PER-PRIORITY FLOW CONTROL (PPFC) FRAMEWORK FOR ENHANCING QOS IN METRO ETHERNET

BAHAREH PAHLEVANZADEH

UNIVERSITI SAINS MALAYSIA

2013

PER-PRIORITY FLOW CONTROL (PPFC) FRAMEWORK FOR ENHANCING QOS IN METRO ETHERNET

by

BAHAREH PAHLEVANZADEH

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

November 2013

ACKNOWLEDGEMENTS

Praise God, because of all these presents and opportunities.

At first, I would like to express my truthful respect and sincere gratitude to my supervisor Associate Professor Dr. Wan Tat Chee, without whose continuous precious guidance, emotional and financial support this work would not have been achieved.

I would like to acknowledge University Science Malaysia for the fellowship and USM-RU-PRGS grant and creating a resourceful academic environment. Special acknowledgements go to all academic and non-academic members of the School of Computer Sciences, Network Research Group (NRG) for their camaraderie, caring and warm hospitality.

Special thanks go to senior research scientists; Dr. Sven-Arne Reinemo (University of Oslo), Dr. Kenjiro Cho (Internet Initiative Japan Inc.), Dr. Vishal Sharma (University of California & Metanoia Inc. USA), and Dr. Alireza Taghizadeh (Universiti Sains Malaysia) for their valuable comments and guidance.

Finally, I have to personally make acknowledgment to my very dear husband, Alireza, for his love, and kind indulgence. I wish to express my gratitude to my parents and my only sibling, Sharareh, for their unflagging understanding and supports over this long journey. I also credit my lovely Malaysian-born daughter, Anousheh, for her accompanies and also her self-sacrifices to stay far from me for thousands of miles during last stages of my PhD. She inspired me in the hardest moments of doing PhD and made me grateful and happy of whatever I have as presents and opportunities. Yes! mommy's book is finished now and I am coming home to read you storybooks.

I have been blessed with the continues support from each and every member of my family. Consequently, I deeply dedicate this thesis to them for all their love, patience and courage.

Bahareh Pahlevanzadeh

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACK	Acknowledgement
AF	Assured Forwarding
AI	Additive Increase
AIMD	Additive Increase Multiplicative Decrease
BE	Best Effort
Buf	Buffer
BW	Bandwidth
CBR	Constant Bit Rate
CDF	Cumulative Distribution Function
CCU	Congestion Control Unit
CSU	Classifier and Shaper Unit
CoS	Class of Service
COV	Coefficient of Variation
DB-EFC	Device Based Ethernet Flow Control
DiffServ	Differentiated Service
EFC	Ethernet Flow Control
EtherValve	OMNET++ Simulator-based EthernetValve Evaluation Framework
E2E	End-to-End

EWMA	Exponential Weighted Moving Average
FPP	Final Prioritized PAUSE
HBH	Hop-By-Hop
HP	High Priority
HRT/HUDP	Hard Real Time based on UDP
HVCMacro-PPFC	Hybrid Virtual Color and Class-Based Prioritized Buffer (Macro Flow Based Scheme)
HVCMicro-PPFC	Hybrid Virtual Color and Class-Based Prioritized Buffer (Micro Flow Based Scheme)
IDM-Unit	Intelligent Decision Making Unit
IETF	Internet Engineering Task Force
IFS	Intelligent Flow Selection
I2CNA	Intelligent Implicit Congestion Notification and Action
IntServ	Integrated Service
IPv6	Internet Protocol version 6
IPDV	IP Packet Delay Variation
IPP	Initial Prioritized PAUSE
ISCD	Intelligent Sampling and Congestion Detection
ITU-T	International Telecommunication Union
LLC	Link Layer Control

LP	Low Priority
NRT	Non Real Time
MAC	Media Access Control
MAN	Metropolitan Area Network
MD	Multiplicative Decrease
Metro Ethernet	Metropolitan Ethernet
MEF	Metro Ethernet Forum
MIDM Unit	Monitoring and Intelligent Decision Making Unit
M-Unit	Monitoring Unit
MP	Medium Priority
NIC	Network Interface Card
NGN	Next Generation Network
NP	NO PAUSE
OWD	One Way Delay
OWPL	One Way Packet Loss
OSI	Open Systems Interconnection
P	Priority
P-AIMD	Prioritized Additive Increase Multiplicative Decrease
PB-EFC	Port Based Ethernet Flow Control
P-EWMA	Prioritized Exponential Weighted Moving Average Algorithm

PPFC	Per-Priority Flow Control
Q	Queue
QoS	Quality of Service
RFC	Request for Comments
RTO	Retransmission Time-Out
RTT	Round Trip Time
SACK	Selective Acknowledgment
SAN	Storage Area Networks
SLA	Service Level Agreement
Std.	Standard
SRT/SUDP	Soft Real Time based on UDP
TCP	Transmission Control Protocol
VBR	Variable Bit Rate
VLAN	Virtual Local Area Network
VOD	Video On Demand
VOIP	Voice Over IP
VT-i	Virtual Traffic-Based Prioritized Buffer (Input)
VT-O	Virtual Traffic-Based Prioritized Buffer (Output)

KAWALAN ALIRAN BERDASARKAN SETIAP PRIORITI UNTUK MENINGKATKAN KUALITI PERKHIDMATAN ETHERNET METRO

ABSTRAK

Hari demi hari komunikasi Internet dan perkhidmatan semakin menghadapi peningkatan yang tinggi dari segi kepelbagaian dan bilangan kapasiti serta permintaan. Oleh itu, pendekatan pengoptimuman Internet bagi tujuan membuat pengurusan trafik dan kualiti perkhidmatan (QoS) menjadi satu bidang penyelidikan mencabar; sementara kajian kawalan aliran dan kawalan kesesakan turut dipertimbangkan sebagai asas-asas penting untuk tujuan kawalan trafik terutamanya di kelajuan tinggi Metro Ethernet.

IEEE telah menetapkan piawaian sebagai satu kaedah (Standard IEEE 802.3x) bagi penyediaan *Ethernet Flow Control (EFC)*, yang mana menggunakan bingkai PAUSE apabila bingkai kawalan MAC di lapisan pautan data yang bertujuan menerima atau menghalang penghantaran bingkai. Walau bagaimanapun, pendekatan ON/OFF konvensional IEEE 802.3x mungkin tidak lagi sesuai dengan Metro Ethernet Carrier. Oleh itu, satu seni bina dan mekanisme baru yang menawarkan fleksibiliti lebih baik, pengaliran yang berkesan, kawalan kesesakan dan perkhidmatan kualiti yang tinggi kini sangat diperlukan.

Penyelidikan ini membentangkan satu rangka kerja skim baru yang dipanggil *Per-Priority Flow Control (PPFC)* iaitu satu kesinambungan kepada ukuran asas IEEE 802.3x di bingkai PAUSE berdasarkan keutamaan baru pengurusan menggunakan kelebihan spesifikasi *Hybrid Parametic Flow Label* di pelbagai lapisan suis. Rangka kerja penilaian prestasi *EtherValve* telah dibangunkan berasaskan perisian simulasi OMNET++ untuk penilaian prestasi kuantitatif bagi tiga skim pedalaman PPFC yang dicadangkan berbanding penyelesaian sedia ada dan skim IEEE 802.3x konvensional.

Konvensional *Device-Based Single Shared Physical Buffer-EFC (DB-EFC)* dan *Port-Based Physical Buffer-EFC (PB-EFC)* dimodelkan berdasarkan lapisan pautan yang berpangkalan di skim asal IEEE 802.3x; di mana skim dalaman PPFC yang dicadangkan mengikut ciri-ciri IEEE 802.3x di mana skim dalaman PPFC yang dicadangkan *Virtual Traffic-Based Prioritized Buffer-EFC (VTi-PPFC dan VT-O)*, dan *Hybrid Virtual Color dan Class-Based Prioritized Buffer (HVCMacro-PPFC dan HVCMicro-PPFC*), mengutamakan dan secara virtualisasi menyeberang skim berlapis. Mempertingkatkan mekanisme dan algoritma standard asal, melalui *Intelligent Sampling dan Congestion Detection (ISCD)*, *Intelligent Flow Selection (IFS)*, *Intelligent Congestion Notification dan Action*, *Intelligent algorithm Reaction*, dan mekanisme pengurusan bingkai PAUSE telah dilakukan bagi menyediakan mengutamakan kawalan aliran dan pengurusan sumber.

