDN by suppressing broadcast traffic”, *Advances in Computing and Communications (ICACC), Fifth International Conference on*, pp. 24–27, 2015.
- [43] H. Cho, S. Kang, and Y. Lee, “Centralized ARP proxy server over SDN controller to cut down ARP broadcast in large-scale data center networks”, *Information Networking (ICOIN), International Conference on*, pp. 301–306, 2015.
- [44] F. Schneider, R. Bifulco, and A. Matusiuk, “Better ARP handling with InSPired SDN switches”, *Local and Metropolitan Area Networks (LANMAN), IEEE International Symposium on*, pp. 1–6, 2016.
- [45] K. Lee, J.-K. K. Rhee, S. Yoo, and P. Cho, “Flush optimization in ethernet ring protection network”, *ICT Convergence (ICTC), International Conference on*, pp. 513–514, 2011.

- [46] D. Lee, J.-K. K. Rhee, K. Lee, and P. Cho, “Efficient ethernet multi-ring protection system”, *Design of Reliable Communication Networks, DRCN. 7th International Workshop on*, pp. 305–311, 2009.
- [47] M. Nurujjaman, S. Sebbah, C. Assi, and M. Maier, “Optimal capacity provisioning for survivable next generation Ethernet transport networks”, *Journal of Optical Communications and Networking*, vol. 4, no. 12, pp. 967–977, 2012.
- [48] D. Lee, K. Lee, S. Yoo, and J.-K. K. Rhee, “Efficient ethernet ring mesh network design”, *Journal of Lightwave Technology*, vol. 29, no. 18, pp. 2677–2683, 2011.
- [49] C. Assi, M. Nurujjaman, S. Sebbah, and A. Khalil, “Optimal and efficient design of ring instances in metro ethernet networks”, *Journal of Lightwave Technology*, vol. 32, no. 22, pp. 3843–3853, 2014.
- [50] “PyXLL: Excel functions written in Python”, [Online]. Available: www.pyxll.com (Last accessed on: 10 Dec 2016).
- [51] “Daniel’s Excel Toolbox”, [Online]. Available: www.xltoolbox.net (Last accessed on: 12 Dec 2016).

Appendix A

Microwave Channel Capacity

Link capacities used in the simulation for microwave (MW) links are from the following table-

Channel Spacing (MHz)	Modulation Scheme	Link Capacity (Mbps)
7	QPSK	12
7	16QAM	24
7	32QAM	29
7	64QAM	37
7	128QAM	44
7	256QAM	51
14	QPSK	23
14	16QAM	48
14	32QAM	59
14	64QAM	76
14	128QAM	90
14	256QAM	104
28	QPSK	48
28	16QAM	97
28	32QAM	125
28	64QAM	150
28	128QAM	180
28	256QAM	210
56	QPSK	97
56	16QAM	190
56	32QAM	240

56	64QAM	310
56	128QAM	360
56	256QAM	420

Appendix B

Python Code used for Graph Generation from the Mathematical Model:

```

1 from pyxll import xl_func
2
3 #This function performs analysis on a ring topology to determine
   overall throughput
4 #It calculates ring throughput based on the proposed optimization on
   ITU-T G.8032
5 #Parameters: L=>Traffic added to each link (array), AFFECTED=>Affected
   Link, RPL=>Ring Protection Link, C=>Link Capacity at Highest
   Modulation Scheme, Cd1d2=>Link Capacity of Affected Link After
   Signal Degradation
6 @xl_func("float [] L, int AFFECTED, int RPL, float C, float Cd1d2: float
   ")
7 def AnalyzedThroughput(L, AFFECTED, RPL, C, Cd1d2):
8     """if z return x, else return y"""
9     FINAL=len(L)
10    count=0
11    Cn1n2=C
12
13    traffic_between_RPL_and_AFFECTED=0.0
14    for i in range(AFFECTED+1,RPL+1):
15        traffic_between_RPL_and_AFFECTED+=L[i-1][0]
16
17    traffic_between_RPL_and_FINAL=0.0
18    for i in range(RPL+1,FINAL+1):
19        traffic_between_RPL_and_FINAL+=L[i-1][0]
20
21    traffic_between_RPL_and_FINAL_capped = min(
        traffic_between_RPL_and_FINAL ,C)
22    free_capacity_x=C-traffic_between_RPL_and_FINAL_capped
23

