Quality of Service Guarantee in the Edge Node of Optical Packet/Burst Switched Networks with Traffic Assembly

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

vorgelegt von

Guoqiang Hu

geb. in Zhouchen, Provinz Shandong, China

Hauptberichter:	Prof. Dr.-Ing. Dr. h. c. mult. Paul J. Kühn
Mitberichter:	Prof. Dr. Javier Aracil, Universidad Autónoma de Madrid
Tag der Einreichung:	22. Januar 2008
Tag der mündlichen Prüfung:	24. Juli 2009

Institut für Kommunikationsnetze und Rechnersysteme
der Universität Stuttgart

2009

To my parents

Abstract

With the fast development of heterogeneous network services and applications as well as the increasing bandwidth demand, the future transport network is confronted with the requirements for flexible and dynamic service provisioning, high throughput and assurance of quality of service (QoS). To deal with the challenges, photonic packet switching aims to realize a fast switch of the traffic in the optical domain in a packet-by-packet manner. The fine granularity in the switching is flexible for the support of dynamic service requirements and allows for an efficient resource utilization by taking advantage of the statistical multiplexing gain. In last years, relatively large advancements have been made in the technologies for the fast configurable optical switch fabric. The switching control unit (SCU), however, will keep relying on electronic technologies in the foreseeable future due to the unsolved technological problems in e.g., optical buffering and optical signal processing. Optical packet switching (OPS) and optical burst switching (OBS) are two representative network architectures for the photonic packet switching.

Because of the limitation in the realization technologies, QoS problems in OPS and OBS networks differ from those in conventional store-and-forward packet switched networks. Typically, due to the deficiency in the optical buffering capability, the optical switch fabric must switch on the fly, which leads inevitably to data loss. The loss performance in OPS/OBS core networks has been intensively studied. Furthermore, to alleviate the processing overhead in the SCUs, large optical data frames are suggested for the OPS/OBS core network. To this end, the client traffic needs to be assembled in the edge node before being transmitted through the core network. The traffic assembly procedure causes additional delay and alters the traffic characteristics. Its influence on the end-to-end (E2E) QoS provisioning must be elaborately analyzed.

This dissertation studies the QoS provisioning in the edge node of OPS and OBS networks. It models, analyzes and evaluates the link-layer performance of the edge node and provides the solution to the admission control for QoS guarantee. In the beginning, the general OPS/OBS network architectures are introduced. Especially, the schemes for the QoS differentiation and assurance with respect to the E2E loss performance for core networks are surveyed and classified. Then, the relevant work for the edge node is discussed with the focus on the traffic assembly and scheduling. The edge node's tasks in controlling the optical header rate and local delay are highlighted.

Before proceeding to solve the QoS problems, the time-scale-dependent traffic characteristics in today's Internet backbone are briefly introduced. To deal with such kind of traffic behaviors as well as further complex traffic patterns introduced by the traffic assembly, a novel approximate method is proposed for the multi-scale queueing analysis on the basis of the time scale

decomposition and the concept of the relevant time scale. An example study in combination with simulations verifies the validity and accuracy of the method.

The performance analysis concentrates on the quantitative determination of the assembly degree and the transmission queueing delay in the edge node. For the former case, a closed-form solution is derived for an approximate estimation of the mean frame size after the assembly, from which the optical header rate can be calculated. In the delay analysis for the transmission buffer, the characteristics of the assembled traffic are first analyzed on small time scales and large time scales, respectively. On this basis, the transmission queueing performance is solved by the proposed multi-scale queueing analysis. Connecting these analytical results together, a comprehensive performance evaluation is carried out for the edge node. It is shown how the combined QoS requirements to constrain the header rate and to keep the local delay bound impose the restrictions on the traffic throughput. An admission control algorithm is further proposed based on the derived quantitative relationship between the traffic profile, the system parameters and the QoS specification. Integrating those QoS models proposed for the core OPS/OBS network, this provides a complete solution to guarantee the E2E delay, the timely header processing in SCUs and the loss performance on the data path.

Kurzfassung

Durch die weite Verbreitung des Internets werden heute wachsende Anforderungen an die Transportnetze im Hinblick auf die Dienste so wie die Bandbreite gestellt. Zukünftige Transportnetze sollen einen sehr hohen Durchsatz ermöglichen und die Dienste dynamisch bereitstellen können, um die heterogenen Dienstanforderungen der oberen Schichten möglichst effizient zu unterstützen. Ebenfalls wichtig für ein Transportnetz ist auch die Fähigkeit, eine Dienstgüte (QoS) zu gewährleisten. Eine vielversprechende Lösung dazu stellt die photonische Paketvermittlung dar. Das Grundprinzip dabei ist, in optischen Netzen jeden einzelnen Datenrahmen transparent (d.h. ohne O/E/O-Wandlung) zu vermitteln. Die feine Granularität der Vermittlung ermöglicht eine flexible Bereitstellung von Diensten sowie eine bessere Ausnutzung der Ressourcen anhand des statistischen Multiplex. In letzter Zeit wurden große technologische Fortschritte erzielt, um eine schnell-konfigurierbare optische Vermittlungsmatrix zu realisieren. Wegen ungelöster Probleme der optischen Signalspeicherung/-Verarbeitung, beruht die Steuereinheit des Switches jedoch weiter auf elektronischen Technologien. *Optical Packet Switching* (OPS) und *Optical Burst Switching* (OBS) sind die zwei bekanntesten Netzarchitekturen der photonischen Paketvermittlung.

Wegen technologischen Einschränkungen unterscheiden sich die Dienstgüte-Probleme in OPS/OBS Netzen von jenen in den herkömmlichen paketvermittelten store-and-forward Netzen. Aufgrund des Mangels an optischen Speichern muss die Durchschaltung der Vermittlungsmatrix mit der Ankunftszeit eines Datenrahmens synchronisiert werden. Dies ähnelt dem Verhalten der Leitungsvermittlung und kann zur Blockierung sowie Datenverlust führen. Die Leistungsfähigkeit der OPS/OBS-Kernnetze wurde hinsichtlich der Blockierungswahrscheinlichkeit bereits intensiv untersucht. Um den Verarbeitungsaufwand der Steuereinheit eines Switches zu reduzieren, sind üblich große Datenrahmen für OPS/OBS Netze vorgesehen. Einkommender Verkehr aus den Kundennetzen wird zuerst im Randknoten in Datenrahmen assembliert, bevor er in das Kernnetz weitergeleitet wird. Die Verkehrsassemblierung verursacht zusätzliche Verzögerungen und ändert die Verkehrscharakteristik, deren Einfluss auf die netzweite (E2E) Dienstgütegarantie eingehend untersucht werden muss.

Diese Dissertation behandelt die Dienstgütegarantie im Randknoten von OPS/OBS Netzen. Die Leistungsfähigkeit des Randknotens wird modelliert und analysiert. Weiter wird ein Verfahren zur Zugangskontrolle entworfen. Zunächst werden die allgemeinen OPS/OBS Netzarchitekturen eingeführt. Die Mechanismen zur QoS-Differenzierung und -Gewährleistung mit Bezug auf die Blockierung im Kerznetz werden besonders untersucht und klassifiziert. Bekannte Forschungsergebnisse hinsichtlich der Verkehrsassemblierung und des -Scheduling im Randknoten werden vorgestellt und diskutiert. Die Aufgaben des Randknotens in der Raten-

kontrolle der Rahmenköpfe und in der Beschränkung der lokaler Verzögerung werden hervorgehoben.

Eine systematische Leistungsbewertung für den Randknoten setzt eine gültige Verkehrsmodellierung und eine angemessene Untersuchungsmethode voraus. Der Verkehr heutiger IP-Backbonenetze besitzt komplexe Struktur auf verschiedenen Zeitskalen. Besonders in einem Randknoten kann das Assemblierungsverfahren zu weiteren Verkehrsmustern führen. Um solche Verkehrscharakteristika zu behandeln, wird eine neue analytische Methodik vorgeschlagen, die auf den *time scale decomposition* und *relevant time scale* Prinzipien beruht. Durch Simulationsstudien werden die Gültigkeit und die Genauigkeit der Methodik verifiziert.

Die Leistungsbewertung konzentriert sich auf die quantitative Bestimmung zweier Metriken: (1) die aus der Assemblierung resultierende Rate der Rahmenköpfe; (2) die Wartezeit der Datenrahmen im Sendepuffer. Dazu wird zunächst der Mittelwert der Rahmengröße anhand der Assemblierungsparameter abgeleitet, aus dem die Rahmenkopf-Rate direkt ausgerechnet werden kann. Zur Bewertung der Wartezeit wird zuerst der assemblierte Verkehr analysiert und seine Charakteristik auf kleinen und großen Zeitskalen identifiziert. Die Wartezeit wird durch die vorher vorgeschlagene Methodik analysiert. Anhand der analytischen Ergebnisse wird eine gesamte Leistungsbewertung für den Randknoten durchgeführt. Der Einfluss der QoS Anforderungen bezüglich der Rahmenkopf-Rate und der Verzögerung auf den Durchsatz wird hervorgehoben. Weiterhin wird ein Algorithmus zur Zugangskontrolle abgeleitet. Der Algorithmus integriert das QoS-Modell für den Randknoten und die bekannten QoS-Lösungen für das Kernnetz. Die Integration der QoS-Modelle entspricht einer vollständigen Lösung zur E2E Dienstgütegarantie in OPS/OBS Netzen.

Contents

List of Figures

List of Tables

Abbreviations and Symbols

Abbreviations

ARIMA	Autoregressive Integrated Moving Average
ATM	Asynchronous Transfer Mode
AWG	Arrayed Waveguide Grating
BHP	Burst Header Packet
bps	bit per second
CBR	Constant Bit Rate
CCDF	Complementary Cumulative Distribution Function
COV	Coefficient of Variation
CT	Continuous Time
DiffServ	Differentiated Services
DRR	Deficit Round Robin
DT	Discrete Time
DWDM	Dense Wavelength Division Multiplexing
DWG	Dynamic Wavelength Grouping
E2E	End-to-End
EDF	Earliest Due First
FBM	Fractional Brownian Motion
FCFS	First-Come, First-Served
FDL	Fiber Delay Line
FEC	Forwarding Equivalence Class

FGN	Fractional Gaussian Noise
FIFO	First-In, First-Out
FRWC	Full-Range Wavelength Converter
GCRA	Generic Cell Rate Algorithm
GII	Global Information Infrastructure
GPS	Generalized Processor Sharing
i.i.d.	independent and identically distributed
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JET	Just Enough Time
JISP	Joint Interval Selection Problem
LAUC	Latest Available Unscheduled Channel
LAUC-VF	Latest Available Unused Channel with Void Filling
LRD	Long Range Dependence
LRWC	Limited-Range Wavelength Converter
MMPP	Markov Modulated Poisson Process
MVA	Maximum Variance Asymptotic
NRZ	Non-Return-to-Zero
OBS	Optical Burst Switching
OCDMA	Optical Code Division Multiplexing
OCS	Optical Circuit Switching
OEO	Optical-to-Electrical-to-Optical
OPS	Optical Packet Switching
OTN	Optical Transport Network
PDF	Probability Density Function
PHB	Per-Hop Behavior

pps	packet per second
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RED	Random Early Detection
RV	Random Variable
SCFQ	Self-Clocked Fair Queueing
SCM	Subcarrier Multiplexing
SCU	Switch Control Unit
SDH	Synchronous Digital Hierarchy
SLA	Service Level Agreement
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical Network
SRD	Short Range Dependent
TCP	Transmission Control Protocol
TWC	Tunable Wavelength Converter
VCR	Virtual Channel Reservation
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
WF2Q	Worst-case Fair Weighted Fair Queueing
WFQ	Weighted Fair Queueing
WRR	Weighted Round Robin
WS-MinMax	Wavelength Sharing with Minimum Provisioning and Maximum Occupancy
XGM	Cross Gain Modulation
XOR	exclusive-OR
XPM	Cross Phase Modulation

Symbols

$A(s)$	Cumulative traffic arrival process: traffic amount arrived up to time instant s
$\arg\inf_t f(t)$	Operator to get the value of t which leads to the infimum of $f(t)$
B	Traffic volume of a session in an M/Pareto model
c	Transmission rate of a channel
c_{a}	Access link rate in an M/Pareto model
c_X	Coefficient of variation of random variable X
D	Inter-departure time of data frames
$E[\cdot]$	Mean value of a random variable
$g(q,t)$	An intermediate function used to define the relevant time scale in the MVA method
H	Hurst parameter
I	Packet interarrival time of client traffic
I_i	Interarrival time of the i-th packet in a data frame
I_{res}	Residual packet interarrival time of client traffic
$\inf_\theta f(\theta)$	Operator to get the infimum of function $f(\theta)$ over the range of θ defined on the subscript
L	Packet length in bytes
L_i	Packet length of the i-th packet in a data frame
$l_{\max}$	Maximal packet length in bytes
m_i	Number of wavelengths occupied by class i instantaneously
m_{link}	Total number of wavelengths per link
$m_{\max,i}$	Maximal number of wavelengths that can be occupied by service class i at one time
$m_{\min,i}$	Guaranteed number of wavelengths that can be occupied by service class i at one time
$\max(a,b)$	Function that returns the maximum of a and b
$N(\mu,\sigma^2)$	Gaussian (Normal) distribution with mean μ and standard deviation σ
N_{b}	Number of packets in a data frame

n	Number of FEC flows in the edge node
$P\{\cdot\}$	Probability of an event
P_i^{L}	Loss probability at switching hop i
$P_{\mathrm{EE}}^{\mathrm{L}}$	E2E loss probability in the network
P_{EE}^{*}	Maximal E2E loss probability specified as a QoS requirement
P_i^{*}	Loss probability specified as a QoS requirement at switching hop i
$p_t(u)$	PDF function of traffic amount in an arbitrary time interval of t
Q	Unfinished work in the buffer in bytes
$Q(s)$	Unfinished work in the buffer at time instant s
Q_{B}	Burst component for the unfinished work Q
Q_{b}	Unfinished work in the buffer in number of data frames
Q_{C}	Cell component for the unfinished work Q
Q_{p}	Unfinished work in the buffer in number of packets
r	General notation for average traffic rate
$r(k)$	Autocorrelation function of lag k
r_{dat}	Data rate of one FEC
$r_{\mathrm{dat},i}$	Data rate of FEC i
$r_{\mathrm{dat},i}^{*}$	Sustainable data rate of FEC i on the E2E path
$r_{\mathrm{req},i}$	Maximal desired data rate specified in the request of FEC i
r_{sig}	Rate of frame headers for one FEC
$r_{\mathrm{sig},i}$	Rate of frame headers for FEC i
$r_{\mathrm{sig},i}^{*}$	Sustainable rate of frame headers for FEC i on the E2E path
S_{b}	Data frame size
$S_{\mathrm{b}}^{\mathrm{ptime}}$	Data frame size under the pure time-based assembly scheme
s_{th}	Size threshold in size-based assembly scheme
$\sup_t f(t)$	Operator to get the supremum of function $f(t)$ over the range of t defined on the subscript
T_{a}	Assembly duration of a data frame
$T_{\mathrm{a}}^{\mathrm{psize}}$	Assembly duration of a data frame under the pure size-based assembly scheme

t	Time instant, time interval or time scale according to the context
t_{th}	Timeout period in time-based assembly scheme
$t_{\text{th},i}$	Timeout period of the assembler for FEC i
U_t	Amount of traffic arrival (in bytes) in an arbitrary time interval of t
V_t	Number of frame arrivals in an arbitrary time interval of t
VAR$[\cdot]$	Variance of a random variable
W	Queueing delay in the buffer
W_{b}	Queueing delay normalized by frame transmission time
W_{p}	Queueing delay normalized by packet transmission time
X_t	Traffic volume measured in an arbitrary time interval of t
z	Total number of hops on an E2E path
α	Shape parameter of the Pareto distribution in an M/Pareto model
β	Asymptotic constant in the exponential approximation of the tail probability of a queue
$\Gamma(\omega,\nu)$	Gamma distribution with shape parameter ω and scale parameter ν
Δ_i	Transit delay in the SCU of the i-th hop in an OBS network
δ^*	Delay budget for the edge node
δ_i^*	Delay budget for FEC i in the edge node
δ_{W}	Statistical bound on the queueing delay W
ε	Transmission duration of a single data unit
η_t	Residual of t divided by E[D]
κ	Minimal traffic volume of a session in an M/Pareto model
$\Lambda(\theta,t)$	Effective bandwidth with system parameter θ and t
λ	Packet arrival rate at an assembly buffer
λ_i	Packet arrival rate at the assembly buffer of FEC i
λ_{s}	Session arrival rate in an M/Pareto model
ξ_u	Residual of $U_t = u$ divided by s_{th}

ρ	Offered load of the system
τ_q	Relevant time scale for $P\{Q > q\}$
$\Phi(u)$	Distribution function of the standard Gaussian distribution $N(0,1)$
ϕ	Mean frame interarrival time normalized by frame transmission time
φ	Mean traffic volume of a session in an M/Pareto model

1 Introduction

The last decade witnessed the large success of the Internet, which had and continues to have its extensive and profound influence on today's information society. Thanks to the flexible protocol stack on the basis of the Internet protocol (IP), the Internet is capable of accommodating a variety of network services. This greatly stimulates the development of new network applications and at the same time motivates the replanting of the conventional voice service from circuit-switched networks to IP-based solutions. The tendency is apparent that the IP layer is becoming a convergence layer in the global information infrastructure (GII). In the upward direction of the protocol stack, it provides a uniform network layer for heterogeneous applications and the respective upper-layer protocols. In the downward direction, a dynamic transport network architecture is preferable in order to support the IP-based client networks in a flexible and efficient manner.

The current transport networks have evolved from telephone networks that were inherently designed for connection oriented services. Electronic circuit switching technologies, with the representatives of the synchronous digital hierarchy (SDH) and the synchronous optical network (SONET) [BC89, RS02], realize the traffic grooming and switching in terms of the hierarchical time division multiplexing/demultiplexing. They generally require accurate network-wide synchronization and expensive optical-to-electrical-to-optical (OEO) equipments in switching nodes. With the introduction of the dense wavelength division multiplexing (DWDM), the transmission bandwidth of individual optical fibers was significantly increased. Optical circuit switching (OCS), also called light path switching, was further developed to realize all optical end-to-end (E2E) connections by means of wavelength paths [NR01, Ram02]. Such a hierarchical structure of the circuit switched transport network leads to a high complexity in the network configuration/maintenance and imposes large control/management overhead, which is inflexible and redundant from the perspective of the IP layer if it is applied to individual narrowband services.

This contradiction motivated the development of optical transport networks (OTN) that switch and route in a finer granularity than that of the circuit switching. Typically, the traffic is switched/routed per individual data frames, following the basic concept of packet switching. Taking advantage of the statistical multiplexing, the network bandwidth is more efficiently utilized for the IP traffic that exhibits a high degree of traffic dynamic. To match with the huge transmission bandwidth of fiber links, however, the OTN nodes must realize a very high switching speed. Electronic packet switching is the mature technology, which can currently support the line rate up to 80 Gbps or even above. In the long term, the photonic packet switching has more potentials by providing for transparent data switching in the optical domain.

1.1 Photonic Packet Switching

Photonic packet switching adopts optical switch fabric that can be fast configured to switch the traffic in the form of optical data frames. Due to the deficiency in the buffering technology for optical signals, the photonic packet switching follows the style of switching on the fly in contrast to the conventional store-and-forward packet switching. This means that the control unit of the switch has to keep track of a good timing so that a forwarding route within the switch fabric is configured before the arrival of the data. The implementation of the control part of the switch will keep relying on electronic technologies in the foreseeable future, because the practical realization of the optical buffering and signal processing is still confronted with many technological challenges. With electronic devices, the processing speed of the header information in the switching/routing decision for individual data frames stands for a crucial performance constraint. To alleviate this problem, in most of the proposed network architectures, large data frames are applied to reduce the header processing overhead. Correspondingly, the client traffic, e.g., flows of IP packets, needs to be assembled/disassembled to/from OTN data frames at the ingress/egress of the OTN. This function can be included into a light-weight adaption layer between the IP layer and WDM layer in the protocol stack. In general, the network architectures with the photonic packet switching have a much flattened layering structure, which is sometimes in literature highlighted by the notation of the IP-over-WDM solution in comparison to, e.g., the IP-over-SDH-over-WDM, in the circuit switched OTN.

Service provisioning in transport networks is mostly bound to a service level agreement (SLA) that specifies the quality of service (QoS) to be delivered. Photonic packet switched network architecture must be able to provide guaranteed QoS so as to meet with the service requirements from the client layers. Due to the speciality in the applied technologies, QoS problems in photonic packet switched networks differ from those in the conventional store-and-forward packet switched networks. Typically, the limited optical buffering capability makes the data loss become the main performance issue in a switching node. The feature of switching on-the-fly poses a timing requirement on the header processing and the switch configuration. Furthermore, the traffic assembly procedure in the ingress edge node causes additional delay as well as jitter. The altered traffic characteristics by the assembly also have an impact on the network performance. These issues must be integrated into the solution to the E2E QoS provisioning.

This dissertation inspects the QoS model for the two most representative network architectures of the photonic packet switching, i.e., the optical packet switching (OPS) and the optical burst switching (OBS). These two architectures share a lot of similarities and can be treated equivalently on many points in the QoS study. The focus is placed on the performance and mechanisms in an edge node in the provisioning of an E2E QoS.

1.2 Organization of the Dissertation

In the remainder of the dissertation, an overview on the OPS/OBS network architectures and QoS models is given in Chapter 2. Firstly, the functionality and realization technologies of the core switching node are introduced with respect to the OPS and OBS, respectively, aiming to show the common features and major diversities in between. Then, the issues on the channel

management, contention resolution and scheduling in the switch are discussed, which are directly related to the link-layer network performance. Finally, the QoS mechanisms in individual switching nodes are introduced and the solutions for the E2E guarantee of the loss performance are presented.

In Chapter 3, the edge node of OPS/OBS networks is closely looked at. The interrelation between an edge node and a core switching node is illustrated. By means of an ingress edge node, the traffic assembly is introduced in detail and the relevant work in the traffic scheduling is briefly surveyed. On this basis, the significance of the edge node on the E2E network performance is outlined and the admission problem for services with guaranteed QoS is posed. This motivates the contributions of this dissertation.

For effective admission control, an efficient performance model is essential to estimate the realizable QoS for service requests beforehand. Valid traffic characterization and appropriate analytical methods are of special importance for the performance analysis. In Chapter 4, the traffic characterization in IP backbone networks is reviewed. It is highlighted that the IP traffic exhibits different characteristics on multiple time scales, which must be taken into consideration in selecting the traffic model and analytical method. The M/Pareto model is introduced to synthesize the client traffic in the ingress edge node. Available analytical methods that are able to deal with the time-scale-dependent traffic characteristics are presented, which are classified into two categories: the time scale decomposition approach and the integrated analysis. They all have their own disadvantages and are difficult to be applied for the QoS analysis in the OPS/OBS edge node directly.

By a combined application of the principles of the time scale decomposition and the integrated analysis, a new approach is proposed in Chapter 5 to tackle the traffic having complex characteristics on multiple time scales. It shows that this method is straightforward to be used and provides a closed-form solution to the queueing performance at a low computational overhead. Its validity and accuracy are further demonstrated by an example study.

In Chapter 6, in-depth performance analysis is carried out for the OPS/OBS edge node, on the basis of which an admission control procedure is constructed. Since the traffic assembly has important impacts on the performance in the core network, the quantitative relationship between the assembly parameters and the resulting traffic assembly degree is derived. The assembled traffic is characterized and the performance of the transmission queue is accordingly analyzed by the method introduced in Chapter 5. This further allows for the evaluation of the total delay in the edge node with respect to the node throughput. Summarizing the analytical results, an admission control algorithm is designed and its practical application for E2E service guarantee is discussed.

In Chapter 7, this dissertation is summarized and a prospect of future work is provided.

2 Network Architectures and QoS Provisioning

QoS provisioning in communication networks has been studied for a long time, typically with respect to public switched telephone networks (PSTN), asynchronous transfer mode (ATM) and IP-based networks. Basic QoS mechanisms for service differentiation, admission control, traffic shaping, policing and scheduling etc. have an extensive applicability in different types of networks. On the other hand, the design of an efficient and robust QoS architecture depends very much on the specific network architecture. Special features in the implementation of the optical packet switching (OPS) and the optical burst switching (OBS) lead to new problems in the channel allocation, contention resolution and scheduling, which are closely related to the QoS provisioning. In this chapter, the general OPS/OBS network architectures are introduced and the respective QoS models are surveyed.

In Section 2.1, the fundamental features and design issues of OPS/OBS networks are presented to outline the differences and similarities between these two architectures. Section 2.2 discusses the problems in channel management due to the on-the-fly switching pattern in OPS/OBS networks. In Section 2.3, fundamental contention resolution schemes are introduced. On the basis of channel reservation and contention resolution, comprehensive scheduling of optical packets/bursts is one of the key issues in the switch design. This is treated in Section 2.4. Section 2.5 reviews and classifies the QoS provisioning mechanisms and architectures proposed for OPS/OBS networks in literature.

2.1 Network Architectures

As the two most representative network architectures based on the photonic packet switching, the OPS and OBS share a lot of similarities. In literature, a uniform definition that clearly differentiates these two architectures is not available yet. In this thesis, the definition in [Gau06] is adopted. The OPS stands for the architectures applying in-band signaling. The OBS, on the contrary, uses out-of-band signaling and hence enables more feasible implementations.

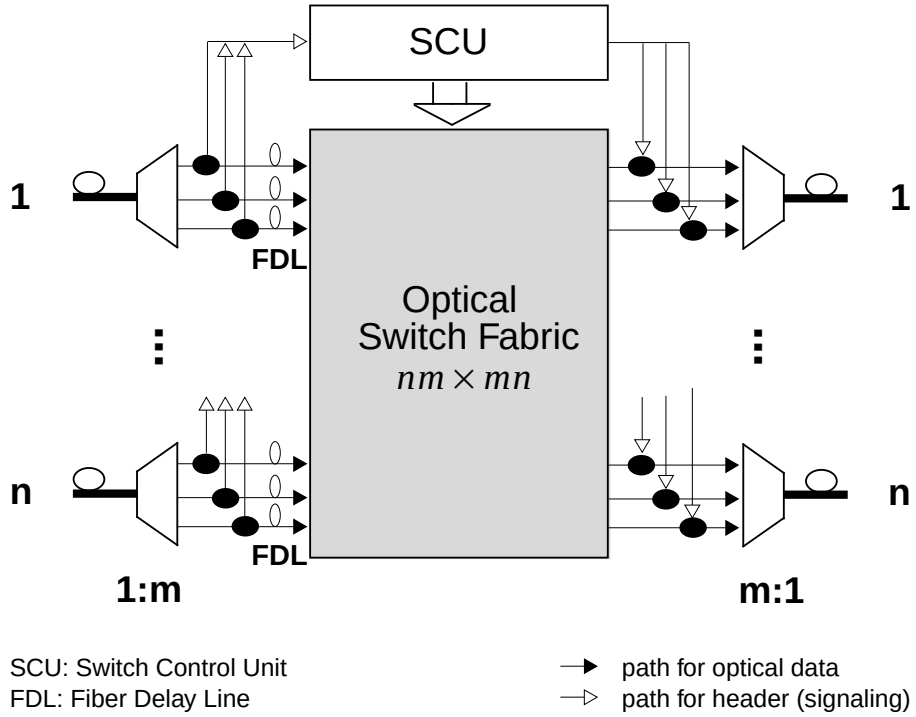


Figure 2.1: OPS node

2.1.1 Optical Packet Switching

In this section, the basic system model for an OPS switch is presented first to illustrate the switching concept of the OPS. Then, specific functionalities and their realization technologies are briefly introduced.

2.1.1.1 Overview

Optical packet is the elementary frame structure transmitted in OPS networks. Each packet is composed of a packet header and a payload part. The packet header contains the control information, like source and destination address or label, packet length, sequence number, time to live, packet type etc.. The payload part carries the data. The notion of in-band signaling means that the packet header and payload are always transmitted together on the same wavelength channel in OPS networks.

As illustrated in Fig. 2.1, an OPS switch can be divided into four parts: an array of input ports, a switch control unit (SCU), an optical switch fabric and an array of output ports. At the input ports, the wavelength-division-multiplexed (WDM) optical signals are split into multiple wavelengths by an optical demultiplexer. By tapping on the wavelength channel, the packet header is read from the optical channel and processed either electronically or optically by the SCU. After decoding the header information, the SCU carries out the table lookup and routing/switching decision, and configures the optical switch fabric to forward the optical packet to the destined output port. To compensate the latency in the header processing, the optical packet is delayed

by a fiber delay line (FDL) before entering the switch fabric. The data switching remains in the optical domain so that the optical transparency is achieved on the data path. At the output port, a new packet header is reinserted in front of the payload part. This is necessary because in general the header information should be updated (e.g., the field of time to live, label swap) at each switching hop [BPS94].

2.1.1.2 Multiplexing of Header and Payload

While multiplexing and demultiplexing of wavelengths are mostly realized by the arrayed waveguide grating (AWG) [Kau02, RS02], there are different ways to combine the packet header and payload on individual wavelengths. In the time domain, this is implemented by letting the packet header lead the payload. Since the switching elements in OPS networks treat the header and payload as separate segments, a guard time must be inserted in between to assure the signal integrity. Alternatively, solutions in the frequency domain take advantage of the subcarrier multiplexing (SCM) technology to send packet headers on a subcarrier of the wavelength [BPS94, EBS02] concurrently with the payload transmission on the wavelength channel.