Hasil analisa rangka kerja penilaian prestasi *EtherValve* menunjukkan bahawa skim dalaman PPFC yang dicadangkan telah meningkatkan mekanisme kawalan kesesakan; dan oleh itu prestasi QoS skim IEEE 802.3x konvensional di satu berbutir adalah stabil, rata, dan aksi baik. Juga, kesaksamaan berkadar di daya pemprosesan untuk rancangan *PPFC* sebagai keberkesanan pengutamaan, pengurusan penimbal maya dan penjadualan berhierarki telah dilihat daripada keputusan. Pemerhatian keseluruhan dari keputusan-keputusan dianalisa menunjukkan bahawa skim dalaman PPFC yang dicadangkan mencapai pengurangan pada bilangan cetusan bingkai *Prioritized PAUSE*, skaligus menkong prestasi QoS tinggi dari segi kelewatan, jitter dan paket hilang yang elok untuk kedua-dua aplikasi trafik nyata dan aplikasi trafik bukan masanyata.

PER-PRIORITY FLOW CONTROL (PPFC) FRAMEWORK FOR ENHANCING QOS IN METRO ETHERNET

ABSTRACT

Day by day Internet communication and services are experiencing an increase in variety and quantity in their capacity and demand. Thus, making traffic management and quality of service (QoS) approaches for optimization of the Internet become a challenging area of research; meanwhile flow control and congestion control will be considered as significant fundamentals for the traffic control especially on the high speed Metro Ethernet.

IEEE had standardized a method (IEEE 802.3x standard), which provides Ethernet Flow Control (EFC) using PAUSE frames as MAC control frames in the data link layer, to enable or disable data frame transmission. With the initiation of Metro Carrier Ethernet, the conventional ON/OFF IEEE 802.3x approach may no longer be sufficient. Therefore, a new architecture and mechanism that offer more flexible and efficient flow and congestion control, as well as better QoS provisioning is now necessary.

This research presents a new scheme-based framework called Per-Priority Flow Control (PPFC) as an extension to the IEEE 802.3x standard based on a new priority PAUSE frame management using the advantages of Hybrid Parametric Flow Label Specification through multi-layer switches. The EtherValve performance evaluation framework have been developed based on OMNET++ simulation software for quantitative performance evaluation of three proposed PPFC interior schemes versus existing solution and the conventional IEEE 802.3x scheme.

Conventional Device-Based Single Shared Physical Buffer-EFC (DB-EFC) and Port-Based Physical Buffer-EFC (PB-EFC) schemes are modelled according to the link layer based scheme of original IEEE 802.3x; whereas the proposed PPFC interior schemes, Virtual Traffic-Based Prioritized Buffer-EFC (VTi-PPFC and VT-O), and Hybrid Virtual Color and Class-Based Prioritized Buffer (HVCMacro-PPFC and HVCMicro-PPFC), are prioritized and virtualized cross layered schemes. Enhancing the mechanisms and algorithms of the original standard, via Intelligent Sampling and Congestion Detection (ISCD), Intelligent Flow Selection (IFS), Intelligent Congestion Notification and Action (ICNA), Intelligent Reaction (IR3) algorithms, and PAUSE frame management mechanism was done in order to provide more effective prioritized flow control and resource management.

The analysed outcome of EtherValve performance evaluation framework showed that the proposed PPFC interior schemes enhance the congestion control mechanism; and hence, QoS performance of conventional IEEE 802.3x scheme in a granular, stable, smooth, and fair manner. Also, the proportional fairness in throughput for PPFC's schemes as the effectiveness of prioritization, virtual buffer management and hierarchical scheduling has seen from the results. The overall observation from the analysed results showed that the proposed PPFC interior schemes achieved lower number of triggered Prioritized PAUSE frames, to support high QoS performance in term of delay, jitter and packet loss that is desirable for both real and non-real time traffic applications.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The following Sections provides research motivation and problem statement, objectives, scopes and thesis boundary and finally this Chapter ends with the theoretical methodology of the study and thesis organisation.

1.2 Background

For a long time Ethernet has been the predominant local area network standard. This popularity is due to its simplicity, maturity, low costs and a wide existing base. Fast and Gigabit Ethernet have brought more bandwidth to the technology and by moving from half-duplex to full-duplex links, Ethernet throughput has increased. With these properties Ethernet is expected to speed up the deployment and use of next generation networks and services (Gai and DeSanti, 2009; Reinemo et al., 2010). Moreover, Full-duplex operation is also ideal for backbones and high-speed server or router links. The recent effort for backplane Ethernet will allow Ethernet to be used in server and I/O backplanes in the future. Ethernet has also gained popularity in the automation world (Sommer et al., 2010; Anghel et al., 2011). All things considered, Ethernet is truly on its way to becoming omnipresent (Chiruvolu et al., 2004; Fang et al., 2011).

With the growth of bandwidth hungry applications such as mobile agents and the increasing number of users connected to the LANs (Reinemo and Skeie, 2005), it can be expected that high speed Metropolitan Ethernet Networks (Metro Ethernet) will be saturated after deployment in 3-5 years as predicted by the proponents (Malhotra, 2008; Reinemo et al., 2010).

Metro Ethernet provides an easy and cheap way to interconnect multiple sites of an enterprise. Moreover, as the penetration of Ethernet increases, Carrier Grade Ethernet aims to overcome the shortcoming of the native Ethernet, in order to meet Quality of Service (QoS) requirements of end applications such as response times, throughput, delay and jitter by effective resource management and flow control (Malhotra, 2008). However, the current activity in building new generation of high speed Metro networks, growing network bandwidth and intensive network applications lacked effective flow and congestion controls and providing a high level of QoS for metro Ethernet became a challenging topic in network design.

The issue of flow control was initially solved with an extension to the standard, named IEEE 802.3x. The IEEE 802.3x extension adds control frames and ON/OFF type flow control to Ethernet (Seifert, 1998), reducing the buffer overflow problem. Thus, the aim of this standard is to improve the throughput, latency and packet loss of data flow, by avoiding deadlocks, overloads and ensure fair allocation of resources among competing users.

1.3 Research Motivation and Problem Statement

Congestion control, resource and traffic management problems have attracted wide attention from the IETF and IEEE research groups in the past few years. The majority of this research in Internet context (Kelly, 2000; Paganini et al., 2001; Kunniyur and Srikant, 2003) as well as in wired and wireless network context (Holland and Vaidya, 2002; Anantharaman et al., 2004; Schoenen and Otyakmaz, 2010; Pahuja et al., 2011) have focused on modeling, analysis, algorithm development of End-to-End (E2E) congestion control schemes, and adaptation of such schemes for different network architectures (Yi and Shakkottai, 2007).

Although various proposals appear in the literature over the past decade in the area of congestion control protocols (Zhuang et al., 2012) for transport layer or network layer, but to the best of researcher's knowledge, except few proposals in wireless field (Wang et al., 2006; Yaghmaee and Adjeroh, 2009), etc.; they mostly focused on end to end congestion control protocols. There was a lack of comprehensive research study on the link layer congestion control, particularly on IEEE 802.3x as a Hop-By-Hop (HBH) flow and congestion control protocol.

Ethernet technology has been an attractive technology to be deployed in MAN for service providers (Ibanez et al., 2004; Xiaocui Sun and Zhijun Wang, 2010). On the other hand, the newly designed network technologies surpass today' Ethernet in both speed and functionality; however, with respect to networking principles, Ethernet has to overcome some of its weaknesses compared to newer technologies. So this has spurred the major Ethernet vendors and IEEE to create different IEEE task groups to adapt Ethernet to high performance networking (Reinemo et al., 2010).

On the other hand, since Ethernet technology was not designed to be deployed as a transport technology, therefore, it is not surprising that the existing Ethernet QoS model is not appropriate to meet the demands of next generation applications (Elby et al., 2007). However, the success of Metro Ethernet depends greatly on its ability to live up to the QoS requirements of the application delivered over it (Malhotra, 2008). For instance, one of the QoS features currently available and enforced by standardization bodies for Ethernet is IEEE 802.3x Ethernet Flow Control (EFC).

IEEE 802.3x-EFC initially aimed to provide a mechanism for devices to throttle incoming traffic to avoid frame loss due to congestion. Therefore, traffic flow regulation can improve the throughput, latency and packet loss. Although IEEE 802.3x-EFC is simple, it is unfair and inefficient and suffer from some criticisms. For instance, it lacks a guarantee against frame loss since its features left as optional with no much effort for enhancing this standard; and more importantly it is coarse granularity because it blocks all traffic indiscriminately and even cause more spreading of congestion in the network (Hagen and Zarick, 2011).