```

```

24 traffic_between_FIRST_and_AFFECTED=0.0
25 for i in range(1,AFFECTED+1):
26     traffic_between_FIRST_and_AFFECTED+=L[i-1][0]
27
28 Cdeg = min(traffic_between_RPL_and_AFFECTED , Cd1d2)
29 Throughput=0
30 Throughput1=traffic_between_RPL_and_FINAL_capped+min(
    traffic_between_RPL_and_AFFECTED-free_capacity_x ,Cd1d2)+min(
    traffic_between_FIRST_and_AFFECTED ,C-Cdeg)+free_capacity_x
31 Throughput2=min(traffic_between_RPL_and_FINAL ,C)+min(
    traffic_between_RPL_and_AFFECTED , Cd1d2)+min(
    traffic_between_FIRST_and_AFFECTED ,C-Cdeg)
32 if traffic_between_RPL_and_AFFECTED>Cd1d2:
33     Throughput=Throughput1
34 else :
35     Throughput=Throughput2
36 return Throughput
37
38
39 #This function performs analysis on a ring topology to determine
    overall throughput
40 #It calculates ring throughput for unmodified ITU-T G.8032
41 #Parameters: L=>Traffic added to each link (array), AFFECTED=>Affected
    Link , RPL=>Ring Protection Link , C=>Link Capacity at Highest
    Modulation Scheme , Cd1d2=>Link Capacity of Affected Link After
    Signal Degradation
42 @xl_func(" float [] L, int AFFECTED, int RPL, float C, float Cd1d2: float
    ")
43 def AnalyzedThroughputUnoptimized(L, AFFECTED, RPL, C, Cd1d2):
44     """ if z return x, else return y"""
45     FINAL=len(L)
46     count=0
47     Cn1n2=C
48
49 traffic_between_RPL_and_AFFECTED=0.0
50 for i in range(AFFECTED+1,RPL+1):
51     traffic_between_RPL_and_AFFECTED+=L[i-1][0]
52
53 traffic_between_RPL_and_FINAL=0.0

```

```

54 for i in range(RPL+1,FINAL+1):
55     traffic_between_RPL_and_FINAL+=L[i-1][0]
56
57 traffic_between_FIRST_and_AFFECTED=0.0
58 for i in range(1,AFFECTED+1):
59     traffic_between_FIRST_and_AFFECTED+=L[i-1][0]
60
61 Cdeg = min(traffic_between_RPL_and_AFFECTED , Cd1d2)
62
63 ThroughputUnoptimized=min(traffic_between_RPL_and_FINAL ,C)+min(
64     traffic_between_RPL_and_AFFECTED , Cd1d2)+min(
65     traffic_between_FIRST_and_AFFECTED ,C-Cdeg)
66
67 return ThroughputUnoptimized
68
69 #This function is used for checking intermediate outputs, see above two
70     functions for parameter details
71 @xl_func("float [] L, int AFFECTED, int RPL, float C, float Cd1d2:
72     string")
73 def AnalyzedThroughputX(L, AFFECTED, RPL, C, Cd1d2):
74     """ if z return x, else return y"""
75     FINAL=len(L)
76     count=0
77     Cn1n2=C
78
79     traffic_between_RPL_and_AFFECTED=0.0
80     for i in range(AFFECTED+1,RPL+1):
81         traffic_between_RPL_and_AFFECTED+=L[i-1][0]
82
83     traffic_between_RPL_and_FINAL=0.0
84     for i in range(RPL+1,FINAL+1):
85         traffic_between_RPL_and_FINAL+=L[i-1][0]
86
87     traffic_between_RPL_and_FINAL_capped = min(
88         traffic_between_RPL_and_FINAL ,C)
89     free_capacity_x=C-traffic_between_RPL_and_FINAL_capped
90
91     traffic_between_FIRST_and_AFFECTED=0.0
92     for i in range(1,AFFECTED+1):

```

```
88     traffic_between_FIRST_and_AFFECTED+=L[i - 1][0]
89
90     Cdeg = min(traffic_between_RPL_and_AFFECTED , Cd1d2)
91     Ret="Result {0},{1},{2},{3},{4},{5},{6}".format(
        traffic_between_RPL_and_FINAL_capped ,
        traffic_between_RPL_and_AFFECTED , free_capacity_x , Cd1d2 ,
        traffic_between_FIRST_and_AFFECTED , Cdeg ,
        traffic_between_RPL_and_FINAL )
92     # Throughput=traffic_between_RPL_and_FINAL_capped+min(
        traffic_between_RPL_and_AFFECTED-free_capacity_x ,Cd1d2)+min(
        traffic_between_FIRST_and_AFFECTED ,C-Cdeg)+min(
        traffic_between_RPL_and_FINAL , free_capacity_x )
93
94     return Ret
```