2.1.1.3 Realization of SCU and Switch Fabric

As indicated in Section 2.1.1.1, the SCU can be realized either in the optical domain or in the electronic domain. An optical SCU promises a very fast processing speed, e.g., 10 G packet per second (pps) [KWS00], and is supposed to be the final solution to the full optical transparency in an optical packet switch. A representative method is to utilize the optical code division multiplexing (OCDMA) technology to realize the address/label matching and switch control [KWS00, RS02, KM03]. However, since the key technologies for optical buffering and signal processing are still in the infancy, the optical SCU is impractical in the foreseeable future. Alternatively, an electronic SCU serves as a more feasible solution that has led to inspiring achievements in the development of testbeds and demonstrations [GRG+98a, GRG+98b, HA00, DDC+03]. This approach requires the optical-to-electrical-to-optical (OEO) conversion of packet headers and the processing speed is around the magnitude of 10^6 pps. Nevertheless, by using large packet sizes, a high switching throughput can be achieved. Traffic assembly is hence necessary in ingress edge nodes to aggregate the client traffic of small data units into large optical packets [OSHT01]. As a by-product, the traffic assembly re-shapes the ingress traffic and can improve the loss performance in core switches [YXM+02].

Depending on the realization of the SCU, the switch fabric can be optically or electronically controlled. The structures and technologies applied in the optical switch fabric are broad and specialized topics. An overview on these can be obtained in [BPS94, RS02].

2.1.1.4 Header Update

The update of the packet header information is generally done by removing the old header and inserting a new header [Blu01, BBWP03]. In case the header is multiplexed by the SCM, the old header is dropped by filtering out the accommodating subcarrier with the application

of frequency selective components, e.g., semiconductor optical amplifier (SOA) exploiting the cross gain modulation (XGM). If the header is temporally multiplexed with the payload on the wavelength, the old header can be blocked out by the timely control of the SOA gate in the switch fabric. Furthermore, if the non-return-to-zero (NRZ) channel code is used for the header, nonlinear fiber cross phase modulation (XPM) wavelength converter can be exploited to suppress the header.

Alternatively, packet headers can be updated without the erasing/inserting procedure. In [FKW+01], an architecture was proposed to generate a new header by the optical exclusive-or (XOR) operation on the old header and a mask sequence.

2.1.1.5 Synchronous and Asynchronous Operation Mode

A further degree of freedom in the switch design is the synchronous/asynchronous operation mode. In a synchronous node, the switch fabric is configured in constant time intervals so that the traffic scheduling and resource allocation can be efficiently carried out on the basis of constant time slot. Packet size in synchronous OPS can be fixed or variable [GRG+98a, OSHT01]. Anyway, the size is specified such that the packet transmission duration including the necessary guard time fits into an integer multiple of the time slot. If necessary, padding should be appended to packets. Optical packets arriving from various input ports are synchronously forwarded through the switch fabric.

In a public transport network, the length of a fiber link is generally not aligned to an integer multiple of the packet transmission duration and the propagation delay is variable due to temperature variation and chromatic dispersion [GRG+98b, EBS02]. Additionally, jitter occurs within the switching node when packets are forwarded to output ports through different internal paths. For these reasons, incoming packets from different links arrive asynchronously at the switching node. Packet synchronization is necessary to align the timing of packet delivery according to the local reference time.

A coarse synchronization is realized by a multi-stage switchable FDL array. Each stage of the array is composed of parallel FDLs of different lengths. By switching a packet through a specific path in the array that concatenates a series of FDLs of various lengths, an appropriate delay can be introduced to synchronize the packet to the desired time alignment. Furthermore, a fine tuning of the delay is possible by exploiting the wavelength dependent propagation delay in a highly dispersive fiber. Different delays are obtainable by converting the packet onto different wavelengths. In [KWS00, RS02], these two approaches are jointly applied in a switching node. Of course, the resolution of the introduced delay by both means is always limited and should be controlled such that the resulting misalignment is able to be compensated by the inter-packet guard time.

Optical packet synchronization is still an expensive technology. Besides, the active switching elements of synchronization devices bring undesirable signal degradations. These problems are avoided in the asynchronous OPS. Furthermore, the asynchronous operation mode allows the variable sizes of optical packets so that padding is not necessary to transport the client data units of variable sizes. However, the flexibility of asynchronous switching leads to additional complexities in packet scheduling and resource allocation in the SCU.

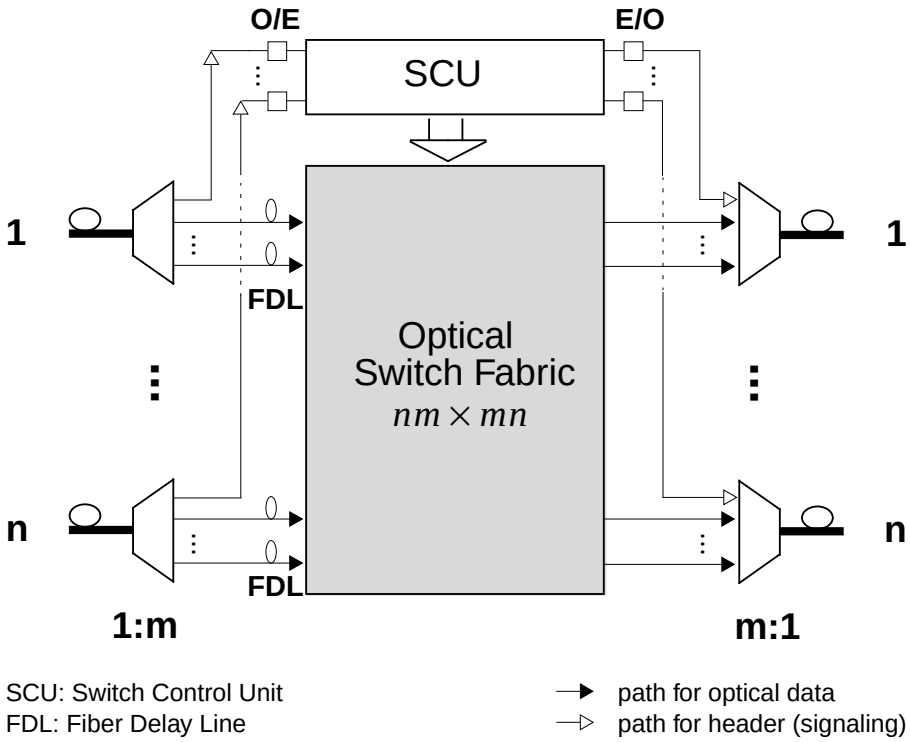


Figure 2.2: OBS node

2.1.2 Optical Burst Switching

OBS was originally proposed [QY99, Tur99] as an intermediate solution to full optical packet switching. Out-of-band signaling, electronic SCU and asynchronous operation mode are key features that alleviate the implementation hardness of the system. Correspondingly, traffic assembly becomes essential in OBS networks. An additional means is possible for the compensation of the header processing time in the SCU. The flexibility in the basic OBS model has given rise to a series of variations of the OBS network architecture.

2.1.2.1 Overview

In Fig. 2.2, an OBS node is illustrated. In contrast to the OPS node, on each link headers of data frames are transmitted through control wavelength channels separated from the data channels, i.e., out-of-band signaling. The control channels are terminated in each switch by O/E devices. Control information contained in headers is then decoded in the electronic SCU. Especially, due to the out-of-band signaling, the data structure of the header includes additionally a wavelength identifier to signal the SCU on which data channel the corresponding data frame is to arrive. Note that the wavelength ID results from the channel allocation for this data frame at the preceding hop. Routing/switching decision and resource allocation/reservation are performed by the SCU according to the decoded control information from the headers. An optical switch fabric is deployed to support transparent optical switching on the data path. OBS only considers asynchronous switching so that the complex optical synchronization procedure is not required.

At the output ports, updated headers are prepared by the SCU and sent to the next hop after E/O conversion. Since headers are received and sent on separate control wavelengths, the realization of header update in the switch is much simplified in comparison to that in OPS networks.

2.1.2.2 Traffic Assembly

To alleviate the requirement on the processing speed of the SCU, traffic assembly is mandatory in OBS edge nodes. For this reason, the optical data frame in OBS networks is called *burst* to highlight the fact that each frame contains a cluster of packets/frames from client networks. For each burst, a burst header packet (BHP) is generated in the edge node and transmitted parallel to the data burst on a control channel to signal the SCUs on the path through the core network. By choosing an appropriate burst size, the BHP rate is limited, so is the processing overhead in SCUs.

2.1.2.3 Compensation Delay

In order to compensate the latency of BHP processing in the SCU, a BHP must reach the SCU of a switch in advance before the corresponding burst is delivered to the switch fabric. Here, the interval between the BHP arrival and the burst arrival is referred to as *compensation delay*. Since the BHP latency is composed of a variable queueing delay and a relatively constant processing time, the compensation delay should be large enough to cover the range of latency as much as possible. Otherwise, a burst is blocked if it arrives at the switch fabric before its BHP has been processed and the switch fabric has been configured to forward it. The compensation delay can be generated either by deferring the burst transmission by an offset time [QY99, QY00] in the ingress edge node or by inserting FDLs in the data channels upon the inputs (cf. Fig. 2.2) of the switch fabric [Tur99, XVC00].

2.1.2.3.1 Offset-Time Approach

As illustrated in Fig. 2.3, after the BHP is sent, the burst transmission is postponed by an offset time in the ingress edge node. The value of the offset time is predefined and carried by the BHP. Hence, the SCU at the next hop can derive the expected arrival time of the optical burst so as to schedule the configuration of the switch fabric correspondingly. This is sometimes referred to as *advance reservation* in literature. At the i-th hop, the BHP experiences a transit delay of Δ_i in the electronic SCU, while the optical burst is relayed transparently. As a result, the offset time between the BHP and the burst is decreased by Δ_i after traversing the switching node. So, the offset time field in the BHP must be updated at each hop. The dashed line from the last hop to the egress node represents an optional signaling in the implementation. Because the BHP should temporally lead the burst throughout the core network, the initial offset time set by the edge node should at least cover the sum of the BHP transit times at all intermediate hops between the ingress edge node and the egress edge node.

Source routing strategy was proposed to enable the determination of the number of hops beforehand. Alternatively, the offset time can be fixed according to a predefined maximal hops of

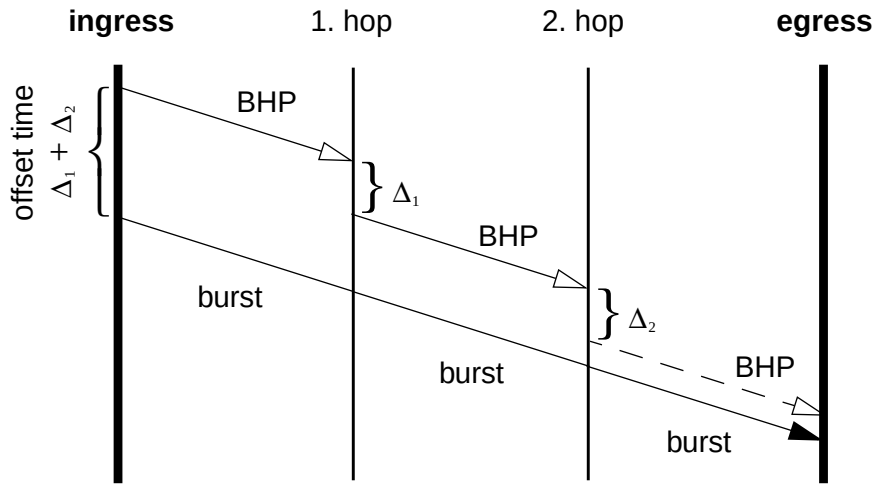


Figure 2.3: Offset time in an OBS network

routing through the network. This allows for more flexibility in the per-hop routing/switching strategy. However, it leads to unnecessary delay of bursts in case there are a smaller number of hops on the actual data path.

2.1.2.3.2 Per-Hop Compensation Delay by FDL

This approach is similar to that applied in OPS networks (cf. Fig. 2.1). The length of the FDL is decided according to the maximal single hop BHP latency. In comparison to the offset time based approach, this scheme neither imposes restrictions on the routing decision nor brings unnecessary delay by overestimating the hop distance. However, there are extra costs for the FDLs.

2.1.2.4 Variations in Signaling Architecture

On the basis of the original OBS architecture, there are various extensions and revisions proposed in literature. In the following, two variations in the signaling architecture are introduced.

2.1.2.4.1 Pipeline E2E Reservation

Ignoring the queueing delay, the latency of BHP processing in an SCU is mainly composed of two parts: the time for routing/switching decision and the time for channel allocation/reservation on the output link. As will be shown later, the channel allocation is a time-consuming procedure if a high channel utilization is required. The rationale of the pipeline reservation is to parallel the channel allocation procedures hop by hop in order to reduce the offset time [PCM+05, BS06].

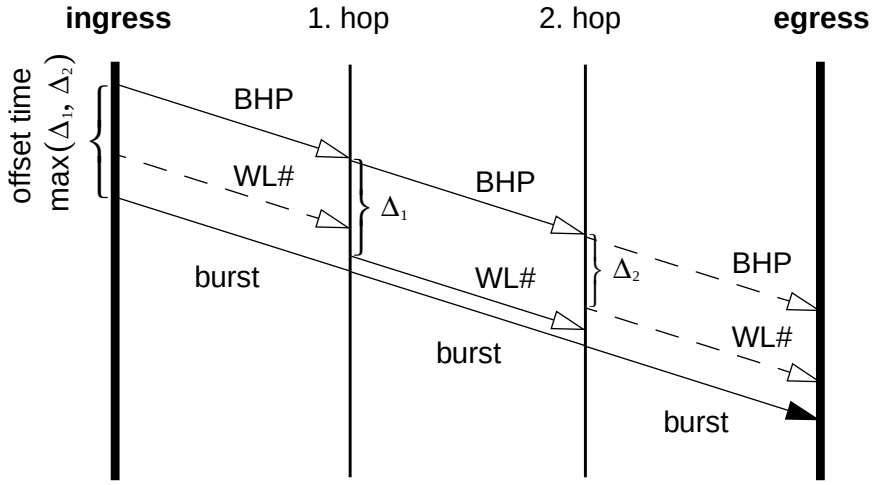


Figure 2.4: Signaling for pipeline reservation

The signaling process of this scheme is illustrated in Fig. 2.4. Here, each BHP is forwarded to the next hop without waiting for the completion of the channel allocation for the burst. Without loss of generality, it is assumed that the source routing strategy is applied so that routing/switching decision takes very little time. Thus, the per-hop BHP latency is negligible. After a processing latency of Δ_i, the output channel is allocated. Then, the current switching node sends another message to inform the next hop about the wavelength ID on which the burst will be transmitted. Note that the incoming wavelength ID is essential to configure the switch fabric, but it is not necessary for the channel allocation on the outgoing link in a strict-sense non-blocking switch [RS02, Küh06a]. In this way, the channel allocation procedures are pipelined in the switching nodes along the path. The shortest offset time that should be configured in the ingress node amounts to the maximum of Δ_i. It is much less than that in the original architecture (cf. Fig. 2.3).

However, additional signaling overhead is required to support the pipeline reservation. If the channel reservation for a burst fails at an intermediate hop, channel reservations in all downstream nodes lead to a waste of bandwidth. To solve this problem, it was proposed in [PCM+05] that a fast admission test is performed upon receiving a BHP to predict whether the channel request can be accepted or not. A BHP is relayed to the next hop only if the channel request passes the admission test.

2.1.2.4.2 Two-Way Reservation

In the original OBS proposal, the data burst is sent after the BHP from the ingress edge node irrespective whether the BHP succeeds in reserving the resources in core switching nodes or not. This is called *one-way reservation* and fits the connectionless packet-switching paradigm. *Two-way reservation* schemes were proposed later for a better support of assured services [DB02, ZWZ+04, HXL+05]. With this kind of schemes, reservation acknowledgement is fed back to the ingress edge node and the burst is sent only when the resource on the edge-to-edge path

is successfully reserved. In essence, this concept is very close to connection oriented circuit switching and is not treated in this dissertation.

2.1.3 Object Networks and Notations

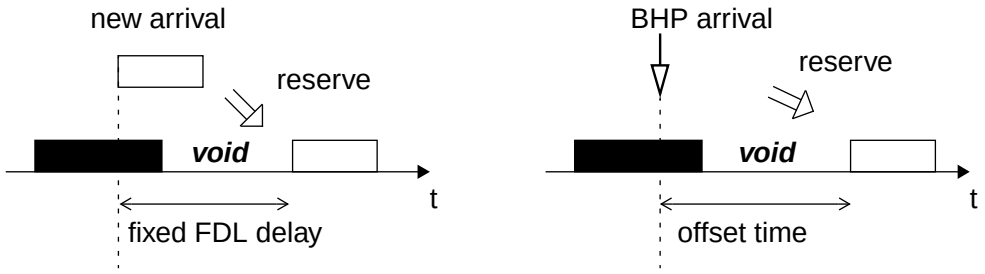
Since the realization of optical SCUs is still a far goal considering state-of-the-art technologies, the attention in this thesis will be confined to the opto-electronic solution of the OPS as well as the OBS. In the rest of the thesis, OPS networks will solely refer to those with opaque SCUs in switching nodes. Because the electronic SCU suffers from the physical limitation in the processing speed, traffic assembly in the ingress edge node is essential in reducing the header processing overhead in the core network. In the OPS/OBS network architectures inspected in this thesis, it is presumed that the client traffic is assembled in the ingress edge node to generate optical packets or optical bursts. For brevity, the optical packet in OPS networks and the optical burst in OBS networks will be uniformly referred to as the *data frame* or *optical frame*, unless explicitly stated otherwise. The data unit of client traffic is uniformly called *packet*, no matter whether it is IP packet, Ethernet frame or other specific data formats of client networks.

2.2 Channel Management

In an OPS/OBS node, channel collision arises when more than one data frame are to be forwarded to the same wavelength of an output port in the overlapping time periods. To assure the successful switching, the SCU tries to reserve the output channel for each data frame upon the header processing. In general, the window size to be reserved is set according to the transmission duration of the data frame. In the synchronous OPS node, this corresponds to the reservation of a fixed or variable number of contiguous time slots on the destined wavelength channel. In the asynchronous OPS/OBS node, a variable time duration has to be reserved on the channel. For OBS networks, this is specially known as the *just-enough-time* (JET) scheme [YQ97].

An important issue in the channel reservation is the handling of the channel voids. A void occurs when an optical frame is delayed through an FDL to resolve the collision, as illustrated in Fig. 2.5(a) for the asynchronous switching. The similar case holds for the synchronous OPS with variable packet sizes, only with the difference that the resulting void is in the unit of time slot because the FDL delay in the synchronous node is generally set to a multiple of time slots. In OBS networks, further voids can result from the channel reservation by the BHP when variable offset time is introduced between the BHP and optical burst, as shown in Fig. 2.5(b). In the case of Fig. 2.5(a), the void between the existing reservation (black box) and the new reservation (white box) can be still usable for other data frame, the transmission of which happens to fall into the void duration. Similarly, in the case of Fig. 2.5(b) if another BHP comes later but the offset time of the correspondent burst is short enough so that the burst arrival and duration fit the void, the void is also potentially usable.

With simple reservation schemes, e.g., *Horizon* [Tur99] and *latest available unscheduled channel* (LAUC) [XVC99], the channel allocation is uniquely characterized by the horizon defined as the latest time instant up to which the wavelength is reserved. The reservation module of the



(a) FDL delay for contention resolution (b) advance reservation by offset time in OBS

Figure 2.5: Occurrence of channel voids

SCU memorizes the horizon of each channel and only the period beyond the horizon is available for new reservations. This approach simplifies the implementation, but the channel voids cannot be exploited. Significant performance gain is achieved by the advanced schemes supporting void filling [TYC+00, XVC00]. For this purpose, it is necessary to keep track of the start and end of each reservation window, or alternatively the start and end of each void. Additional complexities in the data structure and search algorithm are to be dealt with [JG03b, Jun04, Jun05].

2.3 Contention Resolution

Without the relay of wavelength converters, optical switching can be performed only on the same wavelength between the input and output link, which is known as the wavelength continuity constraint. Popular optical switch architectures (e.g., broadcast-and-select [GRG+98a] or tune-and-select [FPS02]) exclude the internal blocking between the input and output channels of the same wavelength. Therefore, for an optical frame arriving on the wavelength λ, the contention arises only when the wavelength channel λ on the destined output link is already occupied.

Channel contention in OPS/OBS networks can be resolved in the wavelength domain, time domain and space domain [Gau03, AAB+07], respectively. To achieve an acceptable performance in a practical network, different schemes are generally combined to resolve the contention in multiple domains. If it fails, the optical frame is blocked and lost.

2.3.1 Wavelength Domain

By using the tunable wavelength converter (TWC) the optical frame can be switched onto another wavelength of the output link that is free during the requested time window. This is equivalent to the sharing of multiple channels in a trunk/bundle [AK93, Küh06c] and leads to a significant performance improvement even in case of detrimental traffic pattern like the self-similar traffic [TYC+00, DGSB01, GDSB01].

Ideally, a strict non-blocking switch can be realized by the full wavelength conversion scheme [FPS02, RT02], wherein each incoming (outgoing) wavelength channel is equipped with a TWC

that supports the conversion to (from) any other wavelengths used by the fiber link, i.e., full-range wavelength converter (FRWC). However, this scheme becomes very expensive when the number of wavelengths grows. More practical solutions applying limited number of TWCs shared per link or per node were proposed in [EL00, ELP03, MZA05]. Reduction of the expense can be also achieved by the use of the limited-range wavelength converter (LRWC) [YLES96, ELS05, PP06, RZVZ06].

2.3.2 Time Domain

Buffering is a traditional means to resolve the channel competition in the data transmission. However, quite different from the electronic buffer, an FDL provides only a fixed delay corresponding to the FDL length. Larger delay can be realized by a cascade of multiple elementary FDLs or by feeding the data to traverse an FDL for several times in a feedback buffer architecture. This leads to a coarse delay granularity that equals to the elementary FDL latency. Voids can be generated in the contention resolution by FDL buffering as shown in Section 2.2.

Channel voids are in principle detrimental for the network performance, because the void cannot always be exploited (e.g., the incoming frame size is larger than the void) even if the channel reservation supports void filling. A large FDL delay granularity on the one hand increases the void size so that the potential of the performance degradation due to the channel segmentation is high. On the other hand, the large FDL length enables large delay of data so that it is more likely that the channel is free for the data frame after being delayed. The FDL buffer should be elaborately dimensioned with respect to the delay granularity and the buffering depth, taking the influence of the frame size statistic and the load situation into consideration. The fundamental performance features of FDL buffer have been thoroughly studied in [LB03, FLB05] for synchronous buffering and in [Cal00, APW05, RLFB05] for asynchronous buffering, with respect to dedicated output buffer per link.

Since FDL buffers occupy additional ports of the switch fabric, the number of FDL buffers should be constrained in order not to expand the switch dimension too much. Different buffer architectures were proposed that deploy FDL buffer pools shared by all links [HCA98, CZC+04, ZLJ05]. Depending on whether the optical data can be fed back to recirculate a buffer stage or not, the buffer architectures are classified into feed-back and feed-forward architectures [CHA+01, Gau02].

2.3.3 Space Domain

Contention in the switch can be solved in the space domain by forwarding the blocked data frame to another output link that is free for the transmission. This is known as the deflection routing [CTT99, CCF01, WMA02, BBPV03]. Since the deviation route is mostly longer (in the number of intermediate hops) than the original route, the deflection routing increases the total network load. In a low load situation, deflection routing can significantly reduce the blocking probability, typically for network topologies with high degrees of connectivity like meshed networks. In high load situations, its effectiveness diminishes. With static deflection policies (e.g., shortest path in choosing the deviation route) the overall network performance can even

degrade at high network loads, because the deflection routing results in a positive feedback in driving the network to more severe congestion [BFB93].

To avoid the aforementioned disadvantage, at high system loads the deflection routing can be disabled in the local node by monitoring the traffic intensity [CWXQ03]. By disseminating the local performance statistics through the routing protocol, adaptive deflection policies [LSKS05, OT05] estimate the contention likelihood in the downstream nodes and select an alternate route with the least blocking probability. These schemes take full advantage of the imbalanced load situation in the network to resolve the local contention.

2.4 Scheduling for Efficient Resource Allocation

Scheduling in OPS/OBS node is concerned with the allocation of outgoing channels for optical data frames. It is a comprehensive problem involving both channel management and contention resolution introduced in previous sections. In this section, scheduling schemes for efficient resource allocation are introduced. Another important application area of the scheduling, i.e., service differentiation and QoS guarantee, will be discussed in Section 2.5.

According to the degree of freedom in a scheduling problem, it can be classified into any of the following categories.

Channel Scheduling: channel selection for a specific data frame. In case of contention, wavelength conversion, FDL buffering and deflection routing shall be taken into consideration in the channel selection.

Header Scheduling: determination of the service order for multiple frame headers in the SCU. The focus of header scheduling is to sort the channel requests according to some criterion. How the channel is decided for an individual request is not an issue here. For that, a channel scheduling scheme can be used.

Comprehensive Scheduling: scheduling multiple channel requests to multiple channels. Due to the additional degree of freedom, comprehensive scheduling is more complicated than channel scheduling and header scheduling.

The respective applications of scheduling schemes in asynchronous OPS/OBS nodes and synchronous OPS nodes are outlined in Table 2.1. Following this structure, various scheduling schemes are surveyed in the following subsections.

2.4.1 Scheduling in Asynchronous Nodes

2.4.1.1 Channel Scheduling

In asynchronous OPS/OBS nodes, frame arrivals follow a point process. So, the channel request can be sequentially processed i.e., in the *first-in, first-out* (FIFO) order. Furthermore, the combination of resolution schemes in different domains facilitates multiple solutions to the channel

Table 2.1: Classification of scheduling schemes for OPS/OBS networks

	Asynchronous OPS/OBS	**Synchronous OPS**
Channel Scheduling	optimization schemes to minimize the delay and void (*LAUC/-VF* and variations)	used for fixed frame length, selection of FDLs in case of contention resolution
	heuristic schemes (*First-Fit, Round-Robin, etc.*)	
Header Scheduling	BHP sorted in the SCU to emulate FCFS for OBS bursts	QoS differentiation §2.5.2.2.1
	QoS differentiation §2.5.2.2.1	
Comprehensive Scheduling	JISP problem and variations with intentional delay of data frames	used for variable frame length, optimization problem as a variation of JISP

allocation for an individual request. Channel scheduling is the typical form of scheduling in an asynchronous OPS/OBS node.

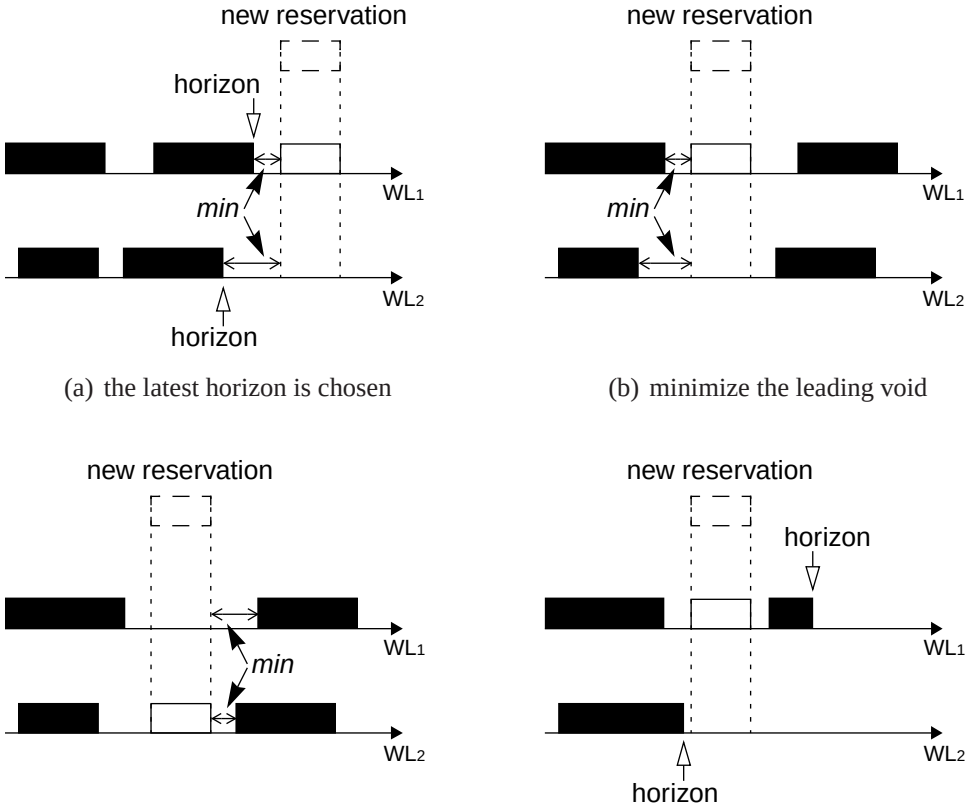
The resource allocation for each request is performed purely according to the arrival time and the data length as well as the current resource occupancy, irrespective of other unscheduled data frames. If an allocation is successful, probably resorting to the contention resolution, all the necessary resources including the output channel, wavelength converter and FDL are immediately reserved and the data is assured for switching. Otherwise, the frame is dropped without further consideration.

Depending on whether the channel selection algorithm carries out traversal search for some optimal criterion, channel scheduling schemes are further classified into optimization schemes and heuristic schemes.

2.4.1.1.1 Optimization Schemes

Wavelength conversion and FDL buffering are jointly applied in many node architectures for contention resolution. In this case, the channel scheduling is a two-dimensional problem that has to determine which wavelength channel and which time segment on this channel should be reserved.

With non-void-filling scheduling like Horizon [Tur99] and LAUC [XVC99], the new reservation can be only scheduled after the horizon time on each channel. The algorithm selects the wavelength that can accommodate the new reservation after the least delay of the data frame through the FDL. When multiple choices are available, the channel has the latest horizon, i.e., the latest unscheduled channel is reserved so that the newly generated void is minimal, as illustrated in Fig. 2.6(a).