On the other hand, this inefficiency is because IEEE 802.3x was designed as a basic flow control solution to a simple point to point connection with no service differentiation, and thus when Ethernet has deployed in MAN areas, some loss in scalability and efficiency has surfaced. Ethernet LAN and its MAC control frame, named PAUSE frame, has no provision for flow, logical and physical selectivity (Kadambi et al., 1998); PAUSE frame *Destination Address* is always set to 01-C2-80-00-00-01 (Seifert, 1998; Nizam et al., 2003). As a result, any upstream node that receives the PAUSE frame will immediately activate its backpressure mechanism that stops all traffic to the *Sending Address* (as indicated in the PAUSE frame). Such scheme is proper only for best effort traffic insensitive to long delays or latency.

Moreover, with the initiation of Gigabit Ethernet as the carrier domain for various broadband services, and the introduction of multiple Virtual Local Area Network (VLAN) connections; the conventional ON/OFF PAUSE approach of IEEE 802.3x may no longer be suitable (Nizam et al., 2003). As mentioned, there are some general inefficiency problems for the IEEE 802.3x standard that are highlighted briefly as follows:

- PAUSE mechanism can result in a high degree of service degradation due to its coarse granularity; in where all the incoming traffic, irrespective of their QoS requirements are affected by PAUSE.
- PAUSE mechanism is not able to distinguish between the misbehaved (aggressive sources that causes congestion) and well-behaved sources.
- Shutdown of a whole link instead of a particular flow causes low bandwidth utilization and sharp oscillation in throughput

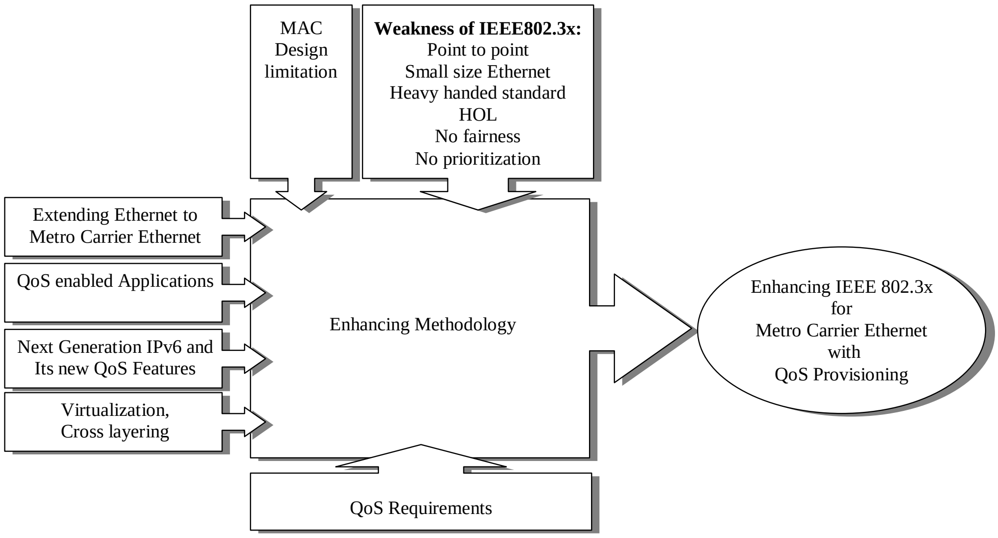


Figure 1.1: Formulation of Problem Domain

- Due to lack of prioritization and granularity in IEEE 802.3x; PAUSE mechanism inevitably causes congestion spreading and or Head-of-Line (HOL) blocking

- The duration of PAUSE, known as pause-time for pausing the transmission needs to be taken into consideration for smoother regulation of traffic flow.

Thus, QoS can not operate properly and there is no guarantees for the QoS provisioning. There are other gaps in the previous research that have not been addressed for the Hop-By-Hop (HBH) IEEE 802.3x in Ethernet technology precisely (discussed in Chapter 2). All these issues provide the motivation to direct the present research study on enhancement of HBH IEEE 802.3x for Ethernet technology. Therefore, a new comprehensive approach as an enhancement to the IEEE 802.3x that can offer more flexible and efficient congestion control with respect to the QoS issues is necessary. Figure 1.1 illustrates the formulation of problem domain.

1.4 Research Questions

The mentioned research problem raises the following questions, which should be answered by the thesis:

i) What are the drawbacks of existing IEEE 802.3x standard (Ethernet Flow and Congestion Control) and other existing schemes in dealing with transient congestion, in Metropolitan Ethernet which affect QoS?

ii) What approaches can be employed to enhance Ethernet flow control to improve QoS performance when the network experiences transient congestion?

iii) How to implement the proposed schemes and evaluate their effectiveness in comparison to existing Ethernet Flow and Congestion Control schemes, in terms of packet loss, delay, jitter, throughput, and queue oscillation?

1.5 Research Objectives

The overall goal of the present research is to enhance QoS in Metro Ethernet and to increase the granularity of Ethernet flow control in order to enhance the IEEE 802.3x standard for Metro Ethernet. To achieve the above goal, the specific objectives are defined as follows:

i) To design an enhanced flexible and granular Ethernet Flow Control framework for Metro Ethernet Network with better QoS performance during congestion

ii) To verify correctness and effectiveness of cross layer algorithms used in the framework in comparison with the existing conventional schemes via simulation

iii) To optimize the proposed algorithms' functionalities by evaluating the respective parameters and their impact on network performance

iv) To investigate the performance of Hop by Hop congestion control in interaction with TCP End to End congestion control for QoS provisioning in Metro Ethernet

1.6 Research Scope

The scope of this research is limited to QoS domain of next generation services and applications in wired metropolitan Ethernet technology; and particularly, it is focused on:

i) Enhancing the HBH flow control scheme in Ethernet in order to meet a scalable QoS provisioning for future Metro Ethernet

ii) Investigating the interaction of the proposed framework interior schemes with E2E congestion control

iii) Investigating the impacts of proposed framework interior schemes for IPv6-based homogenous and heterogenous traffic (real and non-real time traffic).

1.7 Research Methodology

The following outline defines the steps used to develop the proposed framework for addressing the problem statement:

- Quantify the requirements for having an effective granular prioritized flow control scheme in Metro Ethernet
  - Survey of conducted research in flow and congestion control schemes and particularly Metro Ethernet HBH flow and congestion control
  - Survey of conducted research in QoS architectures and particularly IPv6 QoS (by focusing on developments of flow label specification)
- Specify a suitable framework for the specification, analysis and deployment of hybrid prioritized flow control for Ethernet (prioritized PAUSE management)

- Develop suitable schemes and algorithms to ensure that an effective flow control can be accommodated effectively over Metro Ethernet
- Determine the effectiveness of proposed schemes, algorithms and methodologies in achieving given goals
- Compare the performance of proposed schemes and algorithms using computer simulation techniques to determine their suitability, using the OMNET++ simulation platform software
- Analyse the effectiveness of different parameters for optimizing the proposed schemes' performance

1.8 Thesis Organization

After introducing the significant of this research study by providing some evidences and background information, stating existing problems and clarifying its objectives, the rest of thesis is organized into 6 more Chapters as follows:

Chapter 2 includes theoretical background, in where extensively covers the literature survey and discusses the most current and related works in flow and congestion control mechanisms in different layers of OSI model. The researcher will also discuss properties of IPv6 header, QoS models, and principle requirements for enhancing IEEE 802.3x standard in wider environment of Metro Carrier Ethernet. Finally, the Chapter is wrapped up with the summary of existing approaches and their corresponding algorithm and architectures, etc. and provides a perspective for introducing the proposed framework in methodology Chapter.

Chapter 3 covers the methodology discussion on how the proposed framework was designed. It defines the requirements and specifications for a comprehensive hybrid cross-layered PPFC architecture used to support the QoS provisioning for metropolitan Ethernet applications.

This Chapter provides a wide and intermediate perspective of the proposed framework. System model objectives, system operation, system architecture, message transactions along with functional analysis for IEEE 802.3x and proposed hybrid cross-layered PPFC are the topics that introduced in this Chapter. Different enhanced algorithms are discussed in general too. The reasons of choosing hybrid cross-layered PPFC for Metro Carrier Ethernet are justified. In other word, this Chapter provides a high level view for the proposed PPFC framework.

Chapter 4 elaborates proposed framework (PPFC) system architecture along with its corresponding components functionality aspects in detail from lower level view. It provides an introduction to the proposed EtherValve performance evaluation framework and its module components, algorithms and parameters comprehensively. The interior modules of proposed EtherValve framework; named DB-EFC, PB-EFC, VT-i-PPFC, VT-O-PPFC, HVCMacro-PPFC, and HVCMicro-PPFC modules and their simple and compound sub-modules, units, messages, algorithms are presented in this Chapter. The impact of choosing different parameters such as buffer size, thresholds, traffic rate classification, queuing techniques, buffer management and scheduling mechanism, assigned weights are briefly introduced for each scheme.

Chapter 5 introduces simulation environment and the proposed EtherValve performance evaluation framework for PPFC in term of design and analysis. Moreover, it states simulation parameters, scenarios for different experiments, and QoS performance metrics used, while simulation results, analysis and discussion for experiments are presented in **Chapter 6**.

Chapter 7 introduces the research findings; research conclusion, and the possible future work for this study.

In addition to the main Chapters, detailed representing data are included in Appendices.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This Chapter provides a background for the current research. It introduces the work related to Ethernet Flow Control, QoS mechanisms and prioritisation, IPv6 flow label specification principles. And also serves as a justification for importance of research problem (why there is a need for Ethernet QoS), and its possible solutions. The Chapter is organized as follows, Section 2.2 provides a brief explanation about Ethernet, and its QoS requitement for extending Ethernet to Metro Carrier Ethernet.

Section 2.3 reviews some background information of different QoS model. Fundamentals of flow and congestion control through their conventional taxonomy in different levels (OSI layers) are explained; and then by reviewing some background information of HBH IEEE 802.3x and reviewing different related work for enhancing this standard Section 2.4 is ended. Section 2.6 concentrates on important issues of Ethernet Flow Control (EFC) concerning QoS performance and provides a comprehensive discussion of recently proposed methods. Finally, Section 2.9 gives a summary of the review and its connection to the future Chapters. Definition of terminologies used in this research context are presented in Appendix A.

2.2 Metro Ethernet

2.2.1 Ethernet Success History

According to the comprehensive survey done on history of Ethernet and its fields of application (Sommer et al., 2010); first version of Ethernet, named ALOHA, has born in 1972 (Spurgeon, 2000). Then, Ethernet media access control protocol carrier sense multiple access with collision detection (CSMA/CD) (IEEE802.3, 2005) has been created for improving ALOHA. However, the well-known DIX Ethernet II has been published by Intel and DEC, Xerox in the early 1980s. In 1985, the IEEE began to standardize the different versions of Ethernet. The Ethernet brand name has been avoided and instead the technology 802.3 CSMA/CD has been used (Spurgeon, 2000).

Although until the mid 1990-ies various LAN technologies (such as FDDI, Token Ring) were suggested; but Ethernet has been chosen as a dominant LAN technologies (in wired and wireless) (Rathnayaka and Potdar, 2011; Yigitel et al., 2011) due to its simple operation, adaptability with other networks, scalability, simple migration features, as well as its low cost. Consequently, the fast bandwidth evolution of Ethernet, as well as its flexibility, scalability and adaptability enabled this success in a wide range of surroundings. Therefore, Ethernet is an attractive replacement for metro and core network technologies as well as an access technology (Ali et al., 2006; Sommer et al., 2010). Figure 2.1 depicts the Ethernet growth and its fields of application respectively.

2.2.2 The Need for QoS in Metro Ethernet

The term *Metro Ethernet* is used synonymously for *Carrier Ethernet* or *Carrier grade Ethernet*, as the metro area is often considered as the initial deployment target of Ethernet in such networks (Sommer et al., 2010). Termed as *Carrier Ethernet* it is expected to overcome most of the shortcomings of native Ethernet. Therefore, by increasing the Ethernet presentation in public networks, the offered QoS will become increasingly important, and new enhanced Ethernet must possess functionalities to address this issue. Hence, the success of Carrier Ethernet depends greatly on its ability to live up to the QoS demands of the applications delivered over it. In this respect, the inherent variations in user traffic cause unpredictable congestion patterns

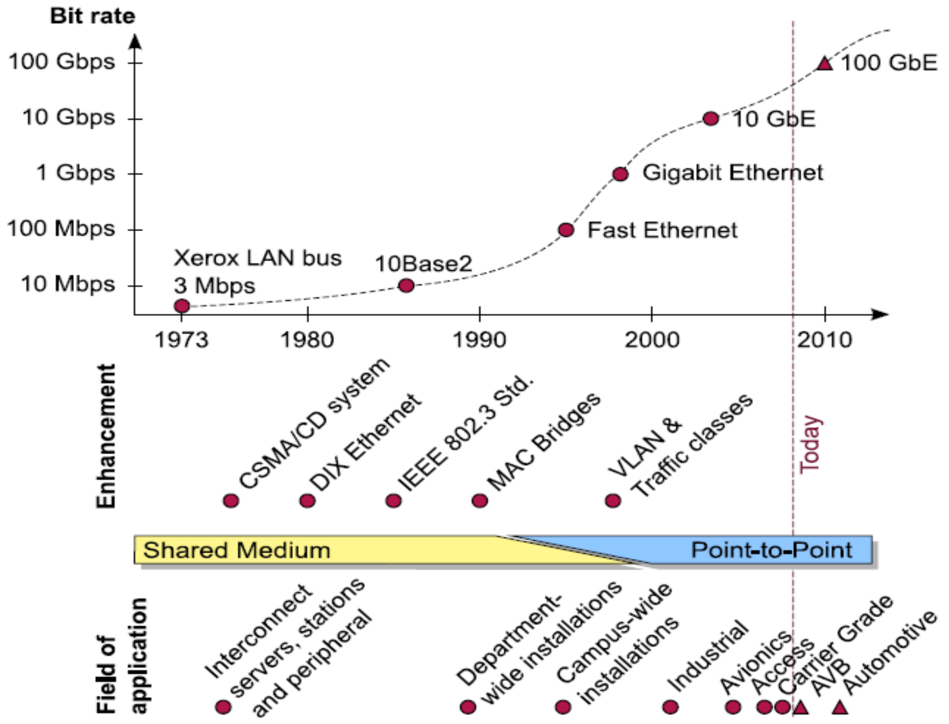


Figure 2.1: Evolution of Ethernet, its enhancements, and its Fields of Application (Sommer et al., 2010)

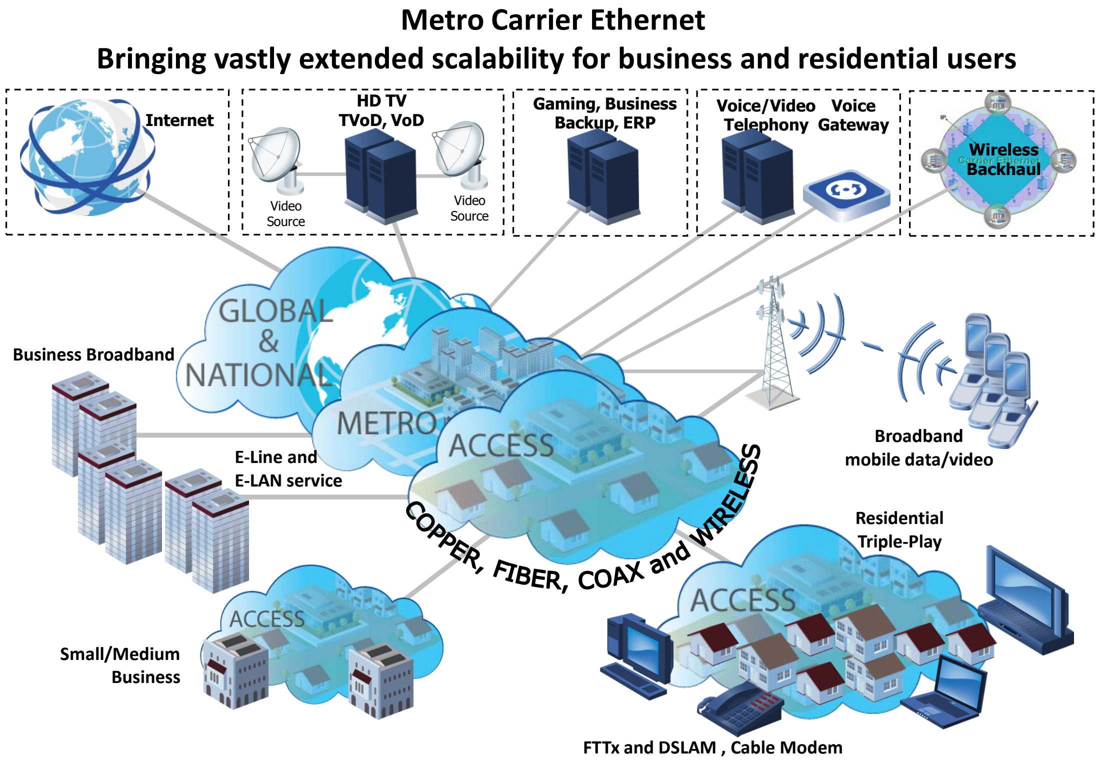


Figure 2.2: Metro Ethernet Forums Vision for Carrier Ethernet (MEF, 2008)

and pose difficulties for QoS provisioning (Malhotra et al., 2009, 2010). Efforts are underway to address this issue for Carrier Ethernet. However, still many challenges remain, which have to be overcome (Elby et al., 2007).