(a) the latest horizon is chosen

(b) minimize the leading void

(c) minimize the subsequent void

(d) void is preferred

Figure 2.6: Arbitration rules in LAUC/-VF and the variations

A representative void-filling scheduling algorithm [TYC+00, XVC00] is the *latest-available-unused-channel with void filling* (LAUC-VF) scheme. In this scheme, the channel voids are also candidates for the channel allocation. The algorithm first tries to place the new reservation on a wavelength while minimizing the delay of FDL buffering. In case more than one void are found on different wavelengths, they are further arbitrated by minimizing the newly generated void leading the new reservation (cf. Fig. 2.6(b)). Alternatively, the new void following the new reservation (cf. Fig. 2.6(c)) can be minimized, as a variation [ISNS02] of LAUC-VF. In further elaborations [ISNS02], a void is more preferred for the reservation than an unscheduled channel under the condition that the necessary FDL delay is the same. An example for this is shown in Fig. 2.6(d). To tackle the high computational complexity, fast algorithms based on computational geometry were proposed in [XQLX04] to emulate the LAUC-VF scheduling.

2.4.1.1.2 Heuristic Schemes

In contrast to those optimization schemes, heuristic algorithms like first-fit and round-robin [XVC00, CC01] aim at a fast determination of the channel allocation without traversal search for the optimum. First-fit with void-filling algorithm, for example, simply adopts the first located void or unscheduled channel that is suitable for the reservation. Here, there is the degree of freedom in the design of the hunting mode [Gau06], i.e., the order in which the resource is searched for the allocation. An FDL-preferred algorithm focuses on one wavelength and

exhaustively try the available FDLs of different lengths to schedule the data frame onto this wavelength, before another wavelength is scanned. On the contrary, a converter-preferred algorithm tries to schedule the frame on any of the wavelength channels with a fixed FDL delay. If it fails, then a larger FDL delay is taken and the wavelengths are checked again in a round-robin order. The converter-preferred algorithm results in smaller delay but higher demand on the number of TWCs in comparison to the FDL-preferred algorithm. The design decision should be made according to the number of shared TWCs and FDLs in the switching node [Gau04].

Deflection routing is mostly used as a supplementary approach to the wavelength conversion and FDL buffering in the contention resolution. With full wavelength conversion, the performance evaluation showed that wavelength conversion should be tried first [Gau04]. Only if it fails, FDL buffering and deflection routing are used. A hunting mode preferring the deflection routing can bring more performance gain than the one preferring FDL buffering at light network loads. However, the situation is inverted at the medium and high loads [GKS04, AAB$^+$07].

2.4.1.2 Header Scheduling

In OBS networks applying variable offset time in the burst transmission, the BHP arrival sequence in a switch can differ from the burst arrival sequence. Request processing according to the BHP arrival sequence can cause a low channel utilization in the channel allocation. This problem is mitigated by intentionally buffering and resorting the BHPs in the SCU according to the burst arrival time [LQXX04, PCM$^+$05]. In this way, the channel reservation is performed in a *first-come, first-served* (FCFS) manner with respect to the actual burst arrival time.

2.4.1.3 Comprehensive Scheduling

In the channel scheduling, the resources are allocated only with respect to each individual request. So, the result can be suboptimal for a cluster of requests arriving within a specific time window. Header scheduling does not change this either. Comprehensive scheduling aims to achieve the optimal allocation scheme by considering a series of channel requests. In [KA05, CEBSC06], the handling of the requests in an SCU is divided into two phases: requests collection and scheduling. In the collection phase, the incoming requests are buffered and decoded. In the scheduling phase, scheduling algorithms are carried out for an optimal allocation scheme taking into consideration of all the collected requests. The optimization is known as joint interval selection problem (JISP) [COR01], the solution of which is, however, generally too complex for on-line applications. Correspondingly, heuristic algorithms [KA05, CEBSC06] were developed to reduce the computational complexity while retaining the advantage of the joint scheduling of a cluster of requests.

2.4.2 Scheduling in Synchronous Nodes

In the synchronous OPS node, there can be multiple frame arrivals in each time slot from the input ports to the same output link. In principle, comprehensive scheduling problem is concerned here.

2.4.2.1 Fixed Frame Size

In case the frame duration is fixed and equal to the time slot, the system does not suffer from the channel segmentation due to the voids. With full wavelength conversion, the channel allocation is trivial. For a channel request, an arbitrary free wavelength in the channel groups can be selected. Vice versa, for a free channel it does not matter which data frame in the request list is to be assigned. Due to this fact, the problem degrades to the channel scheduling. The issue of scheduling lies mainly in the determination of FDL delays to be used in case all wavelengths are allocated and a data frame has to be delayed for contention resolution. Here, different schemes are possible depending on the FDL architectures [CCC+04]. While simple algorithms just choose an available FDL providing the minimal delay, complexer algorithms also take into account the states of the FDLs, i.e., the buffered data in the FDLs. An FDL is chosen to minimize further collisions in the channel access after the data frame is delayed.

2.4.2.2 Variable Frame Size

If variable frame sizes are used, wavelength/FDL allocation for all instantaneously arriving frames becomes a complex optimization problem similar to the case in asynchronous switching nodes. The problem is difficult to be solved with on-line algorithms. For this reason, heuristic algorithms were proposed to achieve a fast computation with a compromised network performance [CCC+04].

2.5 QoS Provisioning

Theories and models for QoS provisioning have been intensively developed in last decades [FBTZ02]. A general overview on this area is provided by classifications of QoS models according to different criteria. Then, schemes for QoS differentiation in a single OPS/OBS node are introduced. On this basis, the architectures for absolute E2E QoS provisioning are discussed.

2.5.1 General Issues

QoS problems in communication networks can be classified in different perspectives. According to the setting of QoS goals, it is distinguished between relative and absolute QoS, deterministic and statistical QoS, respectively. With respect to different degrees of fineness in the flow differentiation, *integrated service* (IntServ) and *differentiated service* (DiffServ) architectures were defined. In view of the system dimension, a problem can be formulated as single-node or E2E QoS provisioning.

2.5.1.1 Relative QoS and Absolute QoS

The concept of relative QoS focuses on the realization of different levels of service qualities between traffic flows. Here, the QoS requirements of a single flow are defined with reference

to those of other flows. For example, as the QoS goal a ratio of the probability for delay bound violation is to be retained [Bod04] between multiple flows. Such kind of QoS paradigms is known as proportional QoS differentiation. In contrast to the relative QoS, absolute QoS cares for the absolute QoS specification for the flow of interests, irrespective of the service quality of other flows.

Scheduling disciplines play a central role in both relative and absolute QoS provisioning. A conserving law would be a very desirable feature for scheduling schemes, which means the average performance with respect to all flows remains at the same level as if the scheduling mechanism is not deployed. Some scheduling algorithms support only relative QoS. However, supplemented with admission control and policing mechanisms for all flows, absolute QoS can also be achieved. Some QoS mechanisms, e.g., channel partition, advanced scheduling schemes like weighted fair queueing (WFQ), support service isolation. That is, they assure that the QoS specification of a guaranteed flow is not violated no matter how the other flows behave. This can be regarded as an extreme case of service differentiation that minimizes the performance interaction between flows. As long as the traffic characteristic of the flow conforms to the agreement, its absolute QoS specification is reliably fulfilled.

2.5.1.2 Deterministic QoS and Statistical QoS

In light of the strictness of QoS guarantee, it is distinguished between deterministic guarantee and statistical guarantee. While deterministic QoS is specified in the form of hard bounds on the performance (e.g., delay bound, free of loss), in statistical guarantee QoS requirements are defined by means of performance statistics (e.g., mean delay, loss probability). To realize deterministic performance bounds, the traffic flow is generally regulated at the ingress of the network to conform to a constraint function. This is a function of time interval and defines the maximal arriving traffic amount in the time interval. The worst-case performance can be estimated by deterministic queueing analysis with respect to the constraint function [Zha95, EM97, Bou98, BT01]. In the statistical QoS paradigm, traffic is characterized by statistical process [1] (e.g., point process, fluid flow model) and conventional queueing analysis based on the Markov theory [Kle75, RMV96] is applicable. Statistical guarantee mostly promises much better resource utilization than deterministic guarantee. However, it relies on an appropriate modeling of traffic characteristics.

2.5.1.3 IntServ and DiffServ

In the Internet, IntServ and DiffServ are the two most important QoS architectures defined by Internet Engineering Task Force (IETF). In IntServ [BCS94], individual E2E flows are identified and parameterized by QoS mechanisms in every switching nodes. This offers a fine granularity in the flow differentiation. However, it is not scalable for large networks. In DiffServ [BBC+98], traffic flows are classified into several predefined service classes and the QoS mechanisms differentiate the QoS between the service classes instead of individual flows. Diverse

[1] Statistical QoS can also be provided with respect to regulated traffic characterized by constraint function [Kan06]. But it is not treated here.

treatments among the service classes, so-called per-hop behavior (PHB) requirements, are specified. These requirements can be realized through various mechanisms, which are left open as a degree of freedom in the implementation. If the absolute QoS is required on the basis of DiffServ, it is necessary to map the QoS requirements of individual flows to those of the service classes defined per hop.

2.5.1.4 Single-Node QoS and E2E QoS

Considering the system boundary, the QoS problem can be solved either for individual network nodes or in an E2E view. Naturally, the E2E QoS is always realized on the basis of per hop QoS guarantee through the path. In the formulation of the problem, it is necessary to deduce QoS requirements at each hop from the overall E2E QoS specification, for example, the delay budget allocation through the network.

Furthermore, per hop QoS mechanism relies on the traffic specification at the input of each node. This sets forth the demand on characterization of the departure flow from each node. In the deterministic QoS paradigm, the theory of network calculus [Bou98, BT01] was developed to deal with this problem. Alternatively, per-hop shaping [SC00] was proposed to provide a renewal flow specification at the input port. In the statistical QoS paradigm, accurate characterization of departure traffic by stochastic process is generally difficult for individual flows. A common practice is to apply the Poisson process as a conservative presumption to model the traffic. Especially in the DiffServ architecture, this assumption is asymptotically true if the traffic of each service class is aggregated from a large number of E2E flows.

2.5.2 QoS Differentiation in Single OPS/OBS Nodes

In OPS/OBS core nodes, the buffering capacity through FDLs is very limited. So, the queueing delay is not a performance issue. On the other hand, data loss is inevitable due to the nature of switching “on-the-fly” of the optical frames. Frame loss probability becomes the major concern in QoS provisioning, which falls in the statistical QoS paradigm. With respect to the scalability, the concept of DiffServ is typically applied in the transport networks.

An overview on the QoS differentiation schemes in individual OPS/OBS nodes is given in Fig. 2.7. While the solutions on the data path perform the admission and channel allocation solely for the data frames, solutions on the signaling path achieve the performance differentiation by handling the frame headers.

2.5.2.1 Solutions on the Data Path

2.5.2.1.1 Channel Partitioning

Basic QoS differentiation can be realized through bandwidth allocation between service classes. The data path of an OPS/OBS node is in principle modeled by a loss system [YQ00, DGSB01, YCQ02a] like conventional circuit switch. Correspondingly, channel partitioning was proposed

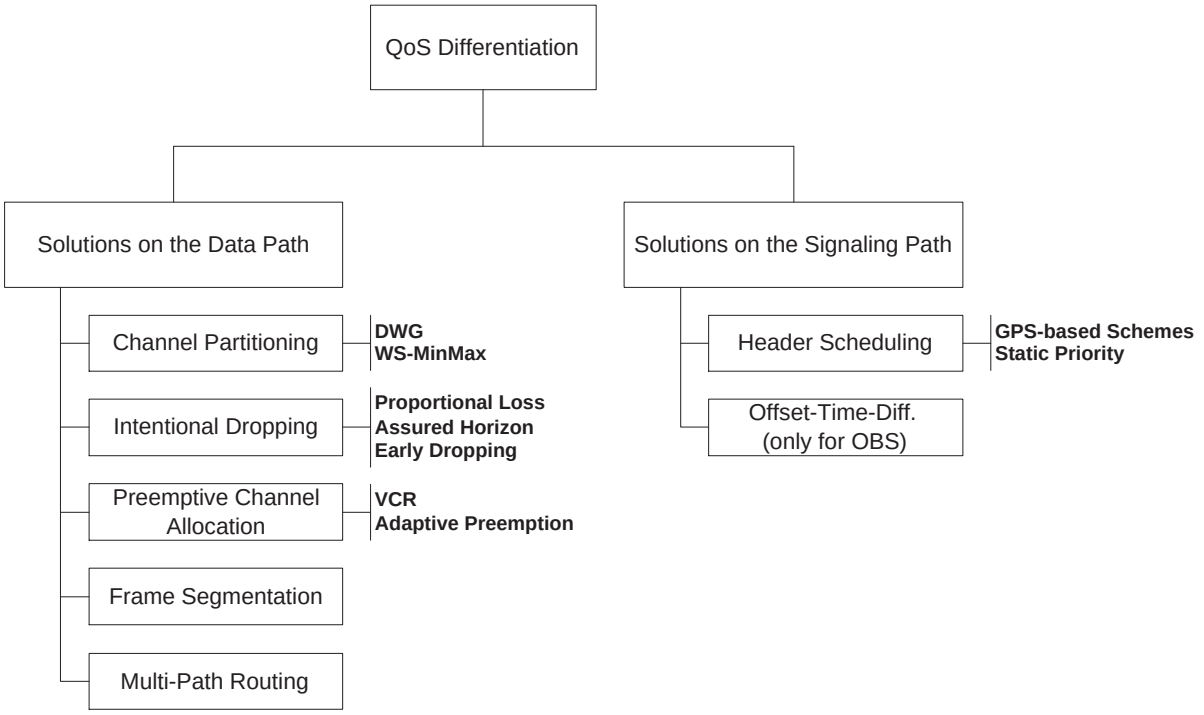


Figure 2.7: Classification of QoS differentiation schemes

to realize bandwidth allocation in such a system. It serves as a fundamental differentiation mechanism and is able to accommodate other schemes like intentional dropping, preemptive channel allocation and multi-path routing (cf. Fig. 2.7).

In a rigorous channel partitioning, wavelengths on the output link are grouped into disjoint sets. The wavelengths in each set serve a specific service class exclusively. This scheme provides a hard service isolation, however, is mostly over-killing and leads to a low channel utilization. In [ZVJC04], *dynamic wavelength grouping* (DWG) was proposed. In this scheme, the set of wavelengths accessible for a service class is not fixed. Instead, it only specifies for a service class i the maximal number of wavelengths $m_{\max,i}$ that can be held at one time. The sum of $m_{\max,i}$ for all i can be greater than the total number of wavelengths of the link m_{link} to allow a certain degree of bandwidth sharing. Especially, a high priority class is assigned with a relatively large $m_{\max,i}$. If an optical data frame arrives and finds its corresponding service class has occupied as many wavelengths as $m_{\max,i}$, there are two options to be taken.

1. A wavelength is reserved for the request if there are free channels. Otherwise, the data frame is dropped.

2. The data frame is dropped no matter whether there are free wavelengths on the output link or not.

Scheme 1) is a natural choice [UOS04] for synchronous OPS with constant frame size. Otherwise, the free bandwidth is simply wasted. For variable frame size or asynchronous operation, scheme 2) is preferable [ZVJC04] if the service differentiation is emphasized. In other words,

over-reservation is not allowed because this can block eligible data frames of other classes that arrive later. However, this excludes the chance to exploit the free bandwidth more efficiently.

To tackle the above dilemma, adaptive schemes were proposed. In [ZVJC04], the algorithm keeps track of the on-line measurement of loss rate for each class and judges accordingly for a concerned frame arrival whether scheme 1) or scheme 2) should be applied. If the respective loss rate is far below the specification, scheme 1) is chosen to allow a better resource utilization. Otherwise, scheme 2) is taken to assure the service differentiation. In [UOS04], the authors proposed to adaptively adjust the channel grouping parameters based on the measurement of loss performance per class.

DWG was further generalized to a scheme called *wavelength sharing with minimum provisioning and maximum occupancy* (WS-MinMax) [YR06]. WS-MinMax retains the parameter $m_{\max,i}$ and uses Scheme 2) to discard the data frame when $m_{\max,i}$ is exceeded. In addition, it specifies for each class a guaranteed number of wavelengths $m_{\min,i} : m_{\min,i} < m_{\max,i}$ under the condition $\sum_i m_{\min,i} < m_{\text{link}}$. The idea is to assure $m_{\min,i}$ wavelengths for each class in any case and allow $m_{\text{link}} - \sum_i m_{\min,i}$ wavelengths to be shared by all classes in a FCFS manner. Like DWG, the algorithm here only counts the number of wavelengths and does not fix the wavelength groups. If a data frame of class i arrives and finds the number of wavelength occupied by class i amounts to $m_i : m_{\min,i} \leq m_i < m_{\max,i}$, it is eligible to reserve a free wavelength as long as the wavelength number $m_{\min,j}$ can still be guaranteed for any of the other classes $j \neq i$. Mathematically expressed, a channel reservation is admitted for a newly arriving frame of class i if and only if [YR06]:

$$m_i < \min\{m_{\max,i}, m_{\text{link}} - \sum_{j \neq i} \max\{m_j, m_{\min,j}\}\} \tag{2.1}$$

Eq. (2.1) shows that the admission procedure of a single data frame involves the parameters of all service classes. So, an appropriate parameter setting is crucial for the overall system performance. In [YR06], a heuristic optimization algorithm was proposed for the determination of $m_{\max,i}$ and $m_{\min,i}$ with the object to minimize the loss probability of the best effort service while holding the specification of loss performance for assured services.

2.5.2.1.2 Intentional Dropping

With intentional dropping, a data frame can be dropped even when there is no contention on the output link.

Proportional loss scheme [CHT01] aims to realize proportional loss performance between different service classes. For this purpose, the traffic loss of each service class is measured in the switch and a data frame can be intentionally discarded in order to retain the ratio of the loss probability between the classes.

In *assured horizon* [Dol04], the system is switched to the congestion working mode as soon as the number of occupied wavelengths on the output link reaches a predefined threshold. In the congestion mode, all data frames of best effort service (e.g., non-conforming traffic classified

by traffic regulator in the edge node) are actively dropped to avoid channel competition against assured services.

Early dropping [ZVJC04] uses the similar concept as random early detection (RED) mechanism. The arriving data frames of low priority classes can be intentionally dropped to increase the chance of successful channel reservation for high priority classes. The intentional dropping is done randomly with a probability assigned to each service class. The value of the dropping probability is dynamically adjusted by comparing the measured loss performance for the classes of higher priority with the predefined threshold levels.

2.5.2.1.3 Preemptive Channel Allocation

The preemptive reservation allows a newly arriving data frame to preempt another data frame being transmitted on the output channel or an existing reservation on that channel. There are various preemptive schemes.

In *virtual channel reservation* (VCR) [GTJL05], preemptive reservation is deployed on the basis of a variation of DWG. Here, the over-reservation (scheme 1 in Section 2.5.2.1.1) is allowed as long as free output channels are available. As a compromise, the channel reservations marked as over-reserved are subject to preemption by eligible frames coming later. Especially, a signaling protocol was designed for OBS networks. When a burst reservation on the output channel is preempted, a signaling message is sent to inform the downstream switching nodes to cancel the reservation of the preempted burst, so as to increase the channel utilization.

In *adaptive preemption* scheme [OS05], a preemptive drop policy was proposed as a self-contained differentiation mechanism. With a certain probability, an arriving frame of a high priority class can preempt an existing reservation or transmission of the low priority class. The preemption probability is dynamically tuned according to the measured loss performance of the high priority class in order to keep the loss under the agreed level.

2.5.2.1.4 Frame Segmentation

Commonly, when the transmission of a data frame is interrupted, the frame is regarded as lost due to the destruction of the information integrity. For the same reason, if the channel reservation of a frame is preempted before the transmission starts, the reservation is removed completely. However, by appropriate frame structure and coding scheme, it is possible in case of interrupted transmission that the segment already transmitted onto the channel is able to be decoded by the receiver. Correspondingly, when a channel reservation is preempted, only the segment overlapping with the contending frame needs to be preempted.

Taking advantage of this feature, *prioritized burst segmentation* [VJ03] permits a data frame to encapsulate traffic from different service classes, with low priority traffic assembled on the tail of the frame. In case of a preemption of a frame transmission or its channel reservation in the switching node, only the tail segment in collision with the contending frame is dropped. In this way, the lower priority traffic experiences a higher loss ratio in a competition situation.

2.5.2.1.5 Multi-Path Routing

Multi-path routing was proposed in [CCRS06] as a supplementary scheme for channel partitioning. In each switching node, the routing algorithm prepares multiple routing paths for flows of low priority class: a default path following the shortest path and alternative paths with larger delay. The channel occupancy of each outgoing link is dynamically monitored. If a congestion state is discovered on the default outgoing link, a low priority optical frame is switched to an alternative path. On the contrary, high priority traffic is always routed through the shortest path. As a result, high priority class obtains a better loss performance at the expense of the increased delay for low priority class.

2.5.2.2 Solutions on the Signaling Path

In the electronic SCU, scheduling schemes were suggested for the header processing to tune the loss performance of different service classes. In case of out-of-band signaling in OBS networks, different offset times in the advance reservation can influence the probability of successful channel reservation considerably.

2.5.2.2.1 Header Scheduling in SCU

In Section 2.5.2.1, the solutions on the data path implicitly assume that the signaling channels as well as the processing capacity of SCUs are sufficiently dimensioned to exclude any impact on the performance in the data plane. Header scheduling, on the contrary, looks at non-negligible queueing latency of frame headers in the SCU.

In [KA04], it is assumed that the queues of frame headers in the SCU grow up when the data channels are heavily loaded. Scheduling scheme (e.g., WFQ scheduling) based on the generalized process sharing (GPS) concept is applied in the SCU to sort frame headers of different classes before they are delivered for processing. The parameters of the scheduling algorithm are configured to reflect the quantitative bandwidth allocation in *data channels* for each class. As a result of the service isolation property of the scheduler, if the traffic intensity of a service class exceeds the agreement, the headers of non-conforming data traffic suffer larger queueing delay. In case of an overdue in the header processing, the optical data frame is dropped. In this way, the switching of conforming traffic is protected against the non-conforming traffic. This scheme, however, takes effect only in heavy load and overload situation.

In [YZV01], incoming frame headers in the SCU are buffered and scheduled by static priority policy. The QoS differentiation is based on the fact that a frame header processed earlier has a better chance in reserving the channel successfully. On the other hand, the prioritization effect turns up only when there are sufficient backlogs in queues. So, the QoS differentiation with this scheme occurs at relatively high load and in the range of high data loss rate.

2.5.2.2.2 Offset-Time-Differentiation for OBS

Special QoS scheme was proposed for OBS networks by means of the out-of- band signaling [YQ00, QYD01, BS05]. Here, additional QoS offset time is inserted between the transmissions of BHPs and optical bursts. Bursts of high priority class have a longer QoS offset time. At the channel reservation by BHP, a larger offset time means a reservation in the farther future. Alternatively, taking the burst arrival as the reference point in time, a larger offset time is equivalent to an earlier reservation. In either perspective, the probability of successful reservation is higher. In OBS switch without FDL buffering, a high priority burst obtains an absolute superiority over a low priority burst if its offset time is larger than that of the low priority burst by a duration equal to the maximal burst transmission time of the low priority burst. In that case, a high priority burst is never blocked due to the channel occupancy by low priority bursts.

This *offset-time-differentiation* scheme holds the conservation law [YQ00], which is an important advantage compared with the intentional dropping and preemptive scheduling. At the same time, its effectiveness does not depend on the BHP queueing in the SCU like the prioritized scheduling of frame headers mentioned above. On the other hand, QoS offset time is generally much larger than the offset time for the compensation of BHP processing latency in SCUs. This brings additional E2E delay and results in large channel voids. Since small burst sizes have better chances to fit into voids than large sizes, it leads to unfairness problem in the channel reservation [YZV01].

2.5.3 Absolute E2E QoS Guarantee

To support absolute E2E QoS guarantee, a complete QoS architecture is built on the basis of the per-hop QoS differentiation mechanism by integrating admission control, signaling and reservation protocols [Dol04] as well as control/management protocols. Note that the admission control and resource request are here discussed on the E2E flow level (i.e., the concept of virtual connections), instead of the channel requests of individual optical frames addressed in previous sections. The focus is placed on the conceptual problem formulation in the absolute E2E QoS guarantee. The paradigms of static traffic engineering and dynamic service provisioning are inspected respectively.

2.5.3.1 Approach in Static Traffic Engineering

For admission control, a conceptually straightforward way is to evaluate the loss performance following the routing path and derive the E2E loss rate for the request. The request is admitted under the following constraints.

1. The estimated E2E probability $P_{\mathrm{EE}}^{\mathrm{L}}$ for this request is smaller than or equal to the maximal allowable loss rate P_{EE}^{*}.

2. The E2E QoS requirements of those admitted flows in service are not violated.

Assuming the route from the ingress node to the egress node is fixed for the request and there are in total z hops. The hops on the route are sequentially numbered by $i : 1 \leq i \leq z$ and the loss probability at the i-th hop is denoted then by P_i^{L}. Assuming statistical independence, the E2E loss probability is calculated by:

$$P_{\mathrm{EE}}^{\mathrm{L}} = 1 - \prod_i (1 - P_i^{\mathrm{L}}) \tag{2.2}$$

As explained in Section 2.5.1, the Poisson process is commonly used to model the input traffic of the switch, which is also justified for OPS/OBS nodes [IA02, YCQ02a]. Per-hop loss probability P_i^{L} is analyzed taking into account the local traffic intensity, channel occupancy and the QoS differentiation scheme deployed. The closed-form solution, if any, is mostly nonlinear. Also note that the computation of P_i^{L} at an individual hop i is inter-correlated with the loss probability at the preceding hops because the input traffic intensity in the current node depends on the loss probability at preceding hops. Furthermore, the E2E loss performance must be also evaluated for those carried flows to assure Condition 2). In total, the problem concerns itself with the solution to a large array of nonlinear equations. Erlang fixed-point approximation [KMK04] is a well known numerical method to solve this kind of problems.

Due to the large computational overhead, this approach is not suitable for on-line admission control, but is applied in static network planning and traffic engineering [RVZW03]. On this basis, maximal offered traffic is specified for each ingress/egress pair and for each service class such that all QoS measures are satisfied at the maximal network load. In the network operation, the on-line admission algorithm just needs to control the traffic load accordingly. In this architecture, a relative low resource utilization is conceivable due to the static resource allocation.

2.5.3.2 Dynamic Provisioning through QoS Budget Partitioning

The complexity of the admission problem in the preceding section is attributed to two aspects. First, the modeling of reduced load after each hop brings dependence between the loss performance of individual hops. Second, taking solely the E2E loss probability as the performance measure allows a large solution space for the combination of per-hop loss probability along the route. To alleviate the computational overhead, additional approximations and constraints are introduced in this section. Considering the very low loss rate (e.g., $10^{-4} \sim 10^{-6}$) required in practical transport networks, the traffic intensity of a flow can be regarded as fixed throughout the network. Also, the maximal allowable E2E loss probability P_{EE}^{*} is partitioned into a series of specifications of per-node loss probability P_i^{*} with $1 \leq i \leq z$. This resolves the E2E QoS problem into the QoS guarantees in individual nodes along the routing path. By this means, on-line algorithms are feasible to admit E2E service requests with respect to the dynamic network status.

Furthermore, the derivation of per-hop loss probability from the E2E QoS requirement can follow either a static rule or a dynamic approach. Further details are given in the following subsections.

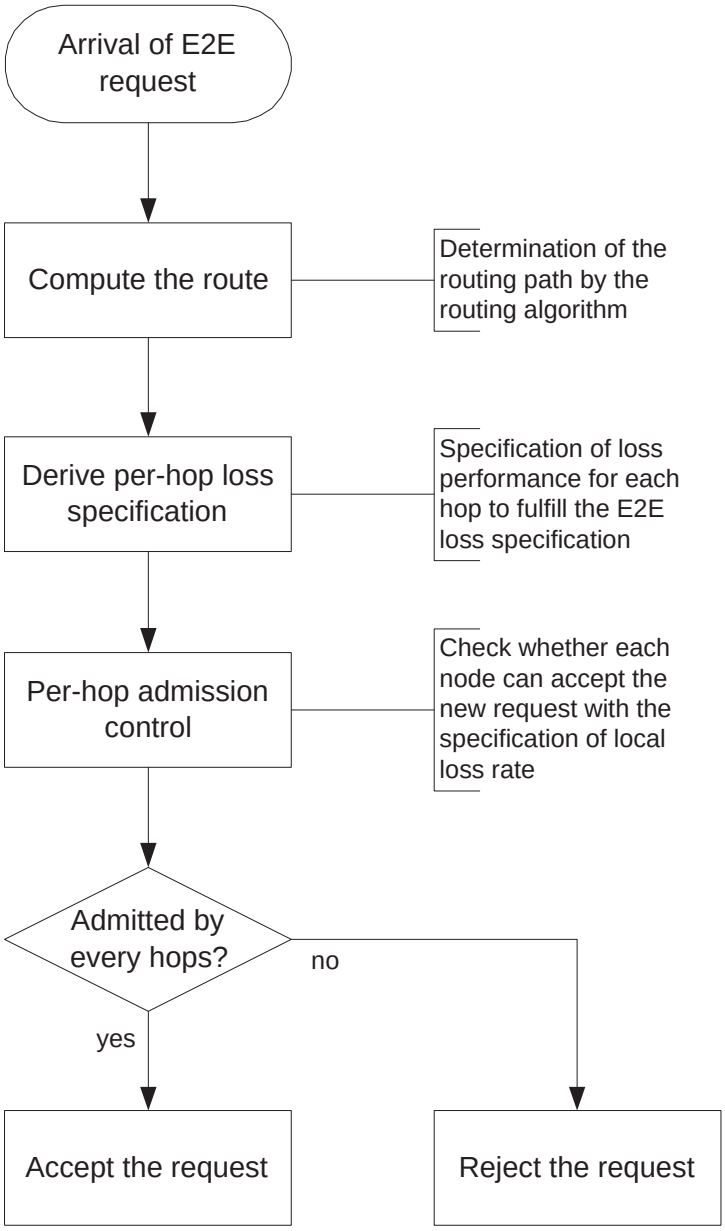


Figure 2.8: Admission control with static specification of per-hop loss rate

2.5.3.2.1 Static Specification

With the static QoS specification along the routing path, a generic admission procedure is illustrated in the flow chart in Fig. 2.8. Upon the arrival of a new E2E flow request in the ingress edge node, the route is first determined by the routing algorithm. Then, the E2E loss specification of the request is split into the requirements at individual hops of the route according to some static rule irrespective of the network status. On the basis of the per-hop QoS specification, each node first classifies the request into a local QoS class and executes the admission algorithm with respect to the traffic description of the new request. If it passes the admission control of every hops, then the new request is accepted. At the same time, the states and configurations in the switching nodes along the route are updated correspondingly. Otherwise, the request is rejected.