The QoS demand is driven by two challenges faced by Metro Ethernet. Firstly, since variety of applications are being supported by Ethernet networks, however it should be able to satisfy application requirements and user perception. Secondly, Metro Ethernet should retain and improve cost-effectiveness of current and future network deployments (Malhotra, 2008).

As mentioned, the challenge is to meet the QoS requirements of applications such as throughput, packet loss, delay and jitter by managing the network resources. Usually the term of QoS comes along with flow and congestion control (Behrouz and Sophia, 2003). Consequently, this thesis aimed to analyse and enhance Ethernet flow and congestion control protocol (IEEE 802.3x) which improves the QoS performance of Ethernet, enabling it to meet the demands of the current and next generation services and applications (Figure 2.2). It should be highlighted that, the terms *Carrier Ethernet* and *Metro Ethernet* are used in this context interchangeably. This is because the research presented in this thesis, on one hand, improves Ethernet and helps it become carrier-class by means of service differentiation (Malhotra, 2008) and on the other hand, its applicability is not restricted to the size or extent of the network (metro, access or core).

2.3 Overview of Quality of Service (QoS)

QoS refers to the ability of a network to provide improved service to selected network traffic and provide a satisfactory experience for users over various underlying technologies including wired-based and wireless-based technologies. As shown in Figure 2.3 the main features of QoS in Data Plane are: Classification, Marking, Shaping, Policing, Flow and congestion Control,

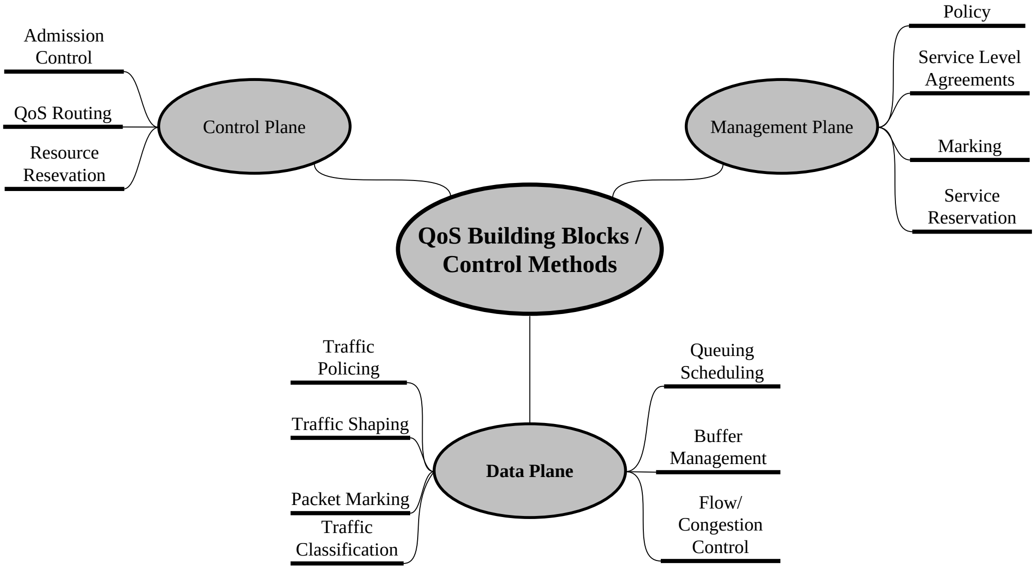


Figure 2.3: QoS Building Blocks

Buffer management, Queuing and Scheduling; to provide some grade of service for different converged networking. Specifically, QoS features provide improved and more predictable network service by providing the following services:

- Supporting dedicated bandwidth
- Improving loss characteristics
- Avoiding and managing network congestion
- Shaping network traffic
- Setting traffic priorities across the network

QoS performance is evaluated based on some main factors such as Throughput, Latency, Jitter and Loss. Depending on the nature of application the QoS requirements are different. For instance, the delay and jitter metrics are important QoS metrics that must be fulfilled for real time applications such as VOIP and VOD in business and commercial environment. Whereas, considering the Metro Carrier Ethernet implementation in data centre (Smoot and Tan, 2011) (Zhang and Ansari, 2011), academic environment (Ren and Jiang, 2010; Anghel et al., 2011), industrial or smart home (Hagen and Zarick, 2011), etc.; with heavy elastic data traffic, the packet loss and delay are two most important QoS metrics that need to be fulfilled.

Although QoS is primarily an IP concept and uses designed tools and protocols to aid the provision of defined predictable data transfer characteristics. But QoS can also be relevant within the Ethernet environment via cross-layering and traffic class mapping issues (explained in Chapter 4).

2.3.1 QoS Models

In terms of QoS models, the most well-known QoS models are *i)* Best effort with no QoS concern, *ii)* Integrated Service (IntServ) and *iii)* Differentiated Service (DiffServ) that are considered as old exemplary architectures proposed.

Best Effort (No QoS): Best Effort is the traditional datagram model. No differentiation between real time and non-real time traffic exists in this model which contributes to unpredictable services. In other words, best effort means that packets are served on a first-come-first-serve (FCFS basis).

Stateful- Integrated Services (IntServ): *Controlled load service* (Braden et al., 1994) and *guaranteed rate service* (White, 1997) have been defined as two services on IP networks which are collectively called Integrated Services (IntServ). Under lightly utilized networks by using controlled load service can approximate the behavior of best effort service. Guaranteed rate service, which mostly is referred to as IntServ, guarantees E2E QoS by means of reserving, allocating, and providing an amount of predefined resource to each flow or session in each server. Also, signaling for resource reservation, while managing hundreds of thousands of flows in a network node requires a great deal of work through RSVP protocol. This complexity inhibits the adoption of IntServ-type QoS architectures in real networks with large scales of flows that is requiring devices to retain state information (Joung et al., 2008; Mohamad et al., 2010).

Stateless- Differentiated Services (DiffServ): DiffServ architecture has been proposed by IETF Blake et al. (1998) to solve the scalability problem of IntServ. It classifies packets or the flows to which they belong into a number of traffic classes. And the packets are marked accordingly at the edge of a network. Therefore, hard work is only necessary at the edge nodes and core nodes (switches) are allowed to do more important processing tasks. Alike IntServ, classes may be assigned with strict priorities, or a certain amount of bandwidth is provisioned for each class (Joung et al., 2008). DiffServ is a highly simplified and scalable version of IntServ, that is well supportive to large flows through aggregation, per-hop behavior (PHB) definition.

Although IntServ and DiffServ have been used for several years, but their restriction on signaling protocols issues and network topology were as obstacle for their accommodation on some networks. However, many advanced QoS architectures (e.g. Quasi-stateful flow-based) have been proposed for different technologies and environment to reduce the complexity of them.

Quasi-Stateful Flow-based Architecture: Scalable core (SCORE), or core-stateless fair queuing (CSFQ) (Stoica et al., 2003) is one of the well enhanced quasi-stateful flow-based architecture (also known as stateful DiffServ). SCORE emulates IntServ based on the state information written in a packet header. Unlike IntServ, the main idea of SCORE is to have packets carry per-flow state, instead of having core routers maintain per-flow state. A tremendous amount of work, MCSFQ (Pelsser and De Cnodder, 2002), ACSFQ (Nabeshima, 2003), CSPFQ & WCSFQ (Cheng et al., 2004), etc., have been proposed based on SCORE using the common concept of Fair Allocation Derivative Estimation (FADE) (Li et al., 2000). The summary of these Architectures are presented in Figure 2.4 and Table 2.1.

Other proposed flow-based QoS management architectures for the next generation network are Flow aware Network (FAN) proposed and utilized by France Telecom (Oueslati and Roberts, 2005, 2006), Flow-state-aware network (FSA) utilized by British-Telecom & Korea-Electronics and Telecommunications Research Institute (ETRI), and it has been recently approved in Recommendation Y.2121 in January 2008, at the ITU-T NGNGSI SG13 meeting held in Seoul; Moreover, Flow Aggregate-based Services (FAbS) proposed by ETRI (Joung et al., 2008). A diagram-based review of mentioned QoS architectures can be found in Figure 2.4 and Table 2.1.