Equal splitting is a commonly used static rule in the per-hop QoS specification [ZVJC04, OS05, YR06]. Replacing $P_{\mathrm{EE}}^{\mathrm{L}}$ and P_i^{L} in Eq. (2.2) with P_{EE}^* and P_i^* respectively, and making P_i^* constant for all i, it is derived:

$$P_i^* = 1 - (1 - P_{\mathrm{EE}}^*)^{1/z} \tag{2.3}$$

The maximal E2E loss probability P_{EE}^* is determined by the service class. With a given routing scheme, the hop distance z between different node pairs can differ from each other, which leads to different specifications of P_i^* even if the corresponding service class is the same. Larger z results in smaller value of P_i^*. Ideally, for each value of P_i^*, a local QoS class should be defined that keeps P_i^* as the QoS goal. In the worst case, the number of classes in a switching node amounts to the product of the number of service classes and the number of node pairs. This leads to a scalability problem in large networks.

In [ZVJC04], a path clustering mechanism was proposed to reduce the number of local QoS classes having to be maintained. The values of z for all ingress/egress node pairs are sorted and grouped with the neighbouring values. For each group of hop distance, P_i^* is calculated by adopting the largest value of z in the group. This leads to a relatively conservative specification of P_i^* to guarantee the absolute QoS. Depending on the degree of the path clustering, the number of necessary QoS classes per hop can be much reduced.

2.5.3.2.2 Dynamic Specification

Static specification does not consider the unbalance in the network load. On the contrary, dynamic scheme distinguishes the workloads of individual switching nodes, and tries to assign relatively loose P_i^* for congested nodes and tight P_i^* for nodes with low load. A representative scheme was proposed in [PCM$^+$04, PCM$^+$07] which is based on a probing mechanism to collect the information of network status.

In this scheme, a set of classes associated with corresponding QoS requirements on the loss performance are defined uniformly for every switching nodes. These classes are not directly related to the service class of a request identified in the ingress edge node. They only represent different levels of loss performance that can be provided by a switching node. Upon the arrival of a new request, the admission algorithm in the ingress node carries out two rounds of signaling through the network, as illustrated in Fig. 2.9. The first round aims to probe the load situation in the network and reserve the resources on the routing path. The request is sent through the route. Each hop on the route checks the best QoS class it can support for this request, and marks the reservation of local resources correspondingly. The class allocation together with the current load in the local node is sent back in a report message to the ingress node. After collecting all feedbacks, with Eq. (2.2) the ingress node is able to decide whether the E2E QoS requirement of the request can be satisfied. Furthermore, if the obtainable E2E loss performance is even better than the requirement, there are opportunities to reallocate a class with a looser loss specification in those switching nodes showing relatively high loads. The instructions of reallocation are sent in a configuration message along the path in the second signaling round. On the other hand,

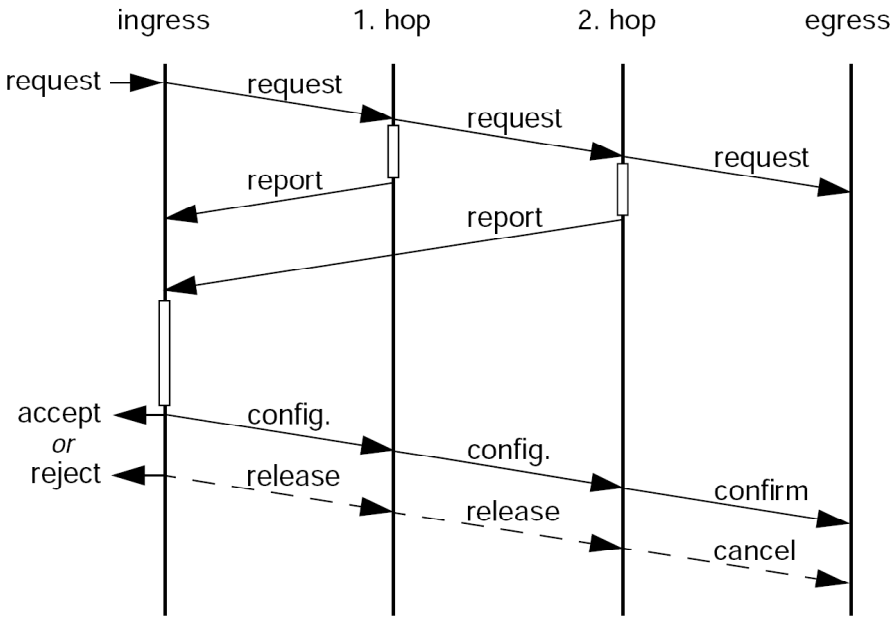


Figure 2.9: Signaling process in the dynamic specification with probing

if the ingress node decides to reject the request, a release message is sent instead to cancel the resource reservation in each node made by the request message in the first round.

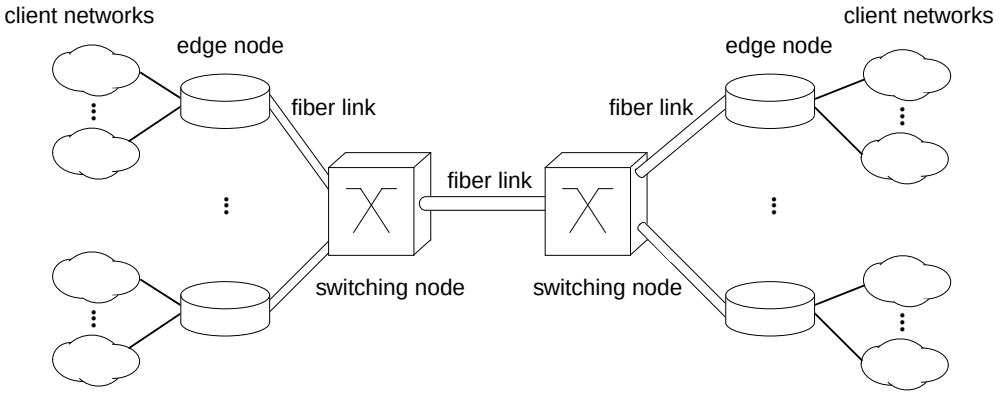
3 Edge Node and its Relevance to E2E QoS

As indicated in Chapter 2, an edge node plays the role of a gateway between the OPS/OBS core network and client networks. As one of the most important tasks, an edge node assembles incoming client traffic into optical data frames and disassembles received data frames to client traffic. With respect to various system and performance constraints, different assembly schemes were proposed and the influence on traffic characteristics as well as network performance was evaluated. Scheduling of frame transmission in the ingress edge node is another important design issue, which can differ from the scheduling in a core switching node due to the availability of electronic buffers in edge node. Furthermore, integrated in the E2E QoS architecture, an ingress edge node is responsible for the admission control, wherein the performance impact of the local edge node (e.g., with respect to delay and jitter) shall also be taken into consideration.

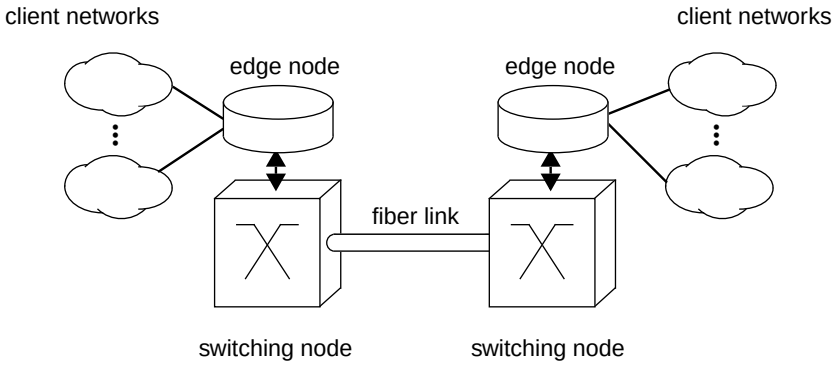
This chapter reviews the related work for the issues mentioned above. On this basis, the QoS problem investigated in this thesis is outlined. In Section 3.1, system models for the edge nodes are introduced with respect to different network layout schemes. Traffic assembly and transmission scheduling are discussed in Section 3.2 and 3.3, respectively. Section 3.4 summarizes the QoS requirements that should be evaluated in the edge node and formulates the admission control problem.

3.1 Network Layout and System Model

According to the deployment environment, there are basically two layout styles at the edge of OPS/OBS networks, as illustrated in Fig. 3.1 with an example network consisting of two core switching nodes. Fig. 3.1(a) shows the approach with distributed edge nodes. Here, edge nodes are geographically interspersed to collect and distribute traffic of client networks in the local area. Traffic collected from client networks is assembled into optical data frames by edge nodes and relayed to a switching node through a point-to-point fiber link. From the perspective of a switching node, however, there is no difference between feeder links from edge nodes and links from other switching nodes. Fig. 3.1(b) depicts the other layout scheme having an edge node co-located with the switching node. Here, the edge node and switching node can be separate devices interconnected with each other by a local fiber cable in a similar way as the case of distributed edge nodes. Alternatively, they can be realized as an integrated system. While the integrated solution provides the opportunity for more efficient scheduling and resource alloca-



(a) distributed edge nodes



(b) co-located edge node with the switching node

Figure 3.1: Network layout on the boundary of OPS/OBS networks

tion, an isolated realization of the edge node and the switching node reduces the implementation complexity and assures a good modularization.

A generic model for an edge node and its connection to the switching node is depicted in Fig. 3.2. The edge node is composed of a transmission module and a receiving module. In the transmission module, the incoming client traffic is classified into different forwarding equivalence classes (FEC) according to their service class and destined egress node. The notation of FEC indicates the fact that packets of the same FEC can be treated equivalently in the core network for routing and switching. After classification, packets are distributed to assembly buffers according to the assigned FECs, where they are aggregated into data frames of larger sizes. Further on, the transmission scheduler delivers the data frames and schedules their transmissions to the switching node through multiple wavelength channels. In the receiving module, the main tasks lie in the disassembly of data frames and distribution of packets to client networks.

The thick dashed data forwarding lines between the edge node modules and the switching node denote the unidirectional fiber links in the case of separate node realization. When an integrated edge/switching node is concerned, the fiber connections as well as the channel multiplexer/demultiplexer are not mandatory. Instead, the Tx/Rx ports of the edge node can be directly connected with the ports of the switch fabric in the switching node. Furthermore, in the integrated solution the assembly control and transmission scheduling in the edge node can be

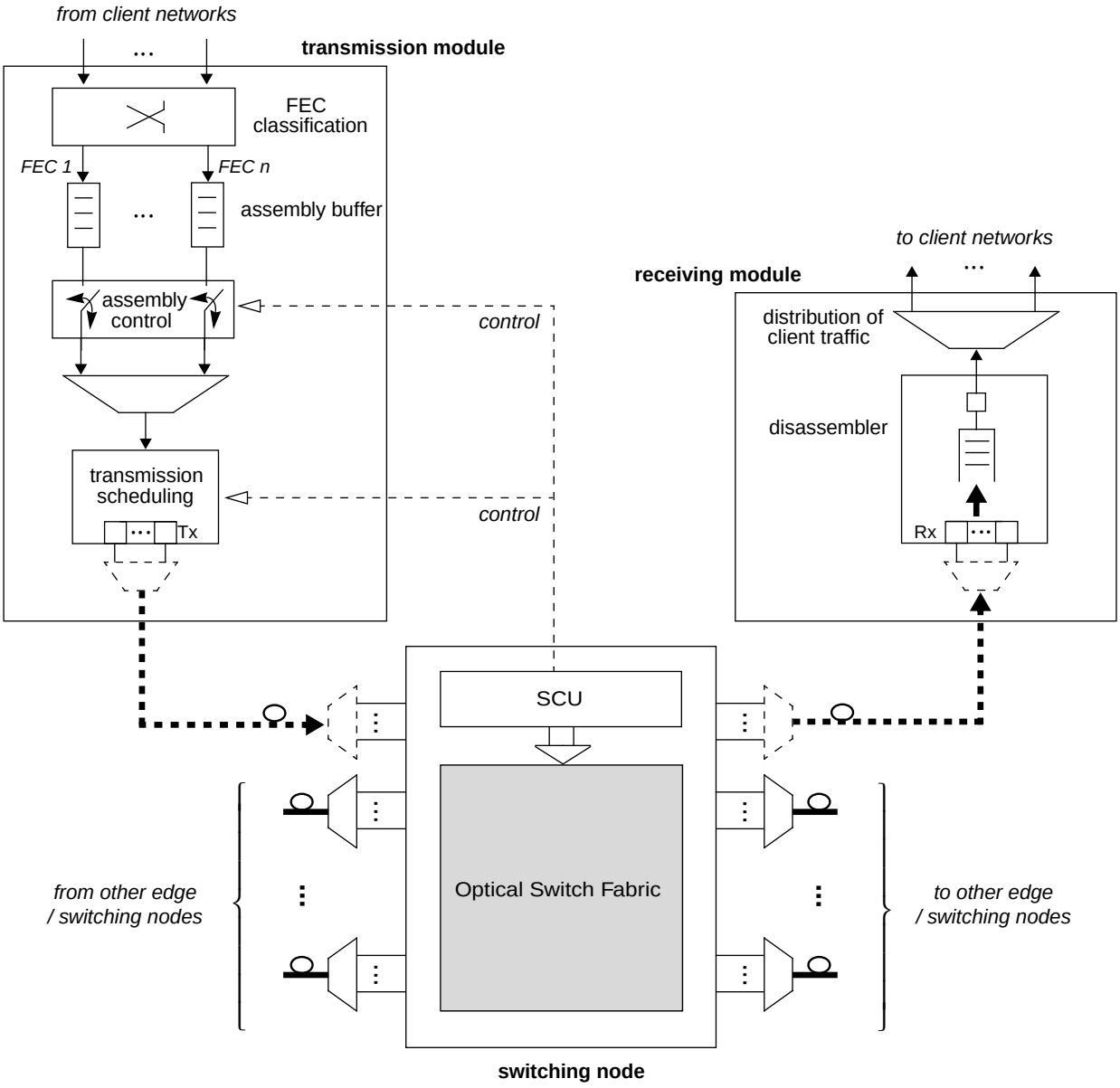


Figure 3.2: System model for the coupling of edge node and switching node

performed with reference to the system status of the switching node, which is denoted by the dashed control line starting from the SCU of the switching node.

The remainder of the thesis concentrates on the transmission module of the edge node since it plays a much more important role in the E2E QoS provisioning than the receiving module. Besides, the isolated realization of edge node and switching node is assumed for the implementation benefit mentioned above. Further details on the study of an integrated node solution can be found in [YXM+02].

3.2 Traffic Assembly

This section first introduces various assembly schemes proposed in literature. Then, the work in traffic characterization for the assembled traffic is reviewed and the impact on the queueing performance in subsequent network elements is discussed. Finally, the performance evaluation for a single assembly buffer is introduced.

3.2.1 Assembly Schemes

The degree of freedom in designing basic assembly schemes is decided by various constraint factors, which are mostly reflected in three basic system parameters of an assembler. Depending on whether the parameters are configured statically or dynamically, assembly schemes can be classified into static schemes and dynamic schemes.

Additionally, special assembly schemes were proposed for QoS provisioning and efficient grooming.

3.2.1.1 Basic Schemes

3.2.1.1.1 Assembly Parameters

Basically, there are three parameters in controlling the traffic assembly: timeout period, frame size threshold and minimal frame size.

Timeout Period: The timeout period specifies an upper bound on the assembly delay. A timer is set upon the arrival of the first packet in an empty assembly buffer. When the timeout occurs, a data frame is generated by aggregating the packets in the assembly buffer.

Frame Size Threshold: The duration of an optical frame in the core network cannot be arbitrarily large. For example, it is generally desired that the maximal frame duration does not exceed the FDL length in order to realize the data storage function in the optical domain more precisely [Buc05]. Furthermore, in the implementation of the channel reservation module for an SCU, it can also be necessary to specify an upper bound for the frame duration [JG03a]. To this end, a size threshold is used in the assembly control to limit the size of a data frame. When the size of the total packets in the assembly buffer exceeds the frame size threshold, the assembly of a new data frame is triggered.

Minimal Frame Size: For photonic switching, there are additional transmission overheads due to the guard gaps between the data segments and the minimal switching time of a switch fabric. In order to assure the switching efficiency, the duration of a data frame cannot be too small. For example, it should be beyond the magnitude of microsecond compared to the typical switching overhead in the magnitude of nanosecond [RS02, Buc05] with SOA switching technology. The parameter of minimal frame size is defined exactly for this purpose. Upon the assembly of a data frame, padding is added, if necessary, to make the frame size not smaller than the minimal frame size. Note here that the minimal frame size

does not trigger the generation of a frame, in contrast to the timeout period and frame size threshold. It only decides whether padding is necessary during the generation of a frame.

3.2.1.1.2 Static Schemes

Basic assembly algorithms are derived by taking different combinations of the assembly parameters into consideration. A brief summary is given in Table 3.1, which is self-explanatory.

For the assembly schemes concerned with the frame size threshold, there are further variations in the specification of the frame size. If fixed frame size equal to the frame size threshold is used, padding is necessary in case the timeout occurs and the collected data do not amount to the frame size. To mitigate the padding overhead in this situation, the frame can be filled with packets from other FECs, as long as they are heading to the same egress edge node [LKA+06]. Furthermore, if the client traffic has variable packet sizes, either padding or packet segmentation [RZ03] is needed to match the possible discrepancy between the size of total collected packets and the fixed frame size.

In case of variable frame size, it is required to define whether the frame size threshold is treated as a loose upper bound or a strict upper bound for the frame size. This decides if the latest arriving packet that triggers the frame size threshold can be included into the current data frame or it has to be delayed and encapsulated into the next data frame [She05]. Nevertheless, if the frame size threshold is relatively large, these variations have only limited influence on the network performance.

Table 3.1: Basic assembly schemes with static configuration

Schemes	**Parameters**	**Description**
pure time-based assembly [Lae02]	timeout period	assembly triggered by timeout
pure size-based assembly [Lae02]	frame size threshold	assembly triggered by frame size threshold
combined time/size-based assembly [XVC00]	timeout period, frame size threshold	assembly triggered either by timeout or by frame size threshold, depending on which occurs first
time-based, min-length assembly [GCT00, YCQ02b]	timeout period, minimal frame size	assembly triggered by timeout; padding is inserted to assure the minimal frame size
time/size-based, min-length assembly [YCQ02b]	timeout period, frame size threshold, minimal frame size	assembly triggered either by timeout or by frame size threshold; padding is inserted to assure minimal frame size

3.2.1.1.3 Adaptive Schemes

As the timeout period is directly related to the delay performance, it is intuitive to have a tunable timeout period in order to improve the delay performance. Also, the timeout period can be configured in a TCP-friendly way.

In [IJAM06], an assembly algorithm with fixed minimal size and adjustable timeout period was proposed. The goal is to dynamically minimize the assembly delay while still keeping the transmission efficiency (i.e., the average ratio of payload size and frame size) on an accepted level. For this purpose, the timeout period is adaptively configured according to the on-line estimation of the first/second-moment statistics as well as the correlation property of the incoming client traffic.

Cao et al. [CLCQ02] studied the pure time-based assembly with adaptive timeout period. When the timer is to be set upon the arrival of an initial packet in the empty assembly buffer, the resulting frame size is predicted based on the recorded history of frame sizes generated by the assembler. The timeout period is then determined from the predicted frame size by keeping a constant ratio of the frame size and the timeout period. This is equivalent to smoothing the traffic according to a constant traffic rate. It was shown that this scheme can effectively improve the good-put of TCP flows.

3.2.1.2 Special Schemes

Based on the basic assembly schemes, more advanced algorithms can incorporate further QoS mechanisms. In [Dol04], the pure time-based assembly was extended to perform the traffic policing for incoming client traffic. When the timeout is triggered, an eligible size is calculated according to the history of frame generation and the amount of reserved bandwidth for the respective FEC. Only the buffer content up to the eligible size is recognized as compliant traffic that can be packed into the current frame and sent immediately. Yuang et al. [YTSC04] aimed to provide delay differentiation and fairness between subclasses of a FEC on the basis of combined time/size-based assembly. Here, each subclass is assigned with a quota defining the eligible data amount of the subclass in an assembled frame. In the frame generation, the assembly algorithm schedules packets of different subclasses to compose individual frames with respect to the quota specification.

Other schemes make a further step in allowing data of different FECs to be assembled together. In [VZJC02, VJ03], the authors suggested to append low priority traffic on the tail of a high priority frame. In combination with the preemptive scheduling with frame segmentation in switching nodes (cf. Section 2.5.2.1.4), QoS differentiation in terms of blocking probability can be realized. Farahmand et al. [FZJ05] looked at assembly with the requirement of minimal frame size. The authors exploited the aggregation of traffic destined to different egress nodes in order to reduce the padding overhead (i.e., efficient grooming). In this case, a packet may experience multi-hop assembly/disassembly before it reaches the destination.

3.2.2 Impact on Traffic Characteristics

Traffic characterization of the assembled frame traffic has its importance in two aspects. First, in the decomposition network performance analysis [Küh79], the stochastic process of the departure traffic serves as the input parameter for the queueing analysis in the subsequent network elements, for example, the derivation of the frame loss probability in core switching nodes. Second, the traffic statistics can be used to directly synthesize assembled traffic in a simulation study instead of having to simulate the assembly process. This appreciably saves the run time in a network wide simulation scenario.

This section outlines the relevant work on the characterization of assembled traffic. The modeling with renewal point processes is introduced first. Then, the correlation property of the traffic is inspected. In both cases, the focus is placed on the three basic assembly schemes: pure time-based, pure size-based and combined time/size-based assembly. They all allow variable frame size. The parameter of minimal frame size is neglected because it generally has a small value and the influence on the network performance is limited.

3.2.2.1 Modeling by Renewal Point Processes

Frame inter-departure time and frame size are the two most important parameters to characterize the assembled traffic, which can be derived by standard probability analysis based on the assumption of the independent and identically distributed (i.i.d.) packet interarrival time as well as the packet size of the incoming client traffic. In [Lae02, YCQ02b, dVRG04, YLC+04], frame inter-departure time and frame size distribution were studied under the assumptions of discrete time (DT) Bernoulli packet arrivals and continuous time (CT) Poisson arrivals, respectively.

Analytical results are reviewed, summarized and extended in the following subsections. Then, the approximate modeling of the departure frame traffic is presented, which has a better applicability in practice due to its simplicity. Finally, the relevance of the traffic characteristic on the network performance is discussed.

3.2.2.1.1 Statistical Analysis

The point process of the assembly procedure is plotted in Fig.3.3 in the CT domain without loss of generality.

Pure Time-based Assembly

With the pure time-based assembly scheme, the frame inter-departure time D is the sum of the constant timeout period t_{th} and the residual packet interarrival time I_{res} of the first arriving packet in this frame, as illustrated in Fig.3.3(a). Due to the memoryless property of the geometric distribution (in DT model) and negative exponential distribution (in CT model) assumed for the packet interarrival time, the distribution of I_{res} is the same as that of the packet interarrival time

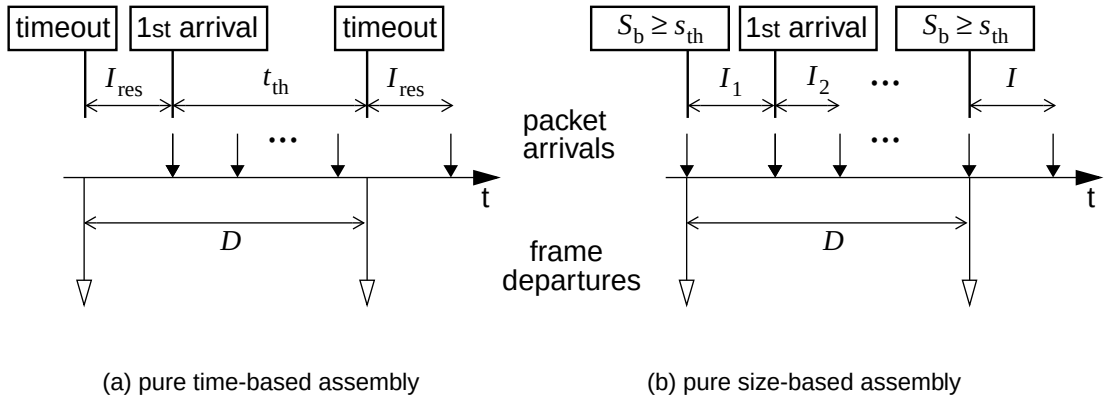


Figure 3.3: Point process for frame assembly

I. Therefore, statistically $D = t_{th} + I$. Since t_{th} is constant, the distribution of D turns out to be the distribution of I with a right shift by t_{th} [dVRG04].

The frame size S_b is the sum of the size of the packets that compose the data frame. Therefore, the frame size distribution is a compound distribution represented by [Lae02, dVRG04]:

$$S_b = \sum_{i=1}^{N_b} L_i. \tag{3.1}$$

Here, L_i denotes the packet length of the i-th packet in the burst, which is i.i.d.. N_b represents the number of packets in the frame. Since an assembly period begins always with one packet arrival, N_b is at least 1. Correspondingly, $N_b - 1$ is the number of packet arrivals within the assembly period t_{th}, which has a Binomial distribution [Lae02] or Poisson distribution [dVRG04] according to the presumption.

Pure Size-based Assembly

With the pure size-based assembly scheme, the frame inter-departure time is the sum of individual interarrival times of all packet of the data frame, as illustrated in Fig.3.3(b) [1] in which s_{th} denotes the frame size threshold. This yields the compound distribution [dVRG04]:

$$D = \sum_{i=1}^{N_b} I_i \tag{3.2}$$

[1] Note, that according to the scenario of Fig. 3.3(b) the assembled data frame size may slightly exceed the frame size threshold.

where I_i is i.i.d. random variable (RV) denoting the interarrival time of the i-th packet in the frame. The probability distribution of N_b can be derived by [Lae02, dVRG04]:

$$\begin{aligned} P\{N_\mathrm{b}=k\} &= P\{S_\mathrm{b}\geq s_\mathrm{th}\} \qquad (3.3)\\ &= P\{\sum_{i=1}^{k} L_i \geq s_\mathrm{th} | \sum_{i=1}^{k-1} L_i < s_\mathrm{th}\} \cdot P\{\sum_{i=1}^{k-1} L_i < s_\mathrm{th}\} \\ &= \sum_{j=1}^{s_\mathrm{th}-1} P\{L_k \geq s_\mathrm{th}-j\} \cdot P\{\sum_{i=1}^{k-1} L_i = j\} \qquad (3.4) \end{aligned}$$

The resulting frame size falls in the interval of $[s_\mathrm{th}-l_\mathrm{max}, s_\mathrm{th}+l_\mathrm{max}]$, where l_max denotes the maximal packet length. In most cases, s_th is much larger than l_max. So, it suffices in the performance evaluation to regard the frame size as constant $S_\mathrm{b}=s_\mathrm{th}$. More precise derivations of the frame size distribution can be found in [Lae02, dVRG04].

Combined Time/Size-based Assembly

This assembly scheme is a combination of the pure time-based and pure size-based assembly. Therefore, the resulting frame traffic characteristic can be derived based on the insight obtained previously.

Let T_a denote the assembly duration defined as the time interval between the arrival of the first packet at the assembly buffer and the time instant at which the frame is assembled. Then, $D=I+T_\mathrm{a}$. With the pure time-based assembly, $T_\mathrm{a}=t_\mathrm{th}$. With the pure size-based assembly,

$$T_\mathrm{a} = \sum_{i=2}^{N_\mathrm{b}} I_i \qquad (3.5)$$

where the distribution of N_b can be derived according to Eq. (3.4). In case of the combined assembly, it can be seen that T_a should follow a similar probability density function (PDF) as that under the pure size-based assembly, except that it is upper-bounded by the timeout period t_th as illustrated in Fig. 3.4(a). This leads to:

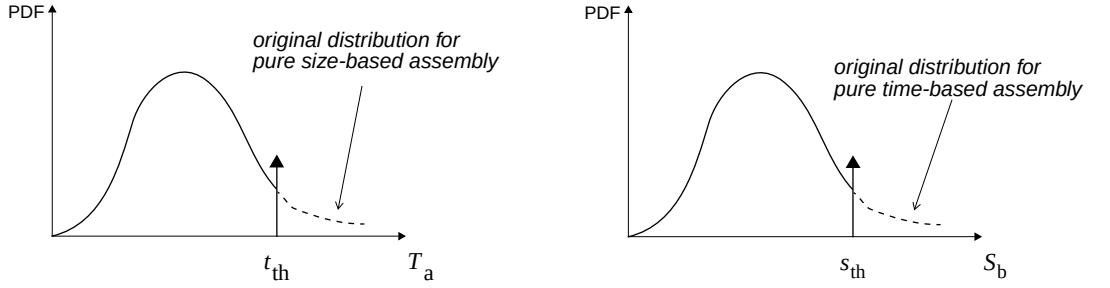
$$P\{T_\mathrm{a}\leq t\} = \begin{cases} P\{T_\mathrm{a}^\mathrm{psize}\leq t\} & \text{for } 0<t<t_\mathrm{th} \\ 1 & \text{for } t\geq t_\mathrm{th} \end{cases} \qquad (3.6)$$

Here, $T_\mathrm{a}^\mathrm{psize}$ denotes the assembly duration in the case of pure size-based assembly derived by Eq. (3.5).

Likewise, the frame size S_b distribution is similar to the case of the pure time-based assembly but upper-bounded by the maximal frame size $s_\mathrm{th}+l_\mathrm{max}$ imposed by the frame size threshold. If $s_\mathrm{th} >> l_\mathrm{max}$, s_th can be regarded as the maximal burst frame approximately. Then, the probability distribution of S_b can be approximated by the truncated distribution function [Lae02] as shown in Fig. 3.4(b):

$$P\{S_\mathrm{b}=s\} = \begin{cases} P\{S_\mathrm{b}^\mathrm{ptime}=s\} & \text{for } 0<s<s_\mathrm{th} \\ \sum_{i=s_\mathrm{th}}^{\infty} P\{S_\mathrm{b}^\mathrm{ptime}=s\} & \text{for } s\geq s_\mathrm{th} \end{cases} \qquad (3.7)$$

$S_\mathrm{b}^\mathrm{ptime}$ stands for the resulting frame size with the pure time-based assembly scheme which is obtainable from Eq. (3.1).



(a) assembly duration T_a

(b) frame size S_b (PDF approximated by continuous function for illustration)

Figure 3.4: PDF under the combined time/size-based assembly

3.2.2.1.2 Approximation for Closed-Form Solution

Although the characteristic of the frame departure process can be theoretically derived as shown in the preceding subsection, a closed-form solution is not always available. That is the typical case for the frame size under the pure time-based assembly and the frame inter-departure time under the pure size-based assembly, wherein a complex compound distribution is concerned. This limits the applicability of these results for a tractable performance analysis and motivates the approximation of the compound distribution by a more regular distribution.

Observing Eq. (3.1) and (3.2), it is noticeable that both compound distributions converge to Gaussian distribution $N(\mu,\sigma^2)$ according to the Central Limit Theory when N_b goes to very large [Lae02, YCQ02b, YLC+04]. The Gaussian distribution can be specified by by matching its μ and σ^2 with the mean and variance of the RV to be approximated. Alternatively, the Gamma distribution $\Gamma(\omega,\nu)$ is proposed [Lae02] as a more accurate approximation because the Gamma distribution assures a sample space of non-negative values in contrast to the Gaussian distribution. Again, the shape parameter ω and scale parameter ν are determined by matching the first and second moment statistics of the Gamma distribution with those of the RV [Ric95, Lae02]. In the following, the analytical results for the mean and variance of the frame size and frame inter-departure time under the respective assembly schemes are presented.

Denote the Poissonian packet arrival rate at the assembly buffer with λ, and use $E[\cdot]$ and $\mathrm{VAR}[\cdot]$ to represent the mean and variance of an RV. From Eq. (3.1), the mean and variance of the frame size under the pure time-based assembly in a CT system can be derived according to the property of the compound distribution [Küh06c]:

$$E[S_b] = E[N_b]\cdot E[L] = (1+\lambda\cdot t_{th})\cdot E[L] \tag{3.8}$$

$$\mathrm{VAR}[S_b] = E[N_b]\cdot \mathrm{VAR}[L] + \mathrm{VAR}[N_b]\cdot E[L]^2 = (1+\lambda\cdot t_{th})\cdot(\mathrm{VAR}[L]+E[L]^2). \tag{3.9}$$

Similarly, from Eq. (3.2) the first two moments of the frame inter-departure time D under the pure size-based assembly is derived:

$$E[D] = E[N_b]\cdot E[I] = \tfrac{1}{\lambda}E[N_b] \tag{3.10}$$

$$\mathrm{VAR}[D] = E[N_b]\cdot \mathrm{VAR}[I] + \mathrm{VAR}[N_b]\cdot E[I]^2 = \tfrac{1}{\lambda^2}(E[N_b]+\mathrm{VAR}[N_b]). \tag{3.11}$$

Here, $E[N_b] \approx s_{th}/E[L]$. VAR$[N_b]$ can be solved numerically from Eq. (3.4) by the transformation method [Küh06c].

As long as the distribution function of S_b under the pure time-based assembly and D under the pure size-based assembly are approximated by the closed-form expressions, the traffic characteristics under the combined time/size-based assembly can be easily derived from Eq. (3.6) and (3.7) by using the truncated distribution function.

3.2.2.1.3 Discussion: Shaping and Burstification Effect

In literature, traffic assembler is from time to time regarded as some kind of shaper that is able to smooth the input traffic and reduce the frame blocking probability in the core network. However, without a deeper understanding of the shaping effect here, the notation of shaping can be misleading. For example, it could be attractive to understand the shaping effect of an assembler in association with those conventional traffic shapers like token bucket and generic cell rate algorithm (GCRA) [Sta02, Küh06b]. Despite of the similarity in the mechanisms at a first glance, they are quite different procedures. While the power of traffic shaper comes from the specification of the traffic envelope, the shaping effect of assembler is just a by-product due to the relatively small variability in the frame inter-departure time. Actually, assembler is very analogous to the packetizer [KMK04] for stream traffic in packet networks, in the sense that they both pack small data units together to form large transmission units. For this reason, the assembly process is sometimes also called *burstification* [2] in the literature. In the following, the shaping effect and burstification effect of the assembler are looked at in more details.

Shaping Effect

In essence, the shaping effect is attributed to the suppressed variability in the frame inter-departure time after the assembly, which can be measured by the coefficient of variation (COV).

With the pure time-based assembly, the frame inter-departure time is the sum of the constant timeout period and one packet interarrival time. Hence, its COV must be less than that of the packet interarrival time, especially when the timeout period is relatively large.

In the case of pure size-based assembly, the COV of the frame inter-departure time c_D can be derived from Eq. (3.10) and (3.11):

$$c_D = \sqrt{\frac{c_I^2}{E[N_b]} + c_{N_b}^2} \tag{3.12}$$

where c_I denotes the COV of the packet interarrival time I and c_{N_b} is the COV of the number of packets N_b in the frame. Generally, the packet size of client traffic is upper-bounded (e.g., 1500 bytes for an Ethernet frame). When s_{th} is much larger than the packet size, c_{N_b} is small. Actually, $c_{N_b} \to 0$ with $s_{th} \to \infty$. On the other hand, $E[N_b] >> 1$. Therefore, it holds that $c_D << c_I$.

[2] The notation of burstification is mainly used in OBS networks. However, from the point of view of traffic analysis, it is a vivid synonym of general assembly procedure.

Since the distribution function of D under the combined assembly scheme is a truncation of the correspondent distribution function under the pure size-based assembly, the resultant COV of the frame inter-departure time is even smaller.

In summary, the assembly gives rise to a smaller variability in the frame inter-departure time. Modeled by the point process that neglects the data size at each event point, the departure process turns out to be smoother than the packet arrival process. When this departure process is directly forwarded to bufferless outgoing channels, the channel contention can be relieved [YCQ02a]. Notice that the statistics of frame size has relatively smaller influence on the loss performance at the same system load. This is known as the insensitivity property in M/G/n loss systems [Küh06c] and also in more generalized queueing systems [Bur81].

In a realistic network scenario, the channels are generally shared by multiple departure flows. So, the aggregated frame arrival process at the channels becomes variable again due to multiplexing. This counteracts the shaping effect to an extent depending on the degree of multiplexing. Performance analysis for WDM link fed with a limited number of assembled traffic flows is given in [HMQ+05], which shows that the frame interarrival time at the channel group can be regarded as constant approximately in the frame loss estimation. In case of large multiplexing degrees, the shaping effect fades out and the aggregated frame arrival process converges to a Poisson process, which justifies the application of the M/G/n loss model for the estimation of frame loss probability [YQ00, DG01, IA02, YLC+04].

Burstification Effect

While shaping effect is mentioned in terms of the frame inter-departure time, burstification effect has its focus on the size of data unit. It refers to the fact that the collected traffic amount in the assembly buffer is released as a whole data burst. From the perspective of a subsequent network element, this means a long duration of data arrivals at a peak rate, e.g., the transmission rate of the interconnection medium or an infinitely large rate in many theoretical studies. From queueing theory [AK93, RMV96], it is known that the burstification has an adverse impact on delay systems, i.e., it requires high buffer capacity and increases the queueing delay. Hence, the burstification effect is relevant for the performance of the edge node scheduler equipped with electronic transmission buffers.

3.2.2.2 Impact on Long Range Dependence

Traffic modeling with renewal point processes provides a good insight into the impact of assembly on traffic characteristics in the finest granularity of time scale. However, it neglects the correlation structure of network traffic. Today's data traffic shows non-negligible correlation spanning over a large time scale, which is known as long range dependence (LRD) [PF95, CB97, WTSW97, WPRT02]. LRD leads to self-similar structure of traffic variability on different time scales. Therefore, LRD traffic is frequently referred to as self-similar traffic as well. LRD can cause either high traffic loss in small-buffer systems or large queueing delay in large-buffer systems and thus needs special attention in the traffic analysis and performance evaluation.

Table 3.2: Impact of traffic assembly on the degree of LRD

	time-based assembly	size-based assembly	combined time/size
X_t in bytes	no change in H (proved)	no change in H (proved)	no change in H (proved)
X_t in frame counting	reduced H (by simulation)	no change in H (proved)	reduced H (by simulation)

The physical significance of the LRD can be well illuminated by the property of variance process. Let X_t denote the traffic volume in an arbitrary time interval of t, the variance of X_t of LRD traffic has the following property:

$$\mathrm{VAR}[X_t] \sim k\, t^{2H} \text{ for } t \to \infty. \tag{3.13}$$

Here, $H : 0.5 < H < 1$ is the Hurst parameter which measures the degree of LRD. k is constant for a specified traffic process. Eq. (3.13) states that the variance of LRD traffic grows polynomially fast on large time scales, in comparison to a linear increase (i.e., $H = 0.5$) in the case of renewal process or short range dependent (SRD) traffic.

Early studies [GCT00, XY02] showed that the Hurst parameter can be reduced after the traffic is assembled. The authors attributed this to the smoothing effect of the assembler. Later work discovered that the assembler only reduces the absolute traffic variability on small time scales. The Hurst parameter, which characterizes on large time scales the growing speed of VAR[X_t] instead of the absolute value of the variance, is not changed by the assembly procedure [YCQ02a, HDG03]. Especially, Hu et al. [HDG03] distinguished the frame departure process X_t measured in terms of bytes and in terms of frame counting, respectively. It was proved by analysis that the Hurst parameter is not changed by any of the three basic assembly schemes in case X_t is measured in bytes. A summary of the study is presented in Table 3.2.

It is worth pointing out that although the Hurst parameter is important for the characterization of LRD, it is not necessarily a decisive parameter in the performance evaluation. For example, in pure loss OPS/OBS nodes, it was shown that LRD has little influence on the frame loss probability [IA02, YCQ02a]. In delay systems, the relevance of LRD on the delay performance depends on the buffer size. Also, the parameter k in Eq. (3.13), which reflects the absolute value of the variance on a specific time scale, also has a crucial influence on the queueing performance. This issue will be further treated in Chapter 4 and Chapter 5.

3.2.3 Performance Evaluation for an Assembler

Relevant performance measures for an individual assembler include the assembly delay and the ratio of padding in case that fixed frame size or minimal frame size is concerned.

When the timeout period is specified, the assembly delay is always bounded by the timeout period. With the pure size-based assembly, the assembly duration T_{a} defined in Eq. (3.5) measures the assembly delay experienced by the first packet in a frame. In [HHA07], the assembly delay distribution was also analyzed with respect to an arbitrary packet in a frame.

The padding ratio can be directly derived from the frame size statistics introduced in Section 3.2.2.1. Izal et al. [IJAM06] specially studied the dependence between the padding ratio and the timeout period for the time-based, min-length assembly. Optimal solution was provided to achieve a large throughput and low assembly delay at the same time.

3.3 Scheduling

In comparison to the intensive studies on the scheduling in switching nodes, relatively little work has been done on the transmission scheduling in edge nodes. For simplicity, many research studies took an implicit assumption that data frames are sent to switching nodes directly from assemblers through channels with infinitely large bandwidth [IA02, YCQ02a] or limited bandwidth [HMQ+05], without considering the scheduling at all. In [CCK05], a bufferless transmission scheduler is included in the edge node model. In that case, however, the transmission scheduling in edge nodes does not differ from the scheduling in switching nodes.

A transmission scheduler equipped with an electronic buffer was analyzed in [LA04, Lee06, Zal06] with respect to the FCFS discipline. Lee [LA04] performed a Markov analysis to solve the queue length distribution. However, the input frame flows were modeled by artificial on/off sources without taking into account the influence from the traffic assembly. In [Lee06], the FIFO transmission queue was modeled by a GI/G/n delay system and the mean queueing delay was given by heavy load approximation. Zalesky [Zal06] investigated a special system with partitioned transmission buffers dedicated to individual wavelength channels. Each wavelength channel is exclusively allocated to one FEC.

Simulation studies were also carried out for the evaluation of delay performance of a transmission scheduler. In [HK06], the complementary cumulative distribution function (CCDF) was evaluated for the queueing delay in a FCFS frame scheduler. The influence of different assembly parameters was investigated. In [Car04], frames from different FECs are queued in separate buffers in the transmission scheduler. FCFS, round robin, weighted round robin (WRR) and deficit round robin (DRR) were inspected for the scheduling of head-of-line frames. A similar multi-queue scheduler was also investigated in [ROB04]. The schemes of *oldest burst first* (equivalent to FCFS), *longest queue first* and *random selection* were taken into consideration. The mean and variance of the frame queueing delay were evaluated.

For OBS networks with variable burst offset time set in the edge node, special scheduling algorithms were proposed in the edge node [Per06, LT07]. In principle, these schemes are the extensions of scheduling schemes in core switching nodes (e.g., Horizon, LAUC/-VF) and exploit the flexibility provided by the electronic buffers in edge nodes.

3.4 Admission Control

This section first introduces the relevant QoS requirements for the admission control in edge nodes. On this basis, the QoS problem investigated in this thesis is formulated.

3.4.1 Relevant QoS Requirements

As the network-to-network interface between a core OPS/OBS network and several client networks, the edge node has the important task in performing admission control for E2E QoS guarantee. Besides the QoS requirement on E2E loss performance associated with data channel contentions, as introduced in Section 2.5.3, performance issues on the signaling path (i.e., SCU performance) and E2E delay constraint should also be taken into account.

3.4.1.1 Performance Issue in the SCU

A crucial performance requirement on the SCU is that the transit delay of a frame header should not exceed the compensation delay inserted through the local FDL or offset time. Otherwise, the SCU cannot finish the header processing before the corresponding optical frame reaches the input of the switch fabric and the frame is lost. The resulting frame loss probability is equivalent to the overdue probability in the header processing. This can be calculated from the complementary probability distribution of the header waiting time in the SCU buffer. An M/D/1 delay model was proposed [JG03b, Zal06] for the waiting time analysis. Alternatively, an exponential approximation for the CCDF of the queueing delay was applied in [CCK05]. Based on the per-hop analysis, the end-to-end frame loss probability due to the overdue in header processing can be computed in the same way as introduced in Section 2.5.3.

3.4.1.2 E2E Delay Requirement

The QoS specification for E2E delay is essential for delay-sensitive services like real-time video/audio services. Here, the mean delay alone is not enough to assure the service quality. Special attention should be paid for the delay variation (or jitter). According to the ITU-T standard [Y.1541], the delay variation of a variable delay component is defined by the difference between the maximal and minimal value. In most statistical queueing analysis, the lower bound of delay is given by constant components such as propagation delay, timer and fixed processing time. However, the maximal delay can be infinite. For this concern, the delay upper bound is defined by a quantile (e.g., 0.999 quantile) of the delay distribution to characterize the jitter. In this thesis, the E2E delay performance is measured by the delay upper bound following the aforementioned notation.

In OPS/OBS networks, the E2E delay is mainly composed of assembly and scheduling delay in the ingress edge node, propagation delay throughout the core network and the sum of compensation delays for header processing at individual hops. Additional delay occurs in case FDLs are applied for the contention resolution. With a static routing strategy, the propagation delay is treated as constant. If the compensation delay is fixed (e.g., due to a fixed length of FDLs or a fixed setting of offset time) at each hop, the total compensation delay is also constant. The FDL delay for the contention resolution is always bounded by the maximal FDL length or constrained by the maximal number of re-circulations in a feedback FDL architecture. Subtracting these predictable delay components from the E2E delay budget, the delay bound requirement can be derived for the edge node, which guides the configuration of assembly parameters.

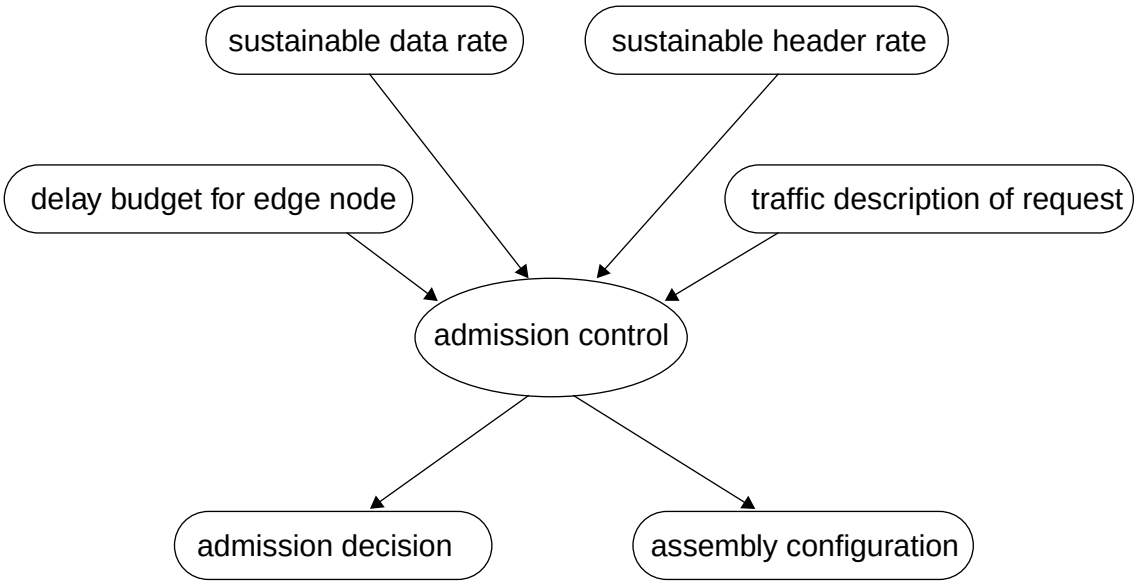


Figure 3.5: Admission problem in edge nodes

3.4.2 Admission Problem

Summarizing, the admission problem in an edge node is reduced to the problem illustrated in Fig. 3.5. The *sustainable data rate* and *sustainable header rate* (i.e., frame rate) reflect the maximal tolerable offered load for a new request on the E2E data path and signaling path respectively, with the purpose to meet with the requirement on the frame loss performance. They can be derived either from the static traffic engineering for each FEC (cf. Section 2.5.3.1) or through the dynamic probing supported by signaling protocols in the control plane (cf. Section 2.5.3.2). The input parameter of the delay budget is specified from the total delay budget as introduced in the previous subsection. The traffic description of the arriving connection request includes not only the maximal offered traffic but also other important traffic characteristics.

With these input parameters, the admission control decides whether the request can be accepted or not. Furthermore, the assembly parameters have to be properly configured to assure the QoS requirements and an efficient resource utilization. This is the central problem that will be investigated in the following chapters.

In [WZV02, CCK05], similar problems were studied without considering the transmission scheduler in the edge node. In this thesis, the queueing delay in the transmission scheduler is specially analyzed with respect to input traffic showing different characteristics on multiple time scales. Also, it is always kept in mind that the QoS requirements should be satisfied not only at the maximal traffic intensity of a request, but also for light traffic. This is an issue because with traffic assembly the worst performance does not necessarily occur at the maximal traffic intensity. An intuitive example is that lower traffic rates can result in larger delay in the edge node due to the increased assembly latency with a size-based scheme [HK06, Kan06].

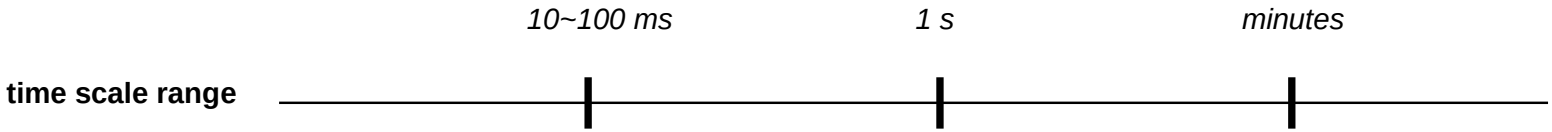
4 Characteristics of Client Traffic and Methods for Queueing Analysis

Today's network traffic is more and more dominated by the data traffic. For the remarkable flexibility of IP-based data networks, the IP layer is becoming a convergence layer for the provisioning of multi-services in a layered protocol architecture. The heterogeneous user behaviors and network control mechanisms/protocols, on the other hand, give rise to complex traffic characteristics on multiple time scales, which have significant impacts on the network performance. This issue must be taken into account in the performance analysis of OPS/OBS transport networks. Correspondingly, analytical methods capable of dealing with various traffic behaviors on multiple time scales are necessary, not only for the inherent time-scale-dependent characteristics of the client traffic, but also for the traffic pattern newly introduced by the traffic assembly in the OPS/OBS edge node.

This chapter gives a review on the traffic characterization in backbone data networks and on the respective methods for the queueing analysis. In Section 4.1, the traffic properties on multiple ranges of time scales are surveyed with regard to their mathematical definitions, causes and performance impacts, etc.. In Section 4.2, the M/Pareto model is introduced as a widely used model for aggregated traffic (cf. Chapter 5). Section 4.3 inspects the analytical methods that are able to capture the performance impacts on multiple time scales.

4.1 Traffic Characteristics

An overview of traffic characteristics in wide area networks (WAN) or backbone networks is illustrated in Fig. 4.1 over the spectrum of time scales. Starting from the small time scale, the traffic shows uncorrelated property, multifractal behavior, long range dependence (or self-similarity), nonstationarity and periodicity in the four consecutive ranges of time scales. Here, the traffic characteristics in a specific range of time scales refer to the statistical features that can be extracted if the traces are analyzed in a time window with its duration located in the time-scale range. Note that the boundary time scales between the ranges in Fig. 4.1 are only roughly specified because they depend very much on the networks under measurement and are also subject to changes with the network upgrade and evolution. In the following subsections, the characteristics are inspected for the time-scale ranges, respectively.



characteristics	uncorrelated property	multifractal property	long range dependence (self-similarity)	nonstationarity and periodic pattern
statistical measures	marginal distributions of interarrival time and size of data unit	high order statistical moments	the first and second statistical moments	macro-level average traffic rates, workload patterns, evolution trend, etc.
causes	multiplexing of independent flows	impacts of flow control protocols (e.g., TCP)	heavy-tailed distribution in the behaviors of individual traffic sources	seasonal effects
traffic models	renewal point processes (e.g., Poisson process)	cascade model, Lindenmayer system	on-off superposition model, M/Pareto model, FGN, MMPP	parametric and non-parametric models
performance impacts	efficient buffering	particularly worse queueing performance	particularly worse performance at high loads, good predictability for dynamic bandwidth allocation	traffic forecasting, network planning

Figure 4.1: Traffic characteristics on different time scales

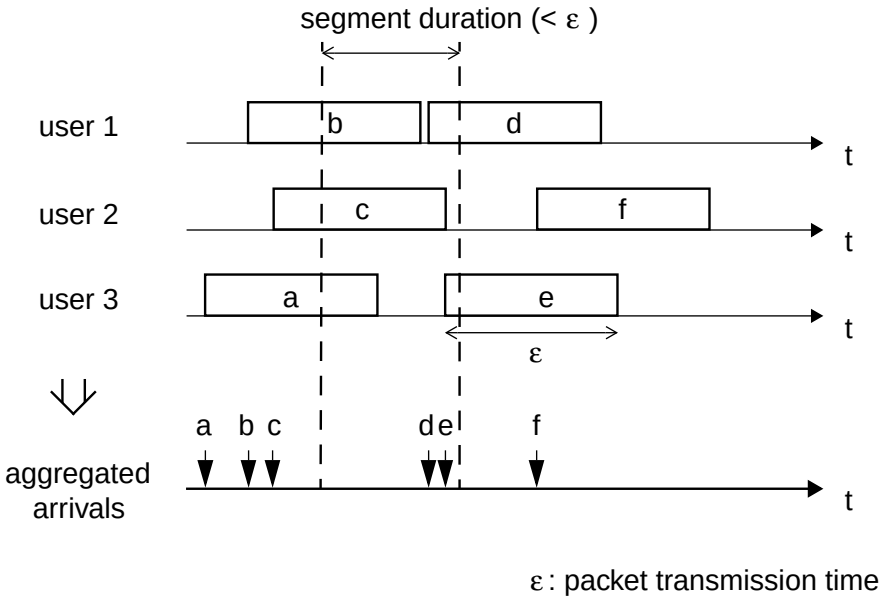


Figure 4.2: Aggregated packet arrivals from three users

4.1.1 Uncorrelated Property

In small time scales up to $10 \sim 100$ ms, the uncorrelated property was reported in the latest measurements of Internet backbone traffic [ZRMD03, KMFB04]. The reason for this phenomenon can be understood by the superposition model depicted in Fig. 4.2.

It is assumed here that three users send packets through their separate access links independently. They have the same link bandwidth and the packet size is constant. Naturally, the packet interarrival time within a flow from an individual user cannot be smaller than the packet transmission time on the access link. Taking the start time of each packet as the arrival instant, an aggregated arrival process is constructed from the superposition of these three individual packet flows. It can be seen that in any arbitrary time interval smaller than the packet transmission time, each flow has maximal one arrival. Therefore, in the aggregated process the arrivals within this time segment are independent of each other because they are originated from independent sources. The number of aggregated arrivals within the small time segment actually follows a Binomial distribution [TWJ02, Hu04]. In case the traffic is aggregated from a large number of users, the distribution converges to a Poisson distribution. Any correlation structures within individual flows come into effect in the aggregated process only on time scales beyond the packet transmission time.

The above discussion also illuminates that the time-scale range of the uncorrelated property depends on the access link rate, which is consistent with the measurement results [ZRMD03]. With the ever increasing bandwidth in access networks, this range is supposed to shift to even smaller time scales in the future.

The uncorrelated property in the small time scales justifies the traffic models on the basis of renewal point processes, typically the Poisson process, depending on the inspected system scenarios. This indicates a considerable buffering efficiency because with Markovian arrival pro-

cesses the overflow probability decreases exponentially fast with the increase of buffer size asymptotically [CT95, Kel96].

4.1.2 Multifractal

Multifractal was proposed to characterize the non-trivial high order statistics in the hundreds of milliseconds range [FGW98, ENNS00]. Following the notation in Section 3.2.2.2, let X_t denote the traffic volume in an arbitrary time interval of t under the assumption that the traffic process is stationary. The traffic is multifractal if X_t satisfies [ENNS00, Gil01] [1] :

$$E[(X_t - E[X_t])^q] = C(q) \cdot t^{\tau(q)+1} \tag{4.1}$$

for t within the time-scale range and any positive values of q up to a certain bound. Here, $C(q)$ and $\tau(q)$ are both deterministic functions. Typical multifractal processes have nonlinear $\tau(q)$ in q. If $\tau(q)$ is linear in q, the process degenerates to so-called monofractal.

The multifractal behavior of data traffic indicates highly irregular local burstiness along the time axis, which is attributed to closed-loop flow control mechanisms such as the congestion control of the transmission control protocol (TCP) [FGHW99]. Typically, the time scales on which the multifractal emerges are correlated to the round trip times within the TCP connections.

Modeling of multifractal traffic generally resorts to iterative procedures such as cascading [FGW98] and construction of stochastic Lindenmayer systems [SNV02]. Performance evaluations [ENNS00] showed that multifractal can lead to very worse queueing behaviors even at low system loads.

However, with the rapidly increasing bandwidth of backbone links, the relevance of the multifractal traffic in the backbone is becoming questionable. More recent measurements [ZRMD03] on OC3/12/48 links negated the existence of multifractal in the highly aggregated traffic.

4.1.3 Long Range Dependence

The concept of long range dependence (LRD) has been briefly introduced in Section 3.2.2.2 as a large-time-scale behavior of stationary traffic processes. In literature, self-similarity is frequently used as a synonym of LRD in the sense of asymptotic second-order self-similarity [WTSW97, PW00], which is also adopted in this thesis. However, it is worth to point out that self-similarity in the general sense is not equivalent to LRD. Especially, self-similarity is a kind of monofractal with $\tau(q) = qH - 1$ in Eq. (4.1), where H is the Hurst parameter. Accurate definitions of self-similarity and LRD can be found in [PW00, WPRT02].

The LRD property in the aggregated traffic is caused by the heavy-tailed distributions of the behaviors of individual traffic sources. From the perspective of the network, a source can be a network terminal, an application program or a real person. Heavy-tailed distribution is a

[1] Multifractal is defined here by means of the RV X_t instead of the cumulative traffic process as the case in [ENNS00, Gil01]. The definitions are equivalent under the condition that the cumulative process has stationary increments. Also, note the implicit condition $E[X_t] = 0$ in [Gil01].

special type of distribution which has finite mean value and infinite variance. Physically, a heavy-tailed distribution means nonnegligible probabilities that an RV takes very large sample values. Willinger et al. showed in the famous pioneering work [WTSW97] that self-similar Ethernet traffic can be modeled by the aggregation of a large number of on-off sources with the on- and off-periods following heavy-tailed distributions. Further on, it was discovered that in many session-based applications such as TCP, FTP and Web, the transmission durations of individual sessions are heavy-tailed distributed while the arrivals of session requests follow Poisson processes [PF95, CB97, Fel00].

A basic theoretical model for X_t of the LRD traffic is the fractional Gaussian noise (FGN) [PW00], or equivalently, fractional Brownian motion (FBM) when the cumulative process is concerned. This model is often preferred in the fluid-flow analysis, not only for the mathematical elegance of its marginal Gaussian distribution, but also for its good conformity with the Gaussianity measured in the real aggregated traffic [ENNS00, KN02, ZRMD03].