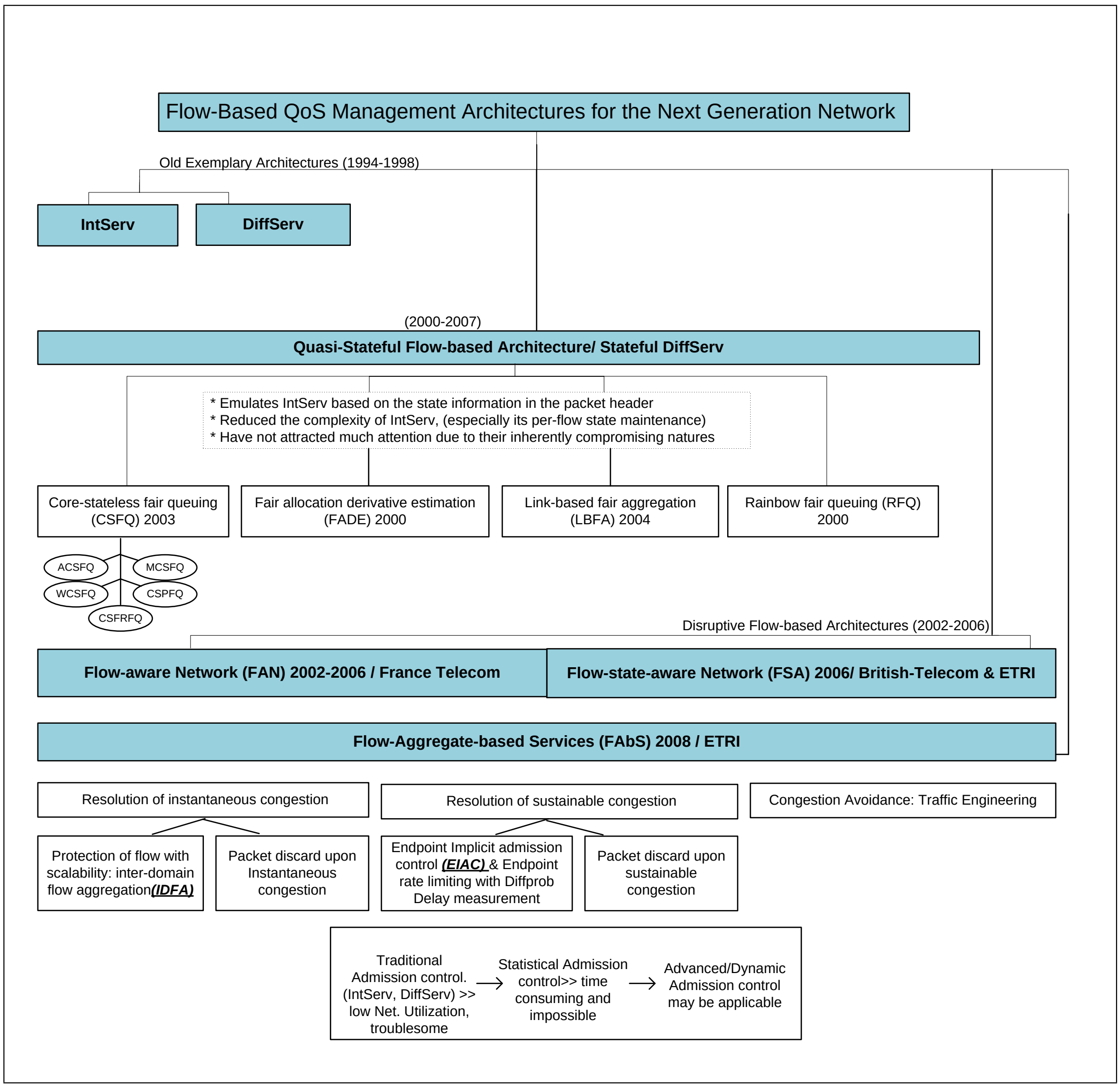
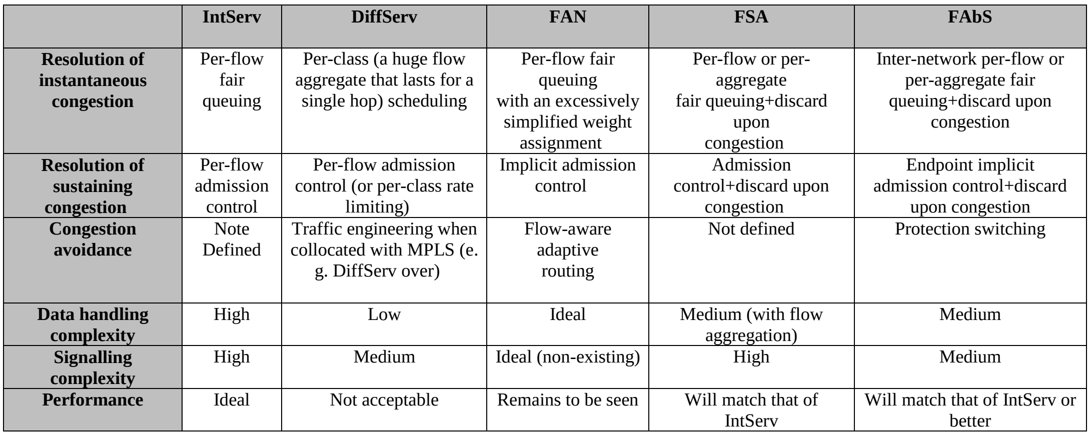

Flow-Based QoS Management Architectures for the Next Generation Network
Old Exemplary Architectures (1994-1998)
IntServ
DiffServ
(2000-2007)
Quasi-Stateful Flow-based Architecture/ Stateful DiffServ
* Emulates IntServ based on the state information in the packet header
* Reduced the complexity of IntServ, (especially its per-flow state maintenance)
* Have not attracted much attention due to their inherently compromising natures
Core-stateless fair queuing (CSFQ) 2003
Fair allocation derivative estimation (FADE) 2000
Link-based fair aggregation (LBFA) 2004
Rainbow fair queuing (RFQ) 2000
ACSFQ
MCSFQ
WCSFQ
CSPFQ
CSFRFQ
Disruptive Flow-based Architectures (2002-2006)
Flow-aware Network (FAN) 2002-2006 / France Telecom
Flow-state-aware Network (FSA) 2006/ British-Telecom & ETRI
Flow-Aggregate-based Services (FAbS) 2008 / ETRI
Resolution of instantaneous congestion
Resolution of sustainable congestion
Congestion Avoidance: Traffic Engineering
Protection of flow with scalability: inter-domain flow aggregation(IDFA)
Packet discard upon Instantaneous congestion
Endpoint Implicit admission control (EIAC) & Endpoint rate limiting with Diffprob Delay measurement
Packet discard upon sustainable congestion
Traditional Admission control. (IntServ, DiffServ) >> low Net. Utilization, troublesome
Statistical Admission control>> time consuming and impossible
Advanced/Dynamic Admission control may be applicable

Figure 2.4: Flow-Based QoS Management Architectures for the Next Generation Network

Table 2.1: Comparison of Existing QoS Architectures (Joung et al., 2008)



	IntServ	DiffServ	FAN	FSA	FAbS
Resolution of instantaneous congestion	Per-flow fair queuing	Per-class (a huge flow aggregate that lasts for a single hop) scheduling	Per-flow fair queuing with an excessively simplified weight assignment	Per-flow or per-aggregate fair queuing+discard upon congestion	Inter-network per-flow or per-aggregate fair queuing+discard upon congestion
Resolution of sustaining congestion	Per-flow admission control	Per-flow admission control (or per-class rate limiting)	Implicit admission control	Admission control+discard upon congestion	Endpoint implicit admission control+discard upon congestion
Congestion avoidance	Note Defined	Traffic engineering when collocated with MPLS (e. g. DiffServ over)	Flow-aware adaptive routing	Not defined	Protection switching
Data handling complexity	High	Low	Ideal	Medium (with flow aggregation)	Medium
Signalling complexity	High	Medium	Ideal (non-existing)	High	Medium
Performance	Ideal	Not acceptable	Remains to be seen	Will match that of IntServ	Will match that of IntServ or better

2.3.2 QoS Building Blocks

Considering the QoS building blocks in Figure 2.3 for packet technologies, can claim that this research mainly focused on *data plane* mechanisms that deal directly with user traffic in a typical Metro Ethernet. And particularly, those mechanisms are included: *traffic classification and packet marking*; *traffic shaping and policing*; ***buffer management***, ***queuing scheduling*** and ***flow/congestion control***. It is expected that, by enhancing the performance of data plane can develop provisioning methods and guidelines. These guidelines can be used for network planning and exploiting the *management plane* and *control plane* functionality.