As for the discrete-event modeling (e.g., applied in simulations) of LRD traffic, two approaches have been taken. The "white-box" approach [WPRT02] takes advantage of the insights into the underlying source-level behaviors and structurally synthesizes the LRD traffic. The on-off superposition model and M/Pareto model fall in this scope. Both models converge to FGN processes in large time scales, which enables the asymptotic fluid-flow analysis for the queueing performance. On the contrary, the "black-box" modeling does not consider the causal relations between the source behaviors and the LRD. It relies on standard traffic models, e.g., Markov modulated Poisson process (MMPP), to approximate the LRD property by parameter matching. Although the matching procedure is generally not trivial [MMM$^+$05], this approach has the advantage to exploit the conventional discrete-event queueing theory for the performance evaluation.

Intuitively, the LRD indicates that in traffic traces the periods with high instantaneous rates tend to cluster together. The same holds for the periods with small traffic rates. Performance studies showed that this leads to a low buffer efficiency at high system loads, i.e., a heavy-tailed CCDF of the queue length [ENW96]. Correspondingly, an efficient performance improvement is achieved by increasing the bandwidth instead of providing a larger buffer [PKC97]. On the other hand, the correlation structure of LRD traffic can be exploited for the traffic prediction in dynamic bandwidth allocations and congestion controls [TP00].

4.1.4 Nonstationarity and Periodicity

Beyond the time scale in minutes, traffic generally cannot be treated as stationary processes any more. This range of time scales can be further divided into several scopes for different traffic characteristics. For example, the inherent nonstationary behaviors from hour to hour in a day, periodic patterns on a daily or weekly basis, and the long-term traffic growth trend in months and years.

Traffic analysis here aims to provide the forecasting of traffic demands to guide the link upgrade and network planning. Therefore, it focuses on the traffic rates averaged over very large time scales from tens of minutes to hours or beyond. Conventional methods for the time-series analysis can be applied [BD02]. For example, the autoregressive integrated moving average

(ARIMA) is a standard parametric model for nonstationary processes, which was applied for long-term forecasting of Internet backbone traffic in [PTZD03]. Non-parametric methods were proposed in [LSX+03] to identify and classify periodic traffic demand patterns.

4.2 M/Pareto Model for the Client Traffic

In order to support the QoS provisioning in OPS/OBS edge nodes, queueing analysis has to be performed with respect to the incoming client traffic. To this end, only stationary traffic processes are considered. Because the traffic inflow in an edge node is supposed to be at a high aggregation level, the multifractal characteristic can be ignored. In order to capture the uncorrelated property on small time scales and the LRD on large time scales concurrently, either the on-off superposition model or the M/Pareto model is applicable. Considering the fact that the on-off superposition model becomes very similar to the M/Pareto model as the number of sources grows large, the attention is concentrated on the M/Pareto model in this thesis.

4.2.1 Parameters

M/Pareto process [NZA99] is a "white-box" structural model reflecting the traffic behaviors of the session-based applications in Internet. With this model, data sessions are generated according to a Poisson process with an average rate of λ_s. The traffic volume of a session is denoted by an RV B following a Pareto distribution:

$$P\{B \leq b\} = \begin{cases} 1 - (\frac{\kappa}{b})^{\alpha} & \text{for } b \geq \kappa \\ 0 & \text{for } b < \kappa. \end{cases} \tag{4.2}$$

Here, κ represents the minimal value of B and α is the shape parameter that measures the degree of the variability. The smaller the shape parameter is, the larger is the variability. For LRD traffic modeling, $\alpha : 1 < \alpha < 2$ is adopted which leads to a heavy-tailed distribution with a finite mean value and an infinite variance. The mean traffic volume φ of a session is calculated by:

$$\varphi = E[B] = \frac{\kappa\alpha}{\alpha - 1}. \tag{4.3}$$

In each session, the data is sent at a constant rate of c_a. c_a can be regarded as the capacity of the access link on which each source transmits the session. In this sense, the M/Pareto process is an inherent model for aggregated traffic.

The above construction results in a fluid-flow M/Pareto model. Its discrete-event variation goes a step further by segmenting each session into a series of packets of equal sizes l_{max} and transmitting them back-to-back. Although the last packet of a session can have a smaller size than the fixed packet size l_{max}, this kind of packets amounts to only a small portion (<10% in general) of the total traffic volume [Hu04]. Approximately, the packet size can be regarded as constant.

In the remainder of the thesis, the notation M/Pareto refers always to the discrete-event M/Pareto model, unless explicitly stated otherwise.

4.2.2 Properties of the Variance

The time-scale-dependent characteristics of the M/Pareto traffic can be illuminated by observing the variance of the RV X_t. VAR$[X_t]$ will be hereafter referred to as *variance structure* or *variance process* interchangeably. It was derived in [NZA99, BC00] for the original fluid-flow paradigm. In case of the discrete-event M/Pareto model, it is conceivable that the discretization leads to the Poissonian property on the small time scales up to the unit packet transmission time on an access link [Hu04], following the principle explained in Section 4.1.1. On larger time scales, however, the variance structure is not influenced by the discretization very much. The time scale $t = l_{\max}/c_{\mathrm{a}}$ is thus called *critical time scale* because it acts as the boundary between the time-scale range in which the traffic exhibits the Poissonian property and the range in which the correlation arises in the traffic trace. Summarizing, it leads to:

$$\mathrm{VAR}[X_t] \approx \begin{cases} \varphi \cdot \lambda_{\mathrm{s}} \cdot t \cdot l_{\max} & \text{for } t < \frac{l_{\max}}{c_{\mathrm{a}}} \\ \lambda_{\mathrm{s}} c_{\mathrm{a}}^2 \cdot \left(\frac{\alpha}{\alpha-1} \cdot \frac{\kappa t^2}{c_{\mathrm{a}}} - \frac{t^3}{3}\right) & \text{for } \frac{l_{\max}}{c_{\mathrm{a}}} \le t \le \frac{\kappa}{c_{\mathrm{a}}} \\ t^{3-\alpha} \cdot \frac{2\lambda_{\mathrm{s}} c_{\mathrm{a}}^2 \kappa^{\alpha}}{(\alpha-1)(2-\alpha)(3-\alpha)c_{\mathrm{a}}^{\alpha}} - t \cdot \lambda_{\mathrm{s}} \kappa^2 \frac{\alpha}{2-\alpha} + \lambda_{\mathrm{s}} \frac{\alpha \kappa^3}{3c_{\mathrm{a}}(3-\alpha)} & \text{for } t > \frac{\kappa}{c_{\mathrm{a}}}. \end{cases} \quad (4.4)$$

It is assumed here that $l_{\max} < \kappa$. The first line on the right-hand side of Eq. (4.4) is derived directly from the property of the Poisson packet arrival process, i.e.,

$$\mathrm{VAR}[X_t] \approx \mathrm{VAR}[\frac{X_t}{l_{\max}}] l_{\max}^2 = E[\frac{X_t}{l_{\max}}] l_{\max}^2 = \frac{\varphi \cdot \lambda_{\mathrm{s}} \cdot t}{l_{\max}} \cdot l_{\max}^2 = \varphi \cdot \lambda_{\mathrm{s}} \cdot t \cdot l_{\max}.$$

It is an approximation because the packet size is assumed to be fixed to $l_{\max}$. The second and third lines are the approximations on the basis of the variance for a fluid-flow M/Pareto model [NZA99, BC00].

In Fig. 4.3, the accuracy of the approximations in Eq. (4.4) is verified by comparing with simulations for the M/Pareto traffic with an average rate of 450 Mbps. Different values of the shape parameter α and the access link rate c_{a} are checked. The approximations show a very good quality. It is noticeable that the parameter c_{a} has very significant impacts on the variance structure.

Observing the third line of Eq. (4.4), the term with the highest order of t dominates when $t \to \infty$. Hence, VAR$[X_t] \sim t^{3-\alpha}$. Compared with Eq. (3.13), this indicates the LRD with a Hurst parameter $H = (3-\alpha)/2$.

The asymptotic behavior can be clearly illustrated by a variance-time plot, which is in practice widely used to measure the Hurst parameter. An example is given in Fig. 4.4 for the synthetic traffic with $\alpha = 1.4$ and $l_{\max} = 1000$ bytes. Here, VAR$[X_t]$ is plotted with respect to discrete time scales $t = 2^j \cdot 0.02$ ms ($j = 1, 2, \ldots$) in a log-log diagram, where 0.02 ms is the basic time scale. For the LRD property, it is straightforward that $\log_2(\mathrm{VAR}[X_t]) \sim 2H \cdot j$ from Eq. (3.13). It can be seen that on the small time scales, the variance follows the dashed reference line that represents a linear relationship between the variance and the time scale, i.e., $H = 0.5$ for the uncorrelated property [2]. On the large time scales, VAR$[X_t]$ deviates from the dashed line and

[2] Strictly speaking, the Hurst parameter should always be measured on the large time scales according to its definition. For the ease of presentation, H is also applied here to describe the growing speed of the variance on the small time scales.

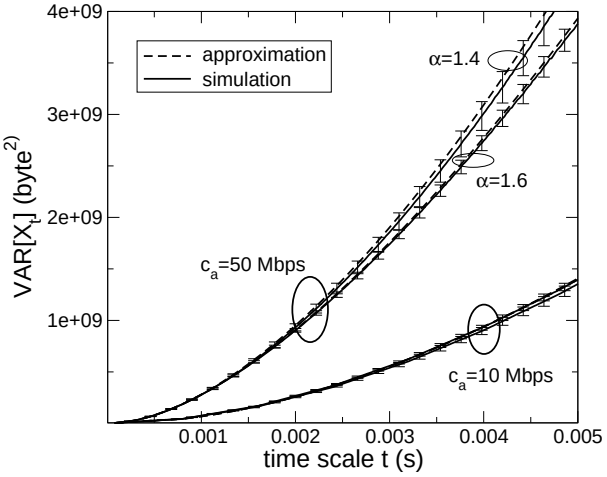


Figure 4.3: Variance with respect to the time scale: approximation vs. simulation

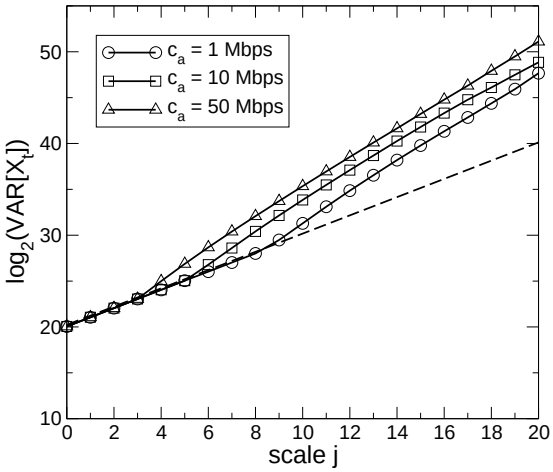


Figure 4.4: Variance-time plot for X_t ($\alpha = 1.4$, $l_{\max} = 1000$ bytes, $t = 2^j \cdot 0.02$ ms)

grows more rapidly, indicating the dominance of the LRD. Applying linear regression on the large time scales, $H = 0.8$ can be derived from the slope of the regression line. The separating point between the time-scale regions of the uncorrelated property and the LRD depends on the access link rate c_a, which is consistent with the boundary condition in Eq. (4.4).

4.3 Queueing Analysis on Multiple Time Scales

The traffic behaviors on multiple time scales can be taken into account of the queueing analysis in two approaches. The time scale decomposition follows the principle of *divide and conquer*. Different traffic models are applied for the different scopes of time scales and solved separately. On the contrary, the methods of integrated analysis explicitly quantify the time-scale-dependent traffic characteristics by a deterministic function of the time scale. From this function, a *relevant time scale* is determined for the level of the queue length of interests, on the basis of which the performance is evaluated. This virtually converts the queueing analysis to an optimization problem.

In the following subsections, the analytical methods are introduced for a single server delay system with FIFO discipline. The mean traffic rate is r. The transmission rate of the server is constant and denoted by c. RV Q stands for the queue length or unfinished work [RMV96] in the paradigm of fluid-flow models.

4.3.1 Time Scale Decomposition

In the performance study for ATM networks, it was recognized that the aggregated cell traffic can be modeled on three time-scale levels: cell scale, burst scale and call scale [RMV96]. On the cell scale, point processes are applied to capture the randomness in the cell arrivals. The traffic behavior is dominated by the effect of the multiplexing of independent user flows, similar to the small-time-scale behavior introduced in Section 4.1.1. For example, $M/D/1$ and

$nD/D/1$ systems are representative cell-scale models. On the burst scale, the traffic variabilities in individual flows exhibit more influences on the queueing performance. Fluid-flow models are typically used on this scale, e.g., the FBM for the LRD property. On the call scale, the request arrival pattern and call holding time become the relevant traffic parameters and the system is modeled by a circuit-switched loss system.

From the perspective of QoS provisioning, the call-blocking probability, which is classified as *call-level QoS* [KMK04], solely relies on the call-scale model. However, the cell queueing delay, or buffer overflow probability is recognized as *in-call QoS* and influenced by both the cell-scale and burst-scale models. Therefore, it is necessary to combine the solutions to the models on both scales. Supposing that the unfinished work Q consists of a cell component Q_{C} and a burst component Q_{B}. The CCDF of Q is formulated as [NRSV91]:

$$\begin{aligned} P\{Q>q\} &= P\{Q_{\mathrm{C}}+Q_{\mathrm{B}}>q\} \\ &= P\{Q_{\mathrm{C}}>q|Q_{\mathrm{B}}=0\}\cdot P\{Q_{\mathrm{B}}=0\}+ \\ &\quad P\{Q_{\mathrm{C}}+Q_{\mathrm{B}}>q|Q_{\mathrm{B}}>0\}\cdot P\{Q_{\mathrm{B}}>0\}. \end{aligned} \tag{4.5}$$

In practice, an accurate computation according to Eq. (4.5) can be complex. For this reason, Norros et al. suggested the following approximation [NRSV91]:

$$P\{Q>q\}\approx \max(P\{Q_{\mathrm{C}}>q\},P\{Q_{\mathrm{B}}>q\}). \tag{4.6}$$

Here, the CCDFs of Q_{C} and Q_{B} are obtained from the cell-scale model and the burst-scale model, respectively.

4.3.2 Integrated Analysis

In this subsection, two important methods are introduced for the integrated queueing analysis. The methods are based on similar principles, which can be understood by looking at the queueing process $Q(s)$ that denotes the unfinished work in the buffer at an arbitrary time instant s.

Let random process $A(s)$ refer to the cumulative traffic amount up to time instant s. The time s can have negative values. According to Reich's formula [RMV96, KMK04], it holds that:

$$Q(s)=\sup_{t\geq 0}(A(s)-A(s-t)-c\cdot t). \tag{4.7}$$

The physical meaning of this equation is not quite straightforward. Though, things become clear when it is realized that a supremum is obtained when $s-t$ corresponds to the start time of the current busy period that covers the time instant s [KMK04]. In other words, it is the case that t is equal to the backward recurrence time of the busy period.

Statistical derivation of $Q(s)$ in the form of Eq. (4.7) is, however, difficult in practice. This motivates the application of the following approximation [RMV96, KMK04]:

$$P\{Q(s)>q\}\approx \sup_{t\geq 0}P\{A(s)-A(s-t)-c\cdot t>q\}. \tag{4.8}$$

Since the stationary traffic process is considered, $A(s)-A(s-t)$ is equivalent to X_t defined as the traffic arrivals in the time interval t in the previous sections. So, the derivation of the CCDF is independent of the time instant s. Consequently, the random process $Q(s)$ can be substituted by the RV Q. Eq. (4.8) is rewritten as:

$$P\{Q > q\} \approx \sup_{t \geq 0} P\{X_t - c \cdot t > q\}. \tag{4.9}$$

This equation suggests that for each value of q a time scale $t = \tau_q$ should be found to maximize the probability of the event $\{X_t - c \cdot t > q\}$. τ_q is thus called the *relevant time scale* for q. Physically, it stands for the time duration over which the event $\{X_t - c \cdot t > q\}$ is most likely to occur. Eq. (4.9) can be solved by the effective bandwidth method and the maximum-variance-asymptotic (MVA) approach, respectively.

4.3.2.1 Effective Bandwidth Method

The effective bandwidth method is based on the theory of large deviations [RMV96]. Applying Chernoff's bound for the term $P\{X_t - c \cdot t > q\}$ on the right-hand side of Eq. (4.9), it is derived [Kel96, RMV96]:

$$P\{Q > q\} \approx \beta \ \exp(\sup_{t \geq 0} \inf_{\theta > 0} (\theta \cdot t \cdot \Lambda(\theta, t) - \theta \cdot (q + c \cdot t))) \tag{4.10}$$

where β is a constant independent of q and is referred to as the *asymptotic constant* following the notation in [CLW96]. $\Lambda(\theta, t)$ is called the *effective bandwidth* of the traffic process and defined as [Kel96]:

$$\Lambda(\theta, t) = \frac{1}{\theta \cdot t} \ln(E[\exp(\theta \cdot X_t)]). \tag{4.11}$$

Note that the effective bandwidth is a deterministic function of θ and t and is very close to the moment generating function $E[\exp(\theta \cdot X_t)]$. In this sense, $\Lambda(\theta, t)$ can be regarded as a transformation function of X_t that comprises the statistical information over the time scale t. Kelly gave in [Kel96] a survey of the effective bandwidths for a variety of traffic processes. $\Lambda(\theta, t)$ can also be determined through traffic measurements [CSS99] so as to avoid the necessity to build parametric traffic models.

The asymptotic constant β in Eq. (4.10) has a relative complex relationship with system- and traffic-parameters, which needs to be specially studied. For simplicity, it is quite common in the admission control that β is set to 1 for a conservative estimation. More elaborate determinations of β can be found in [CLW96, MV96, RMV96, BC00].

In literature, Eq. (4.10) is referred to as *many source asymptotic* of the effective bandwidth theory [CSS99], because the accuracy becomes better when more sources are aggregated in the input traffic. Therefore, it is very suitable for the admission control in backbone networks. However, the computational complexity is relative high due to the sup-inf condition.

4.3.2.2 Maximum-Variance-Asymptotic Approach

The traffic in transport networks is aggregated from a large number of user flows. According to the central limit theory, the distribution of X_t can be approximated by the Gaussian distribution, which has been verified by plenty of traffic measurements [ENNS00, KN02, ZRMD03]. Let r denote the mean traffic rate. Then, X_t follows the Gaussian distribution $N(r \cdot t, \mathrm{VAR}[X_t])$. Since the transmission rate c is constant, $X_t - c \cdot t$ has also a Gaussian distribution $N((r-c) \cdot t, \mathrm{VAR}[X_t])$. In this way, the term $P\{X_t - c \cdot t > q\}$ in Eq. (4.9) is actually the CCDF of the Gaussian distribution. To present the result in the form of a standard Gaussian distribution $N(0,1)$, the following conversion is taken [NW98]:

$$g(q,t) = \frac{q + (c-r) \cdot t}{\sqrt{\mathrm{VAR}[X_t]}}. \tag{4.12}$$

So, Eq. (4.9) is solved as [NW98, CS98]:

$$P\{Q > q\} \approx \sup_{t \geq 0} \frac{1}{\sqrt{2\pi}} \int_{g(q,t)}^{\infty} \exp(-\frac{g^2}{2})\, \mathrm{d}g. \tag{4.13}$$

The supremum on the right-hand side of Eq. (4.13) is obtained by minimizing $g(q,t)$ over the time scale t as long as $\mathrm{VAR}[X_t]$ is specified. The time scale on which the minimum of $g(q,t)$ is found corresponds to the relevant time scale τ_q, i.e.,

$$\tau_q = \arg \inf_{t \geq 0} g(q,t) \tag{4.14}$$

Eq. (4.13) was actually shown to be an asymptotic lower bound for the tail probability of the queue. Choe et al. [CS98] extended the analysis and deduced an asymptotic upper bound:

$$P\{Q > q\} \approx \beta \, \exp(-\frac{g(q,\tau_q)^2}{2}) \tag{4.15}$$

where $g(q,\tau_q)$ stands for the minimum of $g(q,t)$ and β is the asymptotic constant similar to that in Eq. (4.10).

In comparison to the effective bandwidth method, the MVA approach simplifies the analysis by assuming that X_t follows the Gaussian distribution. In this way, the traffic is solely characterized by the variance process $\mathrm{VAR}[X_t]$ instead of the effective bandwidth $\Lambda(\theta,t)$. In [KS99], the performance of different analytical methods for admission control was experimentally compared. The MVA exhibits a very good accuracy not only for the asymptotic tail behavior of the queue, but also in the range of small queue lengths. Choe et al. showed in [CS98] that in spite of the assumption of the Gaussian distribution the MVA upper bound also serves as a good estimation for many non-Gaussian traffic processes. Nevertheless, care should be taken for the situations in which the approximation of the Gaussian distribution becomes questionable. This is the case typically when the traffic of small aggregation degrees (e.g., flows in access networks) is concerned or the system operates at a low load (i.e., light traffic).

5 A Novel Approach for the Multi-Scale Queueing Analysis

In Chapter 4, the time scale decomposition and integrated analysis are introduced as two fundamental approaches for the multi-scale queueing analysis. Each has its advantages and disadvantages. By combining the cell-scale and burst-scale component according to Eq. (4.6), the time scale decomposition is intuitive and simple in the computation to deduce the overall queueing performance. How the system is modeled and solved on the cell scale and burst scale respectively, however, is another problem, which is not necessarily simple in practice. Typically, without an explicit inclusion of the time scale factor in traffic models, it is difficult to quantitatively distinguish the boundary between the cell scale and burst scale for a general traffic process. This problem is solved in the integrated analysis with the introduction of the relevant time scale concept. Noticeably, in both the effective bandwidth method and MVA method, the time scale is modeled explicitly in the traffic description, as referred in Eq. (4.11) and (4.12). Nevertheless, the computational complexity of the integrated analysis is relatively high due to the search for the relevant time scale τ_q for the tail probability $P\{Q > q\}$.

On the basis of the time scale decomposition method and the relevant time scale concept of the integrated analysis, a novel multi-scale queueing analysis is proposed that integrates the advantages of both fundamental approaches. This method will play a central role in the performance analysis in the edge node, as will be seen in Chapter 6. In Section 5.1, the theoretical principle of the method is explained and the solving procedure is introduced. An application example is then given in Section 5.2 for the queueing analysis with the M/Pareto traffic. Section 5.3 summarizes the novelty of the proposed method. For presentation, a single server delay system is studied and the same notations as those in Section 4.3 are to be used.

5.1 Introduction of the Method

5.1.1 Principle

Looking into Eq. (4.10) and Eq. (4.15), it is realized that both effective bandwidth and MVA approach can lead to a closed-form solution of the CCDF of the queue length $P\{Q > q\}$ as long as a closed-form solution to the relevant time scale τ_q is available. This is theoretically feasible when the effective bandwidth $\Lambda(\theta,t)$ or the variance process $\mathrm{VAR}[X_t]$ of the incoming traffic satisfies some conditions. For example, if $\mathrm{VAR}[X_t]$ is differentiable for all $t \geq 0$ and has a continuous derivative, τ_q can be obtained from Eq. (4.12) by solving $\partial g(q,t)/\partial t = 0$.

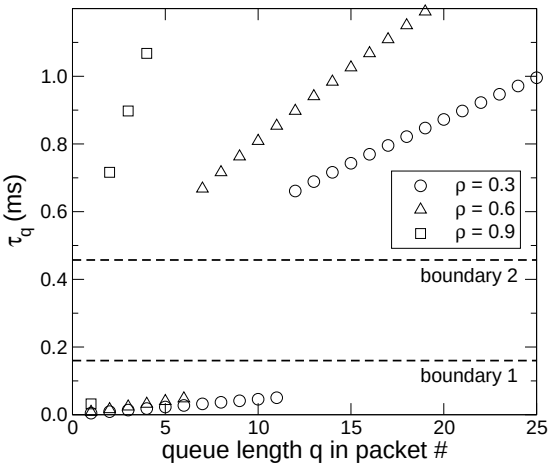


Figure 5.1: Distribution of relevant time scales for M/Pareto model ($c_{\mathrm{a}} = 50$ Mbps)

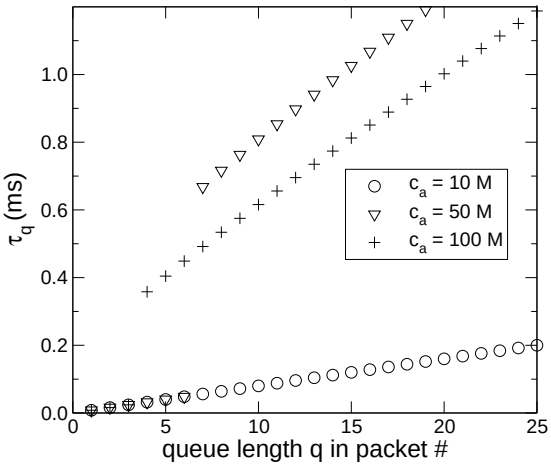


Figure 5.2: Impact of c_{a} on the distribution of relevant time scales for fixed $\rho = 0.6$

In the realistic backbone traffic, the above conditions are not satisfied for general $t \geq 0$. However, according to the classification of traffic properties on the multiple ranges of time scales in Chapter 4, in the respective scopes of time scales, $\Lambda(\theta, t)$ or VAR$[X_t]$ can be piece-wise approximated by such "regular" functions, for example, VAR$[X_t]$ in Eq. (4.4) for the M/Pareto model. In general, the relevant time scale τ_q increases monotonically with q. Therefore, for a scope of queue length $q : a \leq q \leq b$, it leads to the correspondent relevant time scales $\tau_q : \tau_a \leq \tau_q \leq \tau_b$. If τ_a and τ_b are located in the same time-scale region of the piece-wise approximation, a closed-form solution can be obtained for τ_q as well as for the tail probability $P\{Q > q\}$ within the domain $q : a \leq q \leq b$.

In Fig. 5.1, example scenarios are studied to illustrate the above idea. Consider a backbone channel of $c = 2.5$ Gbps with an unbounded FIFO queue. The incoming traffic is modeled by the M/Pareto process with the Hurst parameter $H = 0.8$, the mean traffic volume of a session $\varphi = 10$ KBytes, the maximal packet size $l_{\max} = 1000$ bytes and the access link rate $c_{\mathrm{a}} = 50$ Mbps. The distribution of the relevant time scale τ_q with respect to the queue length q is plotted for the offered loads $\rho = 0.3, 0.6, 0.9$, respectively. Here, τ_q is calculated numerically from Eq. (4.14) following the MVA approach. q is measured in the number of packets. The two dashed horizontal lines mark the boundaries between the time-scale ranges in the piece-wise approximation of VAR$[X_t]$ in Eq. (4.4). Boundary 1 refers to the time-scale level of $l_{\max}/c_{\mathrm{a}}$. Recall that below this level, the traffic exhibits Poisson-alike properties. Boundary 2 denotes the level of κ/c_{a}, above which the LRD starts to appear. The segment between Boundary 1 and Boundary 2 represents a transit phase between the small-time-scale behaviors and large-time-scale behaviors.

It is seen from Fig. 5.1 that τ_q increases monotonically with the increase of q. Taking $\rho = 0.6$ as the example, there is an obvious *critical point* between $q = 6$ and $q = 7$. For $0 \leq q \leq 6$ the relevant time scales are all below Boundary 1 and are distributed in a quite regular way. According to the MVA analysis, this indicates that the CCDF $P\{Q > q\}$ with $0 \leq q \leq 6$ is decided by the traffic characteristics on the time scales below Boundary 1. Similarly, for $q \geq 7$, the relevant time scales are regularly located above Boundary 2. Hence, the large-time-scale traffic behaviors dominate the queueing performance. This leads to the idea to solve the CCDF

by using the pure Poisson model in the domain $0 \leq q \leq 6$ and using the FBM model in the domain $q \geq 7$. As a result, a piece-wise closed-form solution is achieved for $P\{Q > q\}$.

Similar behaviors are also observed for the light load ($\rho = 0.3$) and the heavy load ($\rho = 0.9$). Comparing the three load situations, it is found that the critical point in the τ_q distribution shifts to the small queue length when the system load grows. This means that the large-time-scale traffic characteristics gain more significance in the determination of the queueing performance.

Fig. 5.2 depicts the influence of the access rate c_a of the M/Pareto model on the τ_q distribution. ρ is fixed to 0.6. c_a is set to 10 Mbps, 50 Mbps and 100 Mbps, respectively. It shows that a large c_a makes the critical point located at a small queue length. This feature reflects itself in the resulting CCDF of the queue length, as will be seen in Section 5.2.

5.1.2 Solving Procedure

Section 5.1.1 justifies the piece-wise approximation of $P\{Q > q\}$, which actually generalizes the time scale decomposition method by the introduction of the relevant time scale concept of the integrated analysis. The proposed multi-scale queueing analysis is completed in the following steps:

1. Characterize the traffic in terms of a time-scale-dependent measure, e.g., the effective bandwidth $\Lambda(\theta, t)$ or the variance process VAR$[X_t]$. In case the marginal distribution of the traffic process can be approximated by a Gaussian distribution, VAR$[X_t]$ is preferred for its simplicity.
2. Identify the different traffic behaviors over the time scales and construct the traffic models for different time scales accordingly. For the backbone traffic, the Poisson process is taken to model the small-time-scale behaviors and the FBM is used to model the LRD on the large time scales. These models are called *submodels* of the traffic process.
3. Separately derive the closed-form solution for each submodel either by the effective bandwidth approach or by the MVA method. The solution for each submodel is valid for the system under the study in a specific scope of the queue length.
4. Derive the overall tail probability of the inspected queue by concatenating the results obtained in Step 3.

As a further extension, methods other than the integrated analysis approaches can be applied to solve the submodels of the traffic in Step 3, as long as this simplifies the computation or leads to a more accurate solution. This provides a highly flexible and intuitive way to solve the multi-scale queueing problem on the basis of the standard solutions to individual submodels.