A multi-layer switch operates in two operational planes: *control plane* and *data plane* (Evans and Filsfils, 2010) and has to perform several functions beside simply switching/routing packets from its inputs to its outputs Gebali (2008), namely: Switching or Routing, Traffic Management, Queuing Scheduling, and Congestion Control.

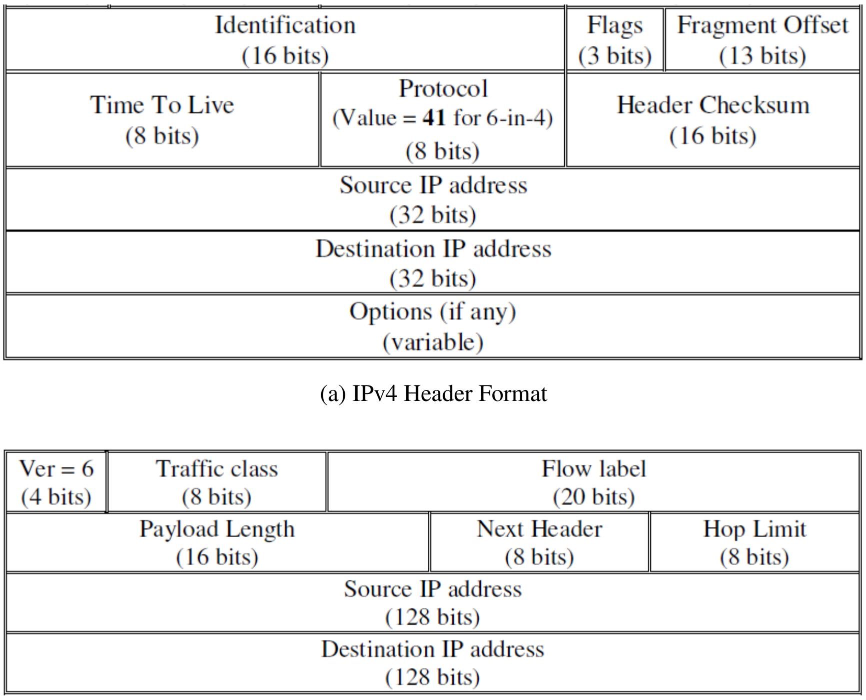
The main function of flow and congestion control is fair allocation of resources among competing users. By employing *Buffer Management and Queueing Scheduling algorithms* in a switch can *i)* provide different QoS to the different types of users and *ii)* protect well-behaved users from misbehaving users that might hog the system resources (bandwidth and buffer space). It should be highlighted that the most costly resource in a network is buffer memory and the second most costly resource is physical channel bandwidth. A queueing scheduling algorithm might operate on a per-flow basis or it could aggregate several users into broad service classes to reduce the workload. Hence, the scheduling goals are typically bandwidth, latency, and jitter control (Gai and DeSanti, 2009). On the other hand, the selection and ordering of arrived data packets for transmission on the outgoing link is the scheduling function that is necessary to control network resources (bandwidth).

In next generation networks with converged applications, each traffic flow has different QoS requirements in terms of allocated bandwidth, maximum delay, jitter and packet loss. Service differentiation is thus a crucial feature to provide QoS provisioning. Among the common QoS support features (Figure 2.3), having an efficient queuing scheduling algorithm which coordinate all other QoS-related functional entities is critical.

Although varieties of queuing scheduling algorithm such as well-known Round Robin (RR), Weighed Round Robin (WRR), Weighted Fair Queuing (WFQ), Strict Priority Queing (PQ), (Katevenis et al., 1991; Demers et al., 1989; Shreedhar and Varghese, 1995), in addition to other developed scheduling have been proposed in many literatures; nonetheless, there was no research study for evaluating the impact of scheduling mechanism in interaction with Ethernet flow control targeting QoS provisioning. Consequently, there was a lack of study on a differentiated prioritized scheduling mechanism of multiple traffic classes, multiple converged traffic type, and or different packet size in simulating a real network situation for Metro Ethernet QoS which may resulted in unfairness and starvation of traffic.

2.3.3 IPv6 and QoS Provisioning

A new version of the Internet Protocol has been designed by IETF, known as IPv6 (Bhatia et al., 2012), in order to address the scalability and service shortcomings of the current IPv4



<table>
<tr><td colspan="2">Identification
(16 bits)</td><td>Flags
(3 bits)</td><td>Fragment Offset
(13 bits)</td></tr>
<tr><td>Time To Live
(8 bits)</td><td>Protocol
(Value = <b>41</b> for 6-in-4)
(8 bits)</td><td colspan="2">Header Checksum
(16 bits)</td></tr>
<tr><td colspan="4">Source IP address
(32 bits)</td></tr>
<tr><td colspan="4">Destination IP address
(32 bits)</td></tr>
<tr><td colspan="4">Options (if any)
(variable)</td></tr>
</table>

(a) IPv4 Header Format

<table>
<tr><td>Ver = 6
(4 bits)</td><td>Traffic class
(8 bits)</td><td colspan="3">Flow label
(20 bits)</td></tr>
<tr><td colspan="3">Payload Length
(16 bits)</td><td>Next Header
(8 bits)</td><td>Hop Limit
(8 bits)</td></tr>
<tr><td colspan="5">Source IP address
(128 bits)</td></tr>
<tr><td colspan="5">Destination IP address
(128 bits)</td></tr>
</table>

(b) IPv6 Header Format

Figure 2.5: IPv4 and IPv6 Header Format

protocol. Figure 2.5 illustrates the IPv4 and IPv6 header formats.

As shown in the previous Section 2.3.1, IntServ and DiffServ are the most utilized QoS architectures for the Internet, while packet classification is a key element in the implementation of both (Tang et al., 2003). An attempt in IPv4 to classify traffic according to a Type of Service (ToS) byte in the IPv4 header did not succeed Internet-wide because the ToS byte was based on fair self-classification of applications with respect to other application traffic and hence, ToS byte was never used widely. On the other hand, as proved by many conventional researches (Fgee et al., 2003, 2004; Tang et al., 2003; Prakash, 2004) as well as recent researchers (Fgee et al., 2008, 2010; Wang et al., 2010) in the field of IPv6 QoS, the new added 20 bits flow label field of IPv6 header can be used to provide various enhanced QoS by an efficient way for packet marking, flow identification and flow state lookup.

The 20 bits Flow Label field in the IPv6 header is used by a source to label packets of a flow. A Flow Label of zero is used to indicate packets not part of any flow. Packet classifiers use the triplet of Flow Label, Source Address, and Destination Address fields to identify which flow a particular packet belongs to (Rajahalme et al., 2011). The usage of the Flow Label field enables efficient IPv6 flow classification based only on IPv6 main header fields in fixed positions (Davies, 2010).

There are various proposals made to the IETF for the Flow Label utilization (Conta and Carpenter, 2001; Conta, 2001; Jagadeesan and Singh, 2002; Banerjee et al., 2002). However, based on the intensive review and comparison of various IPv6 flow label formats for E2E QoS provisioning (Prakash, 2004; Ahmed et al., 2009; Hizwan and Aziz, 2011; Hu and Carpenter, 2011); the *Hybrid Approach of IPv6 flow label format* (Banerjee et al., 2002) is chosen as the best performed approach to be customized in this research study. Since the *Hybrid Approach* can provide the sufficient classification information, unique flow identification, efficient classification mechanism and E2E support across the Internet. Figure 2.6 maintains two types of flow label format: the conventional approach that flow label is just assigned at random as an identifier of flows with no QoS provisioning; and the Hybrid approach, in where flow label fields is divided into several meaningful fields for QoS provisioning. Thus, using the hybrid approach, QoS requirements of different traffic applications are embedded in IPv6 header of each traffic packet.

Therefore, by using Hybrid approach Flow Label is used as an identifier as well a field that provides QoS information to the intermediate nodes. Advantage of IPv6 Flow label for a QoS traffic classier: in where some information used to identify the flow is missing due to fragmentation, encryption or tunneling, the flow label helps to unambiguously classify a flow (Evans and Filsfils, 2010). Consequently, it causes less processing overhead and fast forwarding. The general rules for IPv6 Flow Label use along with more details about the Flow Label field definition, specification and its requirements are presented in latest RFC 6437 (Rajahalme et al., 2011).

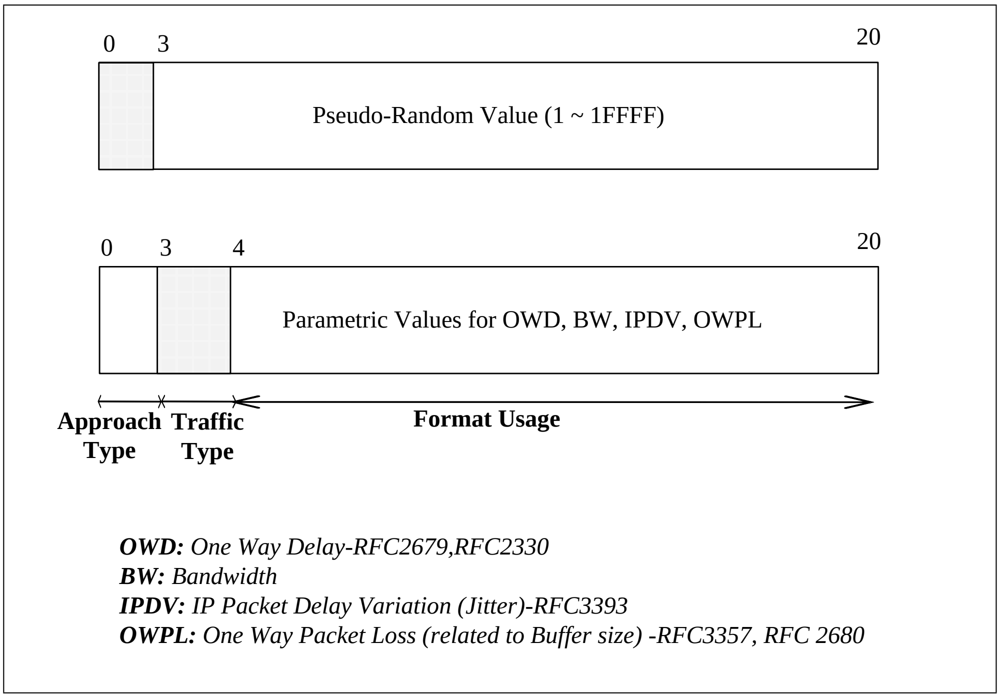

0
3
20
Pseudo-Random Value (1 ~ 1FFFF)
0
3
4
20
Parametric Values for OWD, BW, IPDV, OWPL
Approach Type
Traffic Type
Format Usage
OWD: One Way Delay-RFC2679,RFC2330
BW: Bandwidth
IPDV: IP Packet Delay Variation (Jitter)-RFC3393
OWPL: One Way Packet Loss (related to Buffer size) -RFC3357, RFC 2680

Figure 2.6: Two Types of IPv6 Flow Label Format (Random Versus Hybrid Parametric)

Therefore, IPv6 header along with flow label field provide easier classification of packets with identifiers of traffic (Parra et al., 2011). Additionally, the flow label field has the advantage of being located before the address fields, and hence, can reduce the time of verification and average processing load of the switches in the network, and therefore, reduces E2E delays of the packets.

2.4 Flow and Congestion Control

2.4.1 Flow Control Fundamental and Taxonomy

Generally, there are four levels of flow control that can be exercised at various levels of OSI in a packet network. These four levels that are closely related to the protocol levels are hop (or node-to-node) level, entry-to-exit level, network access level and transport level (Gerla and Kleinrock, 1980). There are varieties of flow control protocols (Tanenbaum, 2002; Behrouz and Sophia, 2003) in transport layer as well as data link layer that are shown in Figure 2.7. The initial IEEE 802.3x were implemented by vendors in the form of none (not able to do any

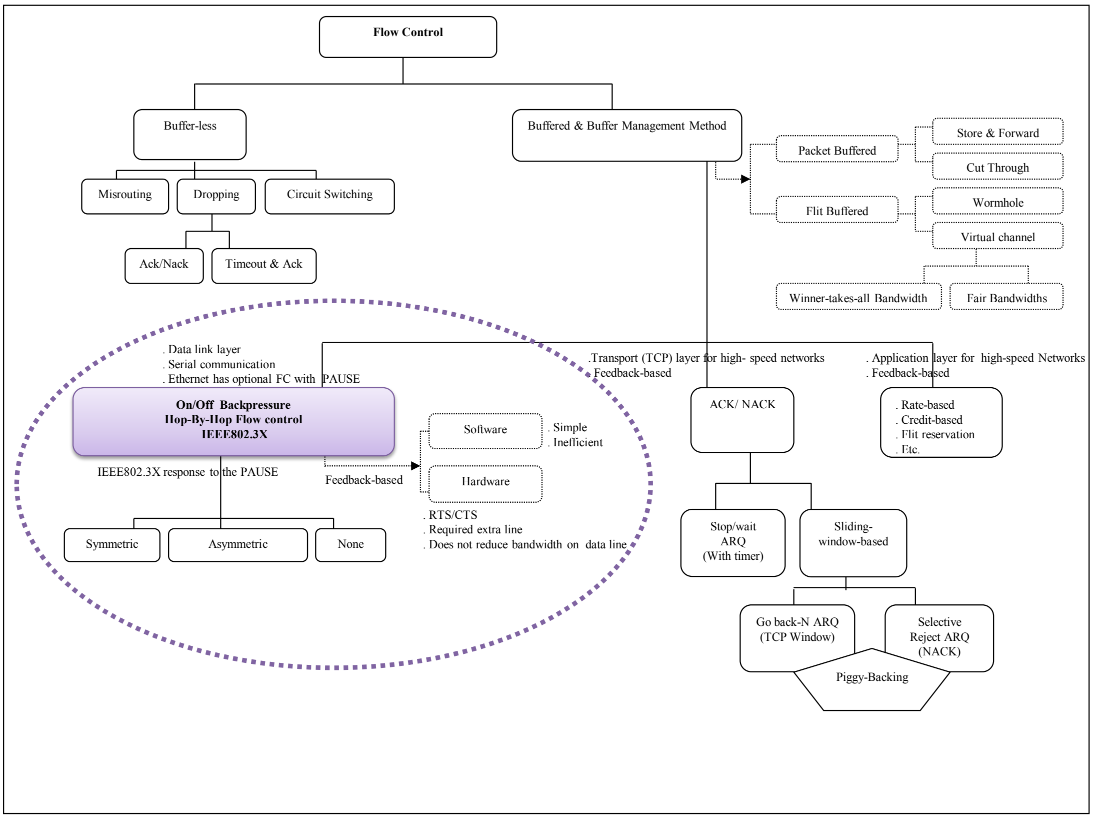


Figure 2.7: Taxonomy of Flow Control Protocol Classification

response), asymmetric (able to transmit or response to the PAUSE, and preferably response), or symmetric (able to transmit & response to the PAUSE) (Dally and Towles, 2004). Consequently, actual networks may not always mechanize all of the above four levels of flow control with distinct procedures. It is quite possible, for example, for a single flow control mechanism to combine two or more levels of flow control. On the other hand, it is possible that one or more levels of flow control may be missing in the network implementation.

Some authors preserve the term ***flow control*** for the transport level, and refer to the other levels of control as congestion control. This terminology is used to emphasize the physical distinction between the first three levels and the fourth level. However, this research study has chosen to use the term flow control for the second layer also, and it may be used interchangeably with congestion control. Any receiving network device has a limited speed (at which it can process incoming data) and a limited amount of memory (in which to store incoming data).

Before these restrictions are reached the receiving device must be able to inform the source device and request to halt transmission until it is once again able to receive. So the purpose of the flow control mechanism is to place the data transfer at an acceptable speed and to solve the incompatibility of the speed of transmission from a fast sender to a slow receiver or on the other hand managing the rate of transmission data between two network devices. However, flow/congestion control has a significant role in the performance of computer networks. In general, the main functions of flow control are deadlock avoidance, prevention of throughput degradation and loss of efficiency due to overload, fair allocation of resources among competing user, and speed matching between network and its attached users; in order to manage the data transfer at an acceptable speed with high bandwidth utilization (high throughput) and low rate of packet loss.

In general, different flow and congestion control techniques have been proposed for computer networks (Vijayaraja and Hemamalini, 2010). The congestion control mechanisms all have the same basic objective: they all try to detect congestion, notify the other nodes of the congestion status, and reduce the congestion and or its impact using rate adjustment algorithms (Wang et al., 2006; Yaghmaee and Adjeroh, 2008b, 2009). As depicted in Figure 2.8, a generic architecture of any flow control protocol is based on three main components namely, Reaction, Notification/Action, and Congestion Detection.

According to the scope of this research, and as shown in Figure 2.7 the two common flow control protocols are E2E Flow and congestion control as used in TCP and similar protocols and the other one is HBH flow control that are discussed in the following Sections.