5.2 Application for the M/Pareto Traffic

According to the properties of the M/Pareto traffic in Section 4.2.2, the $M/D/1$ model is used as the submodel on small time scales. On large time scales, the FBM is taken to model the LRD

property. Both are standard models that can be solved by a variety of methods. In the following subsections, the solutions to the submodels are introduced. Simulations are carried out to verify the accuracy of the analysis in the evaluation of example scenarios.

5.2.1 Solution to the $M/D/1$ Submodel

The $M/D/1$ system is a classical queueing model that can be solved in different ways. While the precise solution is obtained by the Markov theory [Kle75, Küh06c], the effective bandwidth and MVA approach provide asymptotic solutions that are easy to be computed. After comparison, it is found that the solution in [RMV96], which can be also derived from the effective bandwidth theory [Kel96], has a good balance between the estimation accuracy and computational complexity. It suggests an exponential approximation of the tail probability:

$$P\{Q_\mathrm{p} > q_\mathrm{p}\} \approx \beta \, \exp(-\gamma \, q_\mathrm{p}). \tag{5.1}$$

Here, RV Q_p is the queue length measured in the number of packets. γ is obtained by solving the equation $\rho(\exp(\gamma)-1)=\gamma$ where ρ is the offered load. The asymptotic constant β is elaborately calculated as [RMV96]:

$$\beta = \frac{1-\rho}{\rho \cdot \exp(\gamma) - 1}. \tag{5.2}$$

5.2.2 Solution to the FBM Submodel

For the LRD traffic modeled by the FBM, the tail probability of Q is derived either by the effective bandwidth theory or by the MVA approach. Both lead to the equivalent solution [Nor95, CS98, BC00]:

$$P\{Q > q\} \approx \beta \, \exp\left(-\frac{c^{2H} \, (1-\rho)^{2H} q^{2-2H}}{2H^{2H}(1-H)^{2-2H} \, a \, c \, \rho}\right). \tag{5.3}$$

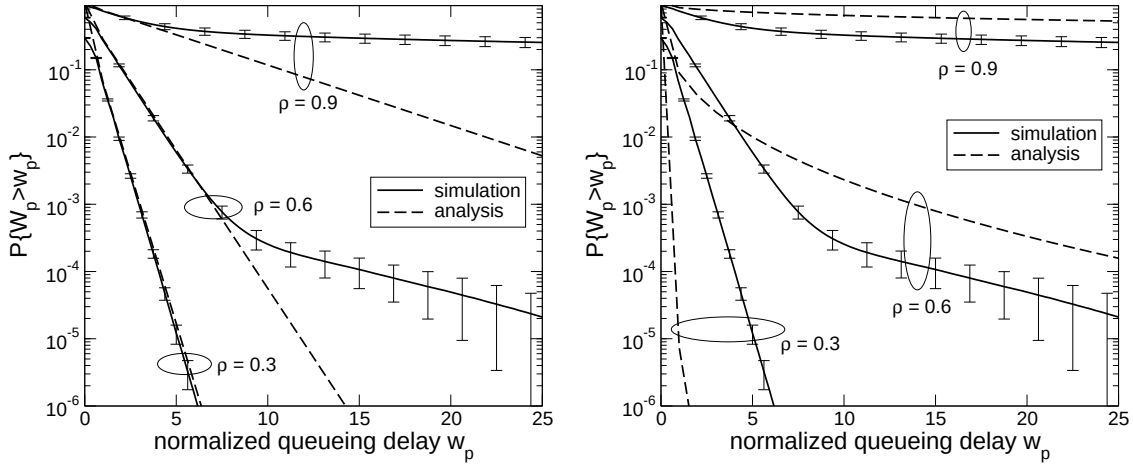
For the application in this thesis, $\beta = 1$ is assumed. Parameter a is further calculated as:

$$a = \frac{c_\mathrm{a}^{2H-1} \, (\frac{2-2H}{3-2H} \, \varphi)^{2-2H}}{(3-2H) \, (2H-1) \, H}. \tag{5.4}$$

Recall that H is the Hurst parameter and c is the transmission rate of the server. c_a denotes the access link rate of individual sources. φ stands for the mean traffic volume of a session (cf. Section 4.2.1).

5.2.3 Evaluation of Example Scenarios

The overall queueing performance of the M/Pareto traffic is obtained by Eq. (5.1) for small queue lengths and by Eq. (5.3) for large queue lengths. For demonstration, the same scenarios to those in Section 5.1 are inspected. The queueing performance is evaluated in terms of the queueing time W normalized by the maximal transmission time per packet. Mathematically



(a) $M/D/1$ submodel (b) FBM submodel

Figure 5.3: Tail probability of normalized queueing time at different loads ($c_\mathrm{a} = 50$ Mbps)

expressed, the normalized queueing time $W_\mathrm{p} = W/(l_\mathrm{max}/c)$. The tail probability of W_p is derived from that of the queue length as [1]:

$$P\{W_\mathrm{p} > w_\mathrm{p}\} = P\{W > w_\mathrm{p}\,\frac{l_\mathrm{max}}{c}\} \approx P\{Q > w_\mathrm{p}\,l_\mathrm{max}\}. \tag{5.5}$$

In Fig. 5.3, the analytical results are compared with simulations for the light, medium and heavy load, respectively. The simulation results demonstrate that the CCDF curve for the small queueing times behaves differently from the part for large queueing times, which results in a *knee point* in the between. Compared with Fig. 5.1, the location of the knee point conforms to that of the critical point in the distribution of relevant time scales. This verifies the crucial influence of the relevant time scale on the queueing performance. Notice that at the light load ($\rho = 0.3$), the behavior of large queueing times is "hidden" in the region of very low probability.

Fig. 5.3(a) shows that the solutions from the $M/D/1$ submodel perform very well in the region of small queueing times. In Fig. 5.3(b), the estimations from the FBM submodel for the large queueing times outline the evolution tendency of the curves, but are relatively conservative. This is due to the rough assumption of $\beta = 1$ in Eq. (5.3). Further improvement in the estimation accuracy is possible, for example, through the application of the Bahadhur Rao asymptotic [MV96], the details of which, however, are not further looked at since the focus of this chapter is on the multi-scale queueing analysis by the piece-wise approximation instead of on the individual submodel.

Corresponding to Fig. 5.2, the influence of the access rate c_a on the queueing performance is illustrated in Fig. 5.4. While the CCDF curves for small queueing times are well approximated by the $M/D/1$ submodel independent of the values of c_a, a large value of c_a worsens the performance significantly in the area of large queueing delay, which is captured by the FBM submodel. When c_a decreases, the knee point shifts to large queueing times and the region in

[1] Theoretically, W here is actually the *virtual queueing time* defined in the sense of a time-average measure, which is not always equivalent to the common *customer-seen* queueing time. In this thesis, it is assumed that the difference is negligible.

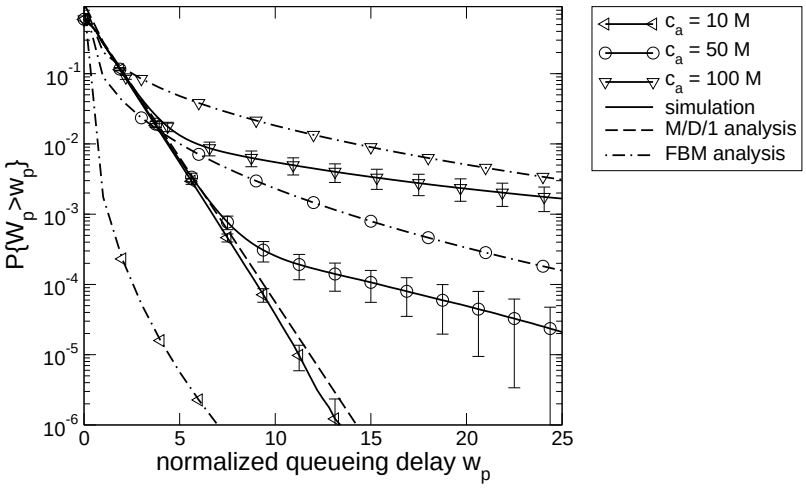


Figure 5.4: Tail probability of normalized queueing time for different access rates ($\rho = 0.6$)

which the $M/D/1$ submodel is valid is expanded. This is a desirable effect that can be exploited in the system dimensioning.

5.3 Summary

In this chapter, the relationship between the multi-scale traffic characteristics and the resulting piece-wise queueing behaviors is inspected by means of the relevant time scale concept on the basis of the effective bandwidth theory as well as the MVA method. This justifies the decomposition of the traffic process into submodels that are valid within their respective ranges of time scales. The queueing performance is thus piece-wise approximated by integrating the solutions of individual submodels.

The novelty of the proposed method lies in the combined application of the time scale decomposition concept and the theory of the relevant time scale from integrated analytical methods, so as to possess the advantages of both sides. On the one hand, the form of the final solution is similar to the closed-form solution of the conventional time scale decomposition method developed in the performance study of ATM networks (cf. Section 4.3.1). There is no need to carry out complex optimization procedures to quantitatively determine the relevant time scales as required by the integrated analysis. On the other hand, the proposed method applies the effective bandwidth or variance process as the generic traffic measure to characterize and identify the traffic properties on multiple time scales. This provides an intuitive general way to determine the time-scale dependent submodels for the time scale decomposition approach, which could otherwise be an intricate task. These advantages make the proposed method very practical for the the system analysis concerning complex traffic patterns on multiple time scales.

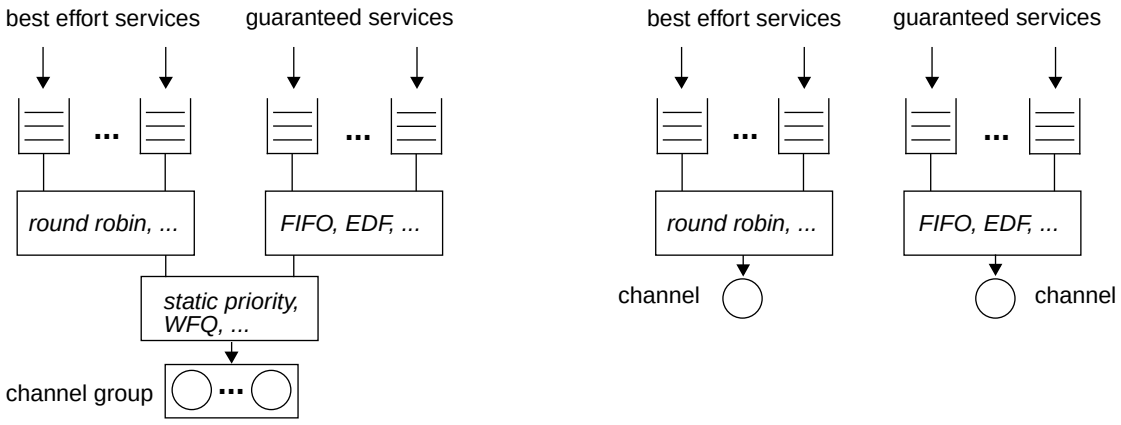
6 Service Guarantee in an Edge Node

In this chapter, the solution to the service guarantee in an OPS/OBS edge node is provided with respect to the QoS requirements outlined in Chapter 3. In Section 6.1, the possible frame scheduling architectures are discussed. The system model to be inspected is derived and the admission control problem is formulated. It is figured out that the mean frame size and the probability distribution of the frame queueing delay are the two most important measures in the QoS provisioning. In Section 6.2, traffic models to be applied in the performance analysis are introduced. The mean frame size is analyzed in Section 6.3 and its relevance to the processing load in SCUs is evaluated. Section 6.4 is devoted to the analysis of the frame queueing delay. The analytical method proposed in Chapter 5 is applied. Simulations are carried out for the verification and more elaborate delay estimation. Based on the results of Section 6.3 and 6.4, a comprehensive analysis is performed in Section 6.5 to study how the system throughput is constrained by the delay budget in the edge node. In Section 6.6, a novel admission control algorithm is proposed. Guaranteed services are supported by meeting with the requirements in all three aspects: the traffic load on the data path of switches, the processing load in SCUs and the delay in the edge node. This chapter is summarized in Section 6.7.

6.1 Overview

6.1.1 Architectures for the Frame Scheduling

Various schemes can be applied for the frame scheduling in an edge node. Since forwarding equivalence classes (FEC) are typically classified according to the service class and egress node address, a hierarchical scheduler [FJ95, Bod04] is therefore a natural choice, an example of which is illustrated in Fig. 6.1(a). Here, the flows of different service types are scheduled according to the static priority allowing the guaranteed services to monopolize the transmission capacity. Alternatively, the disciplines on the basis of the generalized processor sharing (GPS) principle [Sta02, KMK04] can be deployed to assure a minimal bandwidth allocation for each service type to realize the inter-class QoS separation. Weighted fair queueing (WFQ) [Zha95, KMK04], worst-case fair weighted fair queueing (WF2Q) [BZ96] and self-clocked fair queueing (SCFQ) [Gol94] are among this kind of scheduling policies. Within the best effort services, the relative fairness between the flows becomes an issue so that the round robin scheduling as well as its extensions like weighted round robin (WRR) [KSC91] and deficit round robin (DRR) [SV96] is applicable. The flows of guaranteed services are generally subject to the per-flow admission control and traffic policing so that the fairness issue is not so



(a) hierarchical scheduling (b) partitioned channels

Figure 6.1: Frame scheduling and channel sharing in an edge node

crucial as that in the best effort services. The FIFO discipline can serve as a simple and practical solution. If further differentiation with respect to the delay performance is desired, earliest due first (EDF) scheduling [SC99] can be used as the intra-class scheduling of the guaranteed services. However, the realization complexity is relatively high, especially when the number of flows is large.

The hierarchical scheduling scheme assures an efficient use of the bandwidth by sharing a group of wavelength channels among all flows. On the other hand, since each wavelength channel has a very high transmission rate, e.g., up to 40 Gbps with today's mature technologies, the full sharing of multiple channels sets forth the requirement of an ultra high bus bandwidth in the digital design of the electronic edge node. This can be very expensive or even infeasible. On this account, it makes sense to apply the channel partitioning and allocate dedicated channels for each service class, as shown in Fig. 6.1(b). With such a static allocation scheme, the inter-class scheduling is also exempted.

6.1.2 System Model

Both the inter-class scheduling in Fig. 6.1(a) and the channel partitioning scheme in Fig. 6.1(b) assure a good QoS separation of guaranteed services from best effort services. Consequently, the guaranteed services can be modeled independently.

In this thesis, the attention is confined to one individual class of guaranteed services. A system model including multiple FEC flows is depicted in Fig. 6.2. Here, the total n FEC flows are of the same service class and are distinguished from each other only by the address of the destined egress node. Each flow is assembled by the combined time/size-based assembly scheme with the control parameters of the timeout period t_{th} and the size threshold s_{th}. After the assembly, the frames are aggregated to a transmission buffer and sent by the server. The server represents an abstract channel with a transmission rate c corresponding to the allocated bandwidth assured either by the inter-class scheduling in the hierarchical scheduling scheme or by the static

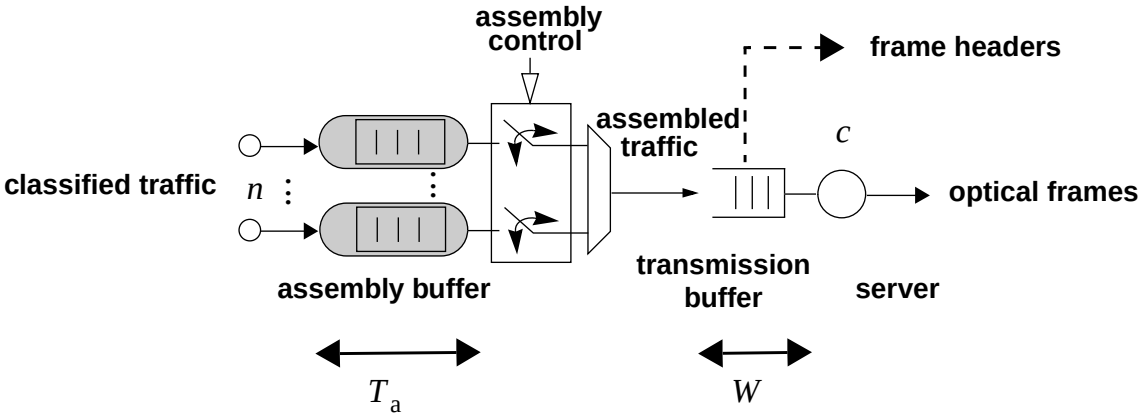


Figure 6.2: System model for QoS analysis

channel allocation in the channel partitioning approach. The frame transmission is scheduled according to the FIFO discipline. In the sections later, it will be shown that FIFO suffices for the scheduling of the assembled traffic. The dashed line in the graph represents the flow of frame headers, which are transmitted either on a separate control channel for out-of-band signaling or on the same channel with data frames in the case of in-band signaling.

Since the frame size has a direct impact on the loss performance in the core network, for the FECs of the same service class it is practical to fix s_{th} to the same value for the sake of fairness. Also, this serves as an appreciative feature to limit the variability in the frame size to improve the overall frame loss performance in core switches. On the other hand, the FECs can have different delay budgets allocated for the edge node because their E2E propagation delay through the core network can differ from each other. Correspondingly, the timeout period can be diverse between the FECs and is denoted by $t_{\text{th},i}$ for individual FEC i. In case of no confusion in the context, t_{th} will also be used to refer to the timeout period of an assembler in general.

As the electronic gateway to the all optical transport network, an edge node should assure a very small traffic loss probability. For the performance study, the data loss is neglected in this thesis. An unbounded transmission buffer is assumed to focus the evaluation on the queue length distribution and the closely related delay distribution. However, it is worth to point out that the obtained results can be further applied to derive the loss probability in case of the limited buffer size [KS98].

Special attention should be paid when an OBS edge node is concerned and an offset time is necessary between each header packet and the corresponding data frame. To this end, the data frame must be delayed on purpose in the transmission buffer. If all the FECs have the same offset time setting, the impact of the offset time on the performance of an edge node is nothing but a constant delay equal to the offset time. In this light, it is not necessary to explicitly model the offset time in a stochastic delay analysis. This thesis pertains to this paradigm. On the contrary, if there are large diversities in the offset times among FECs (e.g., the offset-time-differentiation scheme for QoS differentiation), there is a further degree of freedom in the channel scheduling which is similar to the channel scheduling in core switch nodes. Further details on this issue were reported in [Per06].

6.1.3 Admission Control Problem

In connection with the discussion in Chapter 3, for the guaranteed QoS provisioning the admission control must be realized in the edge node to check whether the service requirements can be satisfied or not. To parametrize the admission control problem defined in Fig. 3.5, each FEC is here regarded as an individual service request without loss of generality. For an arbitrary FEC i, the sustainable data rate $r^*_{\mathrm{dat},i}$ and sustainable header rate $r^*_{\mathrm{sig},i}$ can be specified on the E2E path through the core network. As long as the real data rate $r_{\mathrm{dat},i}$ and header rate $r_{\mathrm{sig},i}$ of this FEC do not exceed the respective specification, the performance requirements on the frame blocking probability and on the timely processing of frame headers can be assured in core switches. In this sense, $r^*_{\mathrm{dat},i}$ and $r^*_{\mathrm{sig},i}$ represent the amount of available resources in the core network for FEC i. Let $r_{\mathrm{req},i}$ denote the requested data rate of FEC i. $r_{\mathrm{req},i}$ is then the upper bound of the real data rate, i.e., $r_{\mathrm{dat},i} \leq r_{\mathrm{req},i}$. Neglecting possible frame overheads (e.g., bits for error correction, etc.) added by the assembly procedure to the data frames, the admission control algorithm should keep track of $r_{\mathrm{req},i}$ such that the following conditions always hold:

1. $r_{\mathrm{req},i} \leq r^*_{\mathrm{dat},i}$
2. the resulting header rate $r_{\mathrm{sig},i}$ under any data rate $r_{\mathrm{dat},i} \leq r_{\mathrm{req},i}$ does not exceed $r^*_{\mathrm{sig},i}$
3. the sum of the assembly delay and queueing delay is bounded by the delay budget δ^*_i.

While Condition 1) is straightforward, Condition 2) assures that the generated header rate $r_{\mathrm{sig},i}$ is always bounded by $r^*_{\mathrm{sig},i}$. $r_{\mathrm{sig},i}$ is derived from $r_{\mathrm{dat},i}$ by the formula:

$$r_{\mathrm{sig},i} = \frac{r_{\mathrm{dat},i}}{E[S_{\mathrm{b}}]} \tag{6.1}$$

where $E[S_{\mathrm{b}}]$ is the mean of the frame size S_{b}. Note that $E[S_{\mathrm{b}}]$ further depends on the data traffic rate $r_{\mathrm{dat},i}$ and assembly parameters, which needs a closer inspection.

The variable delay in the edge node is composed of the assembly delay T_{a} in the assembly buffer and the queueing delay W in the transmission buffer. The delay here is measured by the upper bound of the delay component as described in Chapter 3. Since T_{a} is absolutely limited by the timeout period $t_{\mathrm{th},i}$ according to the assembly mechanism, Condition 3) is reduced to the guarantee of a statistical queueing delay bound that equals to $\delta^*_i - t_{\mathrm{th},i}$. For this purpose, the probability distribution function of W is to be analyzed.

Note that hitherto the assembly parameters are regarded as given parameters that are fixed. In a realistic network scenario, this is generally true for s_{th} because it is closely related to the resource dimensioning (e.g. FDL buffer) as well as the system performance in core switches and can be treated by the edge node as a predefined parameter. $t_{\mathrm{th},i}$, on the other hand, can be adjusted flexibly according to the system status in the operation. The admission control algorithm also includes the task to determine the parameter $t_{\mathrm{th},i}$ for the individual FEC i.

6.2 Traffic Models

Because the M/Pareto model has a good capability in capturing the time-scale dependent characteristics of aggregated traffic, it will be used to model the client traffic in the edge node. The Hurst parameter H is set to 0.8. The mean traffic volume of a session $\varphi = 10$ KBytes and the maximal packet size $l_{\max} = 1000$ bytes.

For comparison, the Poisson process will be applied as a reference model for the packet arrival process of the client traffic. The packet length L is i.i.d. with P{L = 40 bytes} = 0.49, P{L = 576 bytes} = 0.17, P{L = 1500 bytes} = 0.17 and P{L = x bytes} = 0.17/535 for 41≤x≤575 [kcMT98].

6.3 Evaluation of Frame Header Rate

Since the frame header rate is related to the data rate through the mean frame size, an approximate closed-form solution of $E[S_b]$ is derived first in the following subsection. Then, with respect to Condition 2) of the admission control in Section 6.1.3, the evolution of the header rate with an increasing data rate is specially evaluated. In the both studies, only an individual FEC needs to be modeled. Therefore, the FEC index i in the subscript of parameters is omitted in the presentation.

6.3.1 Approximate Analysis of Mean Frame Size

A quantitative estimation of $E[S_b]$ can be performed on the basis of the variance process of the client traffic by approximating the marginal distribution of the traffic arrival with the Gaussian distribution.

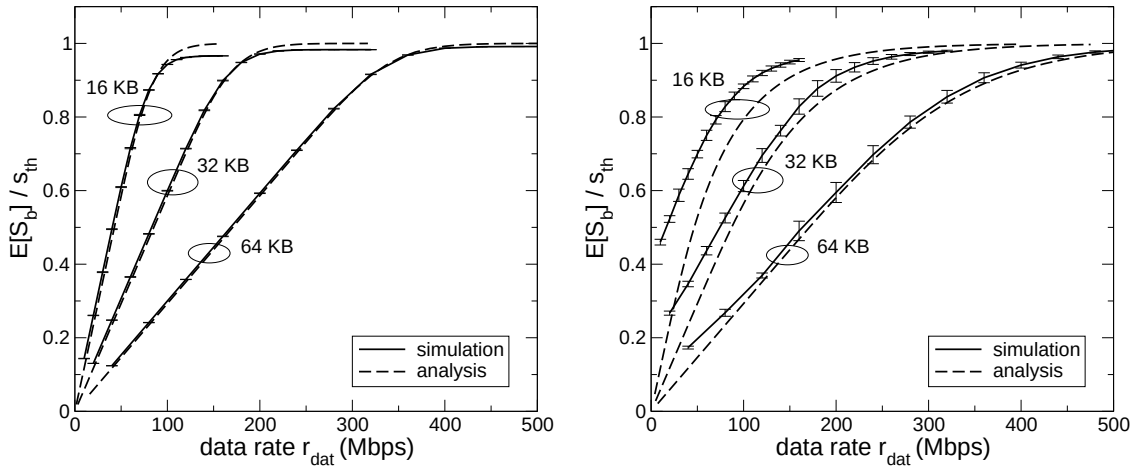
Let U_t denote the amount (in bytes) of client traffic arrival at the observed FEC within an arbitrary time interval of t. The mean traffic rate equals to r_{dat}. The variance process of the traffic is represented by VAR[U_t]. The client traffic is treated as a fluid flow and U_t is a general process with a probability density function (PDF) of $p_t(u)$.

With a pure time-based assembly and the timeout period t_{th}, the frame size S_b is approximately equivalent to $U_{t_{th}}$ with the PDF $p_{t_{th}}(u)$. For the inspected combined time/size-based assembly, the size threshold s_{th} further constrains the frame size. As a result, the PDF of S_b is a truncated function of $p_{t_{th}}(u)$ with an upper bound $u = s_{th}$. The mean frame size is thus calculated as:

$$E[S_b] = \int_0^{s_{th}} u \cdot p_{t_{th}}(u)\mathrm{d}u + s_{th} \cdot P\{U_{t_{th}} > s_{th}\} \tag{6.2}$$

Assume that $p_{t_{th}}(u)$ follows a Gaussian distribution $N(\mu, \sigma^2)$ with $\mu = r_{dat} \cdot t_{th}$ and $\sigma^2 = \mathrm{VAR}[U_{t_{th}}]$. After some derivation, Eq. (6.2) leads to:

$$E[S_b] = \mu\,(1 - \Phi(\frac{\mu - s_{th}}{\sigma})) - \frac{\sigma}{\sqrt{2\pi}}\,\exp(-\frac{(\mu - s_{th})^2}{2\sigma^2}) + s_{th}\,(1 - \Phi(\frac{s_{th} - \mu}{\sigma})) \tag{6.3}$$



(a) Poisson process (b) M/Pareto model (H = 0.8, c_a = 50 Mbps)

Figure 6.3: Comparison between approximate analysis and simulation

where $\Phi(u)$ is the distribution function of the standard Gaussian distribution $N(0,1)$.

To verify the goodness of the approximation, the analytical results are compared with the simulations. In the simulation tool, the size threshold is realized as an absolute upper bound on the frame size. The normalized mean frame size is plotted against the client traffic rate r_{dat} in Fig. 6.3. The incoming client traffic is modeled by the Poisson process and the M/Pareto model, respectively. To determine the parameter $\text{VAR}[U_{t_{\text{th}}}]$, the variance process of the Poisson process is derived by the standard probability theory [AK93, Küh06c] and the variance process of M/Pareto model is obtained from Eq. (4.4). For the assembly parameters, $t_{\text{th}} = 1.5$ ms and s_{th} takes different values of 16 KBytes, 32 KBytes and 64 KBytes.

For the Poisson process, Fig. 6.3(a) shows that the analysis leads to good estimations of the mean frame size in general. A bit deviation appears in the saturation region, typically for the small s_{th} (16 KBytes). This is because the variability of the packet length prevents the frame size fully reaching s_{th} even at high data rates, the effect of which is not captured by the analysis based on the fluid-flow model. For the moderate and large values of s_{th} (32 KBytes, 64 KBytes), however, this effect is negligible.

With the M/Pareto model in Fig. 6.3(b), it can be seen that the analysis results in a certain underestimation. This is attributed to the approximation by the Gaussian distribution that has its domain covering also the negative values. This phenomenon is not obvious in Fig. 6.3(a) because the Poisson process has a relatively small $\text{VAR}[U_{t_{\text{th}}}]$ and the deviation resulting from the negative values fades off. On the contrary, the M/Pareto traffic has the self-similarity property which indicates a much larger variability in the traffic process. This reduces the preciseness of the modeling through the Gaussian distribution. Nevertheless, for the moderate and large values of the size threshold (32 KBytes, 64 KBytes), the analytical results are very close to the simulation results and serve as tight approximations of $E[S_{\text{b}}]$.

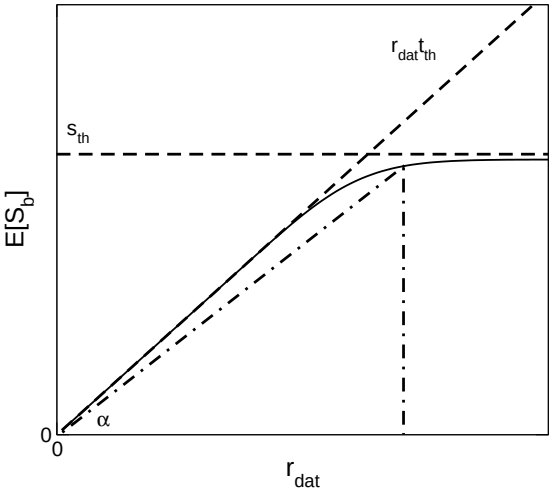


Figure 6.4: Evolution of the mean frame size and header rate

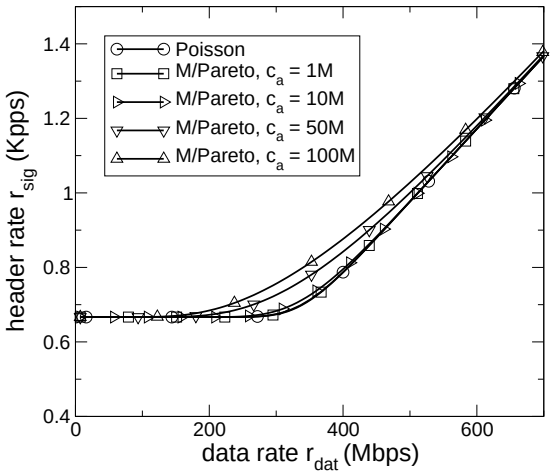


Figure 6.5: Influence of traffic parameters on r_{sig}

6.3.2 Evolution of the Header Rate dependent of the Data Rate

All above results show that the evolution of the mean frame size with the data rate follows a concave curve, which can be interpreted in an intuitive way. At small r_{dat} the frame assembly is mostly triggered by the timeout t_{th}. From Eq. (3.8), $E[S_{\text{b}}]$ under a pure time-based assembly is linearly proportional to r_{dat}. When r_{dat} gets large, the probability that the frame size reaches s_{th} increases. The increase of $E[S_{\text{b}}]$ with r_{dat} is gradually suppressed until at last $E[S_{\text{b}}] \approx s_{\text{th}}$ asymptotically. This evolution process is illustrated in Fig. 6.4.

Connecting an arbitrary point on the curve with the point of origin and denoting the angle to the x-axis with α (cf. Fig. 6.4), from Eq. (6.1) it is easy to see that the header rate r_{sig} equals to $\text{ctan}(\alpha)$. Since the curve is concave, $\text{ctan}(\alpha)$ is non-decreasing with r_{dat}. So is r_{sig}. This means that if the requested data rate is r_{req} and the corresponding header rate under fixed assembly parameters satisfies $r_{\text{sig}} \leq r^*_{\text{sig}}$, then for all $r_{\text{dat}} \leq r_{\text{req}}$ the resulting header rate does not exceed r^*_{sig}. In other words, to assure Condition 2) of the admission control (cf. Section 6.1.3), it suffices to check the condition $r_{\text{sig}} \leq r^*_{\text{sig}}$ only for the requested data rate r_{req}.

In Fig. 6.5, r_{sig} in header packet per second (pps) is plotted with respect to different traffic parameters. The access link rate c_{a} of the M/Pareto model is specially inspected. The results are calculated according to Eq. (6.1) and (6.3) for $t_{\text{th}} = 1.5$ ms and $s_{\text{th}} = 64$ KBytes. It is shown that the large access link rate leads to a fast increase of r_{sig} because the large traffic variability postpones the convergence of $E[S_{\text{b}}]$ to s_{th}. By analyzing the critical time scale (cf. Section 4.2.2) of the M/Pareto model, it can be further figured out that r_{sig} increases as slowly as the Poissonian traffic when the critical time scale is larger than (e.g., 8 ms for $c_{\text{a}} = 1$ Mbps) or close to (0.8 ms for $c_{\text{a}} = 10$ Mbps) the time threshold t_{th}. The cases with relatively small critical time scale (0.16 ms for $c_{\text{a}} = 50$ Mbps and 0.08 ms for $c_{\text{a}} = 100$ Mbps) have obvious faster increases in the header rate. Therefore, it is beneficial to set t_{th} at least equal to the critical time scale of the client traffic as long as the delay budget allows for it.

6.4 Queueing Delay Analysis

The queueing performance in the transmission buffer is much influenced by the assembly parameters of each FEC. Since the assembly time threshold is typically different between the FECs, a general analysis taking into consideration of all these parameters would be difficult and not scalable with respect to the number of the FEC flows aggregated.

In this section, the impact of the time threshold and size threshold is studied first. It shows that the worst queueing performance occurs when the assembly is dominated by the size threshold so that the frame size is maximized. This corresponds to the maximal degree of assembly. Under this condition, the variance process of the frame departure flow from an individual assembly buffer is derived, which highlights the impact of the assembly procedure on traffic characteristics over multiple time scales. The queueing performance for the frame departure flow assembled by the pure size-based scheme is analyzed by the multi-scale analytical method introduced in Chapter 5.

6.4.1 Worst Case Assembly

As mentioned in Section 3.2.2.1.3, significant influence of the assembly procedure on traffic characteristics lies in two aspects. On the one hand, the variability of the assembled frame inter-departure time is smaller than that of the packet interarrival time in the client traffic. On the other hand, the frame size is much larger than the packet size. In particular, the large frame size means large instantaneous peak of workload in the arrival process, which is detrimental for the queueing performance. In [BPR01], it is heuristically shown that the transform of a packet arrival process by aggregating packets into large frames results in a more variable traffic process. This indicates that the assembly can lead to a worse performance in the transmission buffer.

The above intuitive conception is verified by extensive simulation studies for both Poisson and M/Pareto traffic model. Representative results are shown in Fig. 6.6 in terms of the CCDF of the queueing delay W in the transmission buffer with respect to different values of the timeout period. Here, the system load $\rho = 0.9$, the FEC number $n = 20$ and the size threshold $s_{\text{th}} =$ 64 KBytes. The total offered traffic is equally distributed to the 20 FECs and all FECs have the same timeout period t_{th}. For the M/Pareto model, the access link rate $c_{\text{a}} = 50$ Mbps.

It can be seen that with both traffic models the queueing delay continuously increases with the increasing t_{th} and finally converges to be constant irrespective of the further changes of t_{th}. This is due to the fact that the increase of t_{th} leads to larger sizes of the assembled frames. However, when t_{th} becomes large enough, the parameter s_{th} dominates the assembly control and the frame size reaches its maximum. Further changes of t_{th} have no more influence. So, the worst queueing performance appears when the size threshold dominates the assembly control. In this case, the assembly behaves like a pure size-based scheme.

It is further noticeable that the queueing delay only gets worse in the range of small w with the increasing t_{th}. For large w, t_{th} causes little difference. This is because the assembly procedure

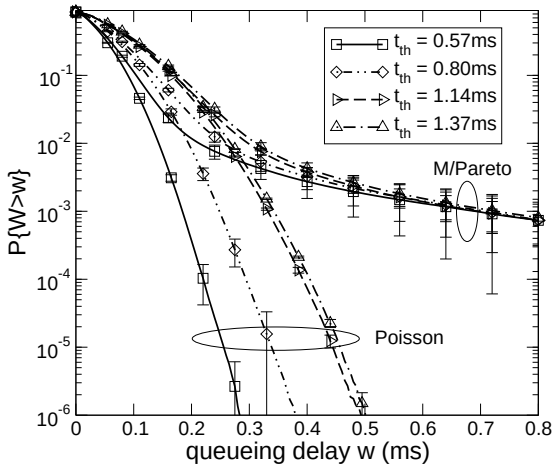


Figure 6.6: Impact of timeout period on the queueing delay in the transmission buffer

has little impact on the large-time-scale characteristics of the passing traffic, which will become clearer in the following sections.

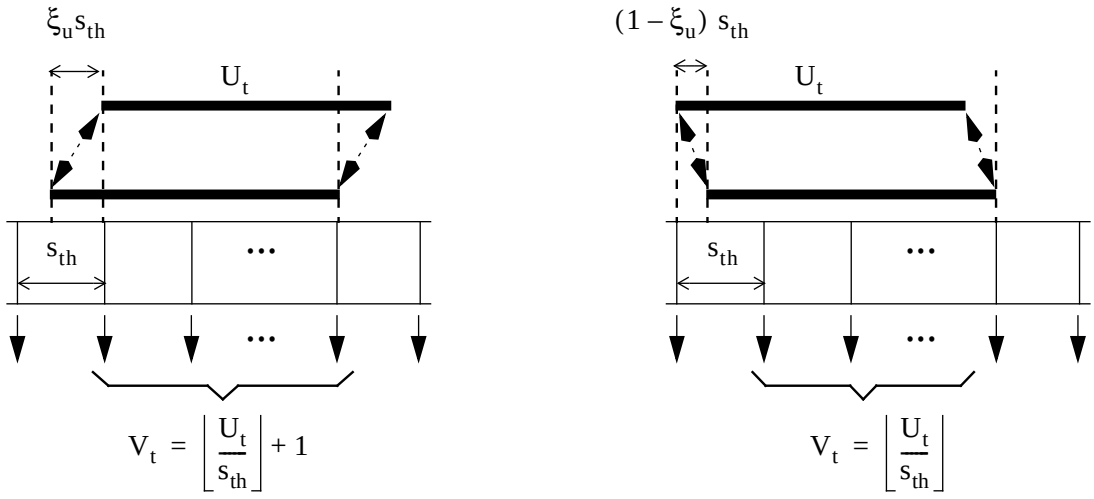
With the pure size-based scheme as the worst case, the assembler is quite similar to a packetizer. However, previous studies on the performance impact of packetization were typically for periodic bit flows [BhH89, KMK04] or in the paradigm of the deterministic queueing analysis [BT01]. Those results are not applicable for the assembly scenario in the OPS/OBS edge node.

The following sections will concentrate on this worst assembly case and the influence of the parameter t_{th} will be neglected.

6.4.2 Characterization of Assembled Traffic

In this section, the variance process of the output traffic from an individual pure size-based burst assembly is analyzed. The same notations as defined in Section 6.3.1 are applied: U_t denotes the amount of the fluid-flow client traffic within an interval t and $p_t(u)$ is the PDF of U_t. The mean value $\mathrm{E}[U_t] = r_{\text{dat}} \cdot t$. V_t stands for the number of frame departures from the assembly buffer within the interval t. The frame inter-departure time is denoted by D.

With the pure size-based assembly, the frame size can be approximated with a constant size equal to s_{th}. Therefore, $\mathrm{E}[V_t] = \mathrm{E}[U_t]/s_{\text{th}}$. In contrast to the mean value, a sample value of the RV V_t must always be integer. It is either equal to $\lfloor U_t/s_{\text{th}} \rfloor + 1$ or $\lfloor U_t/s_{\text{th}} \rfloor$, depending on how the arbitrarily selected segment of U_t spans over the multiple assembled frames, as illustrated in Fig. 6.7. In the picture, the linear horizontal axis represents the data volume and the series of downward arrows stand for the departures of frames with the size s_{th}. Let ξ_u denote the residual of the division U_t/s_{th}, i.e., $\xi_u = U_t/s_{\text{th}} - \lfloor U_t/s_{\text{th}} \rfloor$. Fig. 6.7(a) shows that when the portion of the first frame in V_t covered by U_t does not exceed ξ_u, it turns out that $V_t = \lfloor U_t/s_{\text{th}} \rfloor + 1$. Otherwise, as depicted in Fig. 6.7(b), $V_t = \lfloor U_t/s_{\text{th}} \rfloor$. As the RV ξ_u is actually uniformly distributed between



(a) cases for $V_t = \lfloor U_t/s_{\text{th}} \rfloor + 1$ (b) cases for $V_t = \lfloor U_t/s_{\text{th}} \rfloor$

Figure 6.7: Relation between U_t and V_t

0 and 1, it can be derived [BPR01]:

$$V_t = \begin{cases} \lfloor U_t/s_{\text{th}} \rfloor + 1 & \text{with probability } \xi_u, \\ \lfloor U_t/s_{\text{th}} \rfloor & \text{with probability } 1-\xi_u. \end{cases} \tag{6.4}$$

And the variance process VAR$[V_t]$ is calculated as [BPR01]:

$$\begin{aligned} \text{E}[V_t^2] - \text{E}[V_t]^2 &= \int_0^\infty \text{E}[V_t^2 | U_t = u] p_t(u)\,\text{d}u - \text{E}[V_t]^2 \\ &= \int_0^\infty (\xi_u (\lfloor \frac{u}{s_{\text{th}}} \rfloor + 1)^2 + (1-\xi_u)(\lfloor \frac{u}{s_{\text{th}}} \rfloor)^2) \cdot p_t(u)\,\text{d}u - (\frac{\text{E}[U_t]}{s_{\text{th}}})^2 \end{aligned} \tag{6.5}$$

Notice that Eq. (6.5) can be rewritten as:

$$\begin{aligned} \text{VAR}[V_t] &= \int_0^\infty ((\lfloor \frac{u}{s_{\text{th}}} \rfloor + \xi_u)^2 + (\xi_u - \xi_u^2)) p_t(u)\,\text{d}u - (\frac{\text{E}[U_t]}{s_{\text{th}}})^2 \\ &= \int_0^\infty ((\frac{u}{s_{\text{th}}})^2 + (\xi_u - \xi_u^2)) \cdot p_t(u)\,\text{d}u - (\frac{\text{E}[U_t]}{s_{\text{th}}})^2 \\ &= \int_0^\infty (\xi_u - \xi_u^2) \cdot p_t(u)\,\text{d}u + \frac{\text{VAR}[U_t]}{s_{\text{th}}^2} \end{aligned} \tag{6.6}$$

A closed-form solution for Eq. (6.6) is hard to obtain due to the floor function involved in the computation of ξ_u. However, it clearly illuminates the relationship between VAR$[V_t]$ and VAR$[U_t]$. Since $0 \leq \xi_u < 1$, the integral term in Eq. (6.6) is always positive. That is to say, the traffic variability in terms of the variance process is increased after the assembly process, indicating that a worse queueing performance can be resulted in the subsequent transmission buffer.

In the following subsections, the approximate solution to VAR$[V_t]$ is to be derived for small and large t, respectively. It will be shown that on small time scales the variance structure VAR$[V_t]$

resembles that of a constant bit rate (CBR) flow with the constant interarrival time equal to E[D]. On large time scales, the assembly procedure makes no great change in the variance process other than a constant increment.

6.4.2.1 Small-Time-Scale Approximation

On the small time scale $t \to 0$, the probability for large values of U_t becomes very small such that $P\{U_t \leq s_{\text{th}}\} \approx 1$. Because $\xi_u = U_t / s_{\text{th}}$ for $0 \leq U_t < s_{\text{th}}$, it is obtained from Eq. (6.6):

$$\begin{aligned} \text{VAR}[V_t] &\approx \frac{E[U_t]}{s_{\text{th}}} - \frac{E[U_t^2]}{s_{\text{th}}^2} + \frac{\text{VAR}[U_t]}{s_{\text{th}}^2} \\ &= \frac{E[U_t]}{s_{\text{th}}} - \frac{E[U_t]^2}{s_{\text{th}}^2} \\ &= \frac{r_{\text{dat}}\, t}{s_{\text{th}}} - (\frac{r_{\text{dat}}\, t}{s_{\text{th}}})^2 \end{aligned} \tag{6.7}$$

Taking the fact that the mean frame inter-departure time $E[D] = s_{\text{th}}/r_{\text{dat}}$ and defining $\eta_t = t/\text{E}[D] - \lfloor t/\text{E}[D] \rfloor$. Eq. (6.7) leads to the small-time-scale approximation:

$$\text{VAR}[V_t] \approx \eta_t - \eta_t^2 \tag{6.8}$$

where $0 \leq \eta_t < 1$. It can be shown based on the results in [NRSV91] that Eq. (6.8) is actually the variance process of a CBR flow with a constant interarrival time equal to E[D]. This indicates that on small time scales the variance structure of the assembled traffic resembles the pattern of a CBR flow.

6.4.2.2 Large-Time-Scale Approximation

Since $0 \leq \xi_u < 1$ is a periodic function of u with the period equal to s_{th}, the integration operation in Eq. (6.6) can be represented by a sum of piecewise integration:

$$\text{VAR}[V_t] - \frac{\text{VAR}[U_t]}{s_{\text{th}}^2} = \sum_{k=0}^{\infty} \int_0^{s_{\text{th}}} (\frac{\zeta}{s_{\text{th}}} - (\frac{\zeta}{s_{\text{th}}})^2) p_t(k\, s_{\text{th}} + \zeta)\, \text{d}\zeta \tag{6.9}$$

On large time scales, both E[U_t] and VAR[U_t] are large and the span of PDF $p_t(u)$ is much larger than s_{th}. Therefore, $p_t(k\, s_{\text{th}} + \zeta) \approx p_t(k\, s_{\text{th}})$ for $0 < \zeta < s_{\text{th}}$. Eq. (6.9) is approximated as:

$$\begin{aligned} \text{VAR}[V_t] - \frac{\text{VAR}[U_t]}{s_{\text{th}}^2} &\approx \sum_{k=0}^{\infty} p_t(k\, s_{\text{th}}) \cdot \int_0^{s_{\text{th}}} (\frac{\zeta}{s_{\text{th}}} - (\frac{\zeta}{s_{\text{th}}})^2)\, \text{d}\zeta \\ &= \sum_{k=0}^{\infty} p_t(k\, s_{\text{th}}) \cdot \frac{s_{\text{th}}}{6} \approx \frac{1}{6} \end{aligned} \tag{6.10}$$

This leads to the large-time-scale approximation:

$$\text{VAR}[V_t] \approx \frac{1}{6} + \frac{\text{VAR}[U_t]}{s_{\text{th}}^2} \tag{6.11}$$

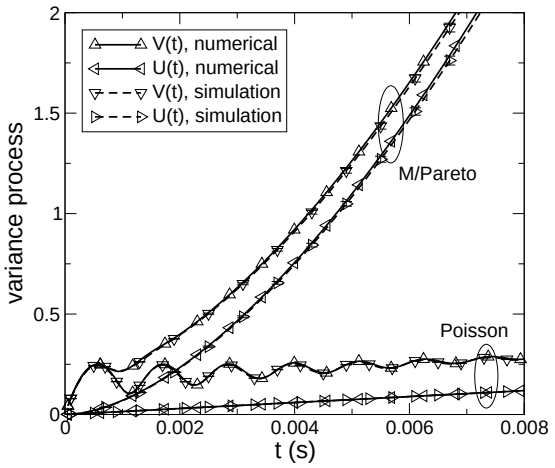


Figure 6.8: Variance process: numerical solution vs. simulation

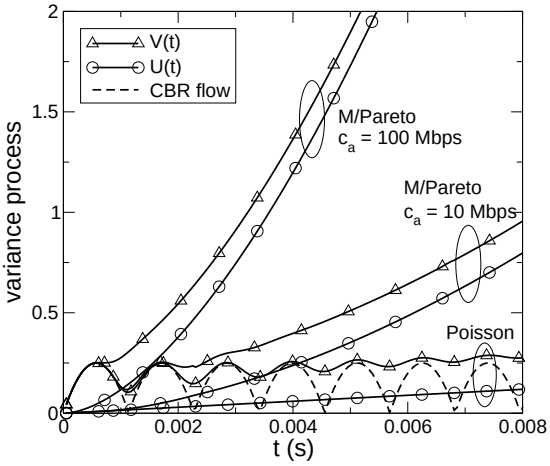


Figure 6.9: Numerical solution to the variance process of different input traffic

6.4.2.3 Simulation and Numerical Solution

To have a complete view of the variance structure and verify the small-/large-time-scale approximation of VAR$[V_t]$, the variance process is studied by solving Eq. (6.6) numerically and also by simulations. For the numerical computation, the PDF $p_t(u)$ is assumed to be a Gaussian distribution and is specified by matching its mean and variance to those of the Poisson process and the M/Pareto model.

In Fig. 6.8, the numerical solution of VAR$[V_t]$ is compared with simulation results. The normalized variance processes VAR$[U_t]/s_{\text{th}}^2$ of the respective packet traffic are also plotted. For both Poisson process and M/Pareto model, the mean traffic rate is 450 Mbps. In the M/Pareto model, the access link rate c_{a} = 50 Mbps. The size threshold is 64 KBytes. It can be seen that the numerical solutions and the simulation results conform with each other so well that they cannot be distinguished in the figure. This justifies the application of the Gaussian Distribution for $p_t(u)$ and the computation precision.

In Fig. 6.9, only the numerical solutions are inspected with respect to different traffic parameters. For comparison, the variance process of the corresponding CBR flow from Eq. (6.8) is also shown. The setting of the traffic parameter is similar to the case of Fig. 6.8 except that the influence of the access link rate of the M/Pareto model is specially studied. For this purpose, c_{a} is set to 10 Mbps and 100 Mbps, respectively.

It can be observed that for both traffic models the variance structure of the assembled traffic looks similar to the normalized variance process of the client traffic incremented by a positive factor. This factor fluctuates periodically on small time scales and converges to be constant with the increasing time scale. This phenomenon is exactly explained by the form of the Eq. (6.6). On small time scales, VAR$[V_t]$ is tightly consistent with the variance process of the CBR flow, which verifies the small-time-scale approximation. Especially, the decreasing parts of the variance structure in the assembled traffic and the CBR flow are an indication of the existence of negative correlation, which is caused by the floor function. On large time scales, the variance processes of the assembled traffic and client traffic tend to be parallel. A closer inspection

discovers that the difference between them is about 0.16. This is consistent with the large-time-scale approximation in Eq. (6.11).

Noticeable difference between the two traffic models lies in the dramatic increase of VAR[V_t] with t in the case of the M/Pareto model due to the LRD property. As a result, VAR[V_t] converges faster to the large-time-scale approximation. On very large time scales, the constant item of $1/6$ in Eq. (6.11) becomes negligible, meaning that the assembly procedure has little influence on the large-time-scale traffic characteristics. This coincides with the previous work [HDG03] showing that LRD is immune to the traffic assembly. Also, it explains the overlap in the tails of the delay CCDF curves in Fig. 6.6, i.e., the traffic behaviors on the respective *relevant time scales* are insusceptible to the different settings of the timeout period t_{th}.

The access link rate c_{a} of M/Pareto model has a significant impact on the traffic characteristic. A large value of c_{a} (e.g. 100 Mbps) makes the traffic variance grow very fast with the time scale and the CBR-similar behavior in V_t diminishes quickly. On the other hand, a small access link rate (e.g. 10 Mbps) results in a relatively mild evolution of V_t and the negative correlation structure on small time scales is more apparent.

6.4.3 CCDF of the Delay

According to the theory of the relevant time scale introduced in Chapter 4 and 5, the small-/large-queue performance of a delay system is dominated by the traffic characteristics on small/large time scales, respectively. Corresponding to the different traffic characteristics of the assembled traffic on small and large time scales, the CCDF of the queue length in the transmission buffer is obtained by the approximation for small queue and large queue, respectively.

For the derivation of the delay CCDF, it is assumed that the client traffic flows distributed to the n FECs are independent of each other. The focus of this chapter is placed on the cases with homogeneous traffic, which means that all FEC flows have the same setting in the traffic parameters (e.g., the traffic intensity) as well as in the size threshold s_{th}. In Appendix A, the scenarios with FEC flows of different traffic intensities are studied and compared with the homogeneous cases. It shows that with a fixed system load and a fixed number of FECs, the worst tail behavior in the delay CCDF is resulted by the homogeneous traffic.

The queue length is measured either in the number of buffered bytes Q or in the number of backlogged frames Q_{b}. As introduced in Chapter 5, the queueing delay W is further calculated by $P\{W > w\} \approx P\{Q > c \cdot w\} = P\{Q_{\text{b}} > c \cdot w/s_{\text{th}}\}$. In the illustration, the normalized queueing time defined as $W_{\text{b}} = W/(s_{\text{th}}/c)$ is also adopted.

6.4.3.1 Small-Queue Approximation

As shown in Section 6.4.2.1, on small time scales, the variance process of an individual frame departure flow is analogous to a CBR flow, the arrival period of which is equal to the mean frame inter-departure time E[D]. As a result, the superposition of the assembled traffic behaves like the multiplexing of n CBR flows on the small time scales. According to the multi-scale queueing analysis in Chapter 5, the standard queueing model for the superposition of periodic

traffic can be applied for small queue lengths as long as the relevant time scale is located in the range of small time scales.

Since the n FEC flows are homogeneous, the frame flows have the same E[D]. So, the tail probability of the backlogged frames is estimated according to the $nD/D/1$ model [NRSV91]:

$$P\{Q_b > q_b\} \approx \sum_{q_b<k\leq n} \binom{n}{k} (\frac{k-q_b}{\phi})^k (1-\frac{k-q_b}{\phi})^{n-k} (\frac{\phi-n+q_b}{\phi-k+q_b}). \tag{6.12}$$

Here, $q_b \in \{0,1,...,n-1\}$ and ρ denotes the total system load. ϕ is the ratio of E[D] and the frame transmission duration, i.e., $\phi = E[D]/(s_{th}/c)$. It can be derived that $\phi = n/\rho$.

Note that the small-queue performance derived here does not depend on specific models of the client traffic, because the small-time-scale characteristic of the assembled traffic obtained in Section 6.4.2.1 is a quite general result.

6.4.3.2 Large-Queue Approximation

Eq. (6.11) illuminates that on large time scales the assembly does not change the form of the variance process, but just results in a constant increase in the variance. With the independent and homogeneous traffic for the n FECs, the variance process of the aggregated traffic X_t (in bytes) at the input of the transmission buffer (cf. Fig. 6.2) is:

$$\mathrm{VAR}[X_t] = n\, s_{th}^2\, \mathrm{VAR}[V_t] \approx \frac{n\, s_{th}^2}{6} + n\, \mathrm{VAR}[U_t]. \tag{6.13}$$

To explore the influence of the assembly on the large-queue performance, the MVA upper bound in Eq.(4.15) is applied. Insertion of Eq. (6.13) into Eq. (4.12) leads to:

$$P\{Q > q\} \approx \beta\, \exp(-\frac{g(q,\tau_q)^2}{2})$$

where $g(q,\tau_q)^2$ is computed as:

$$\begin{aligned} g(q,\tau_q)^2 &= \inf_{t\geq 0} \frac{(q+(c-r)\cdot t)^2}{\mathrm{VAR}[X_t]} \\ &= 1/\sup_{t\geq 0} \frac{\mathrm{VAR}[X_t]}{(q+(c-r)\cdot t)^2} \\ &\approx 1/\sup_{t\geq 0} \frac{n\, s_{th}^2/6 + n\, \mathrm{VAR}[U_t]}{(q+(c-r)\cdot t)^2}. \end{aligned} \tag{6.14}$$

Here, $r = n \cdot r_{dat}$. The first term $n \cdot s_{th}^2/6$ in the numerator of the sup-operation is constant and reflects the affection of the increased variance through the assembly. The second term $n \cdot \mathrm{VAR}[U_t]$ is exactly the influence of the original client traffic on the queueing performance in case that the traffic would not experience the assembly. Because traffic processes generally have the monotonically increasing VAR[U_t] with t, the second term becomes dominate for $t \to \infty$. This happens when large q is concerned, in which case the relevant time scale τ_q is located in

the range of large time scales. That is to say, the exponent part in the MVA upper bound of the queue distribution is decided by the original characteristics of the incoming client traffic U_t. The impact of the assembly can be ignored on this point.

On the basis of this insight, a short view is given regarding the specific large-queue approximations for the Poissonian traffic arrival and M/Pareto model, respectively.

Poisson Process

For the Poisson arrival process, the variance process is a linear function of time scale t. After multiplexing, the variance process still has the form of $v \cdot t$ with v:

$$v = \sum_{i=1}^{n} \lambda_i \, \mathrm{VAR}[L] + \sum_{i=1}^{n} \lambda_i \, \mathrm{E}[L]^2. \tag{6.15}$$

L denotes the packet length of client traffic. λ_i refers to the packet arrival rate at FEC i and $\sum_{i=1}^{n} \lambda_i = \rho \; c/\mathrm{E}[L]$. Insert the linear function $v \cdot t$ into Eq. (4.15) and apply the MVA analysis, it is derived that:

$$P\{Q > q\} \approx \beta \; \exp\left(-\frac{2 \, c \, (1-\rho) \, q}{v}\right). \tag{6.16}$$

The accurate determination of the asymptotic constant β is known to be a complex problem [CLW96, CS98]. Assembled from the client traffic modeled by the Poisson process, the frame inter-departure time of each FEC flow is i.i.d. and is less variable than that of the negative exponential distribution (cf. Chapter 3). According to [CLW96], the asymptotic tail probability of the queue injected by the superposition of such flows shall have $\beta > 1$. Especially, β depends on the number of the aggregated flows n and the total system load ρ. This thesis does not go into the technical details in [CLW96] to analyze β. Instead, simulations will be carried out to supplement the study on this issue.

M/Pareto Model

Since the M/Pareto process converges to the FBM process on large time scales, which is not influenced by the traffic assembly, the solution in Eq. (5.3) in Section 5.2.2 is to be applied for the large-queue approximation. For an analytical determination of the asymptotic constant β with the LRD traffic, even less knowledge is available in comparison with the above Poissonian case in which the frame departures still follow a renewal process. So, it will be studied by means of simulations.

6.4.4 Evaluation of System Scenarios

In this section, systematic scenario studies by means of simulations are presented to verify and supplement the analysis in the preceding section. At the same time, it also provides an illustrative overview of the queueing performance in the OPS/OBS edge node with respect to different system parameters. In the evaluation, homogeneous FEC flows are considered. The client traffic is synthesized with the Poisson process and M/Pareto model, respectively.

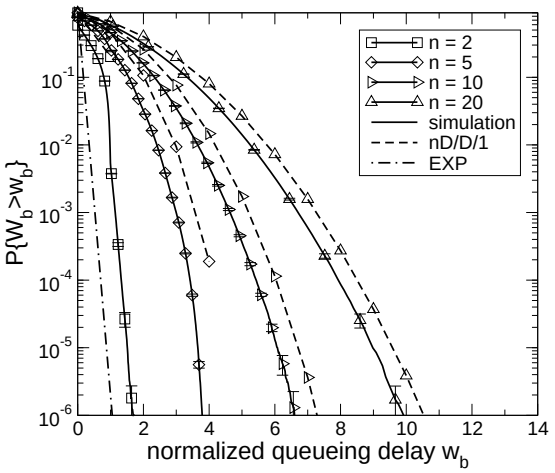


Figure 6.10: Normalized queueing delay wrt. n for the Poisson process

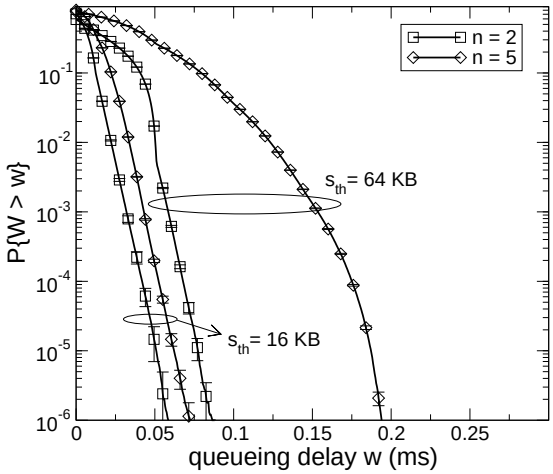


Figure 6.11: Influence of the frame size on the queueing delay for the Poisson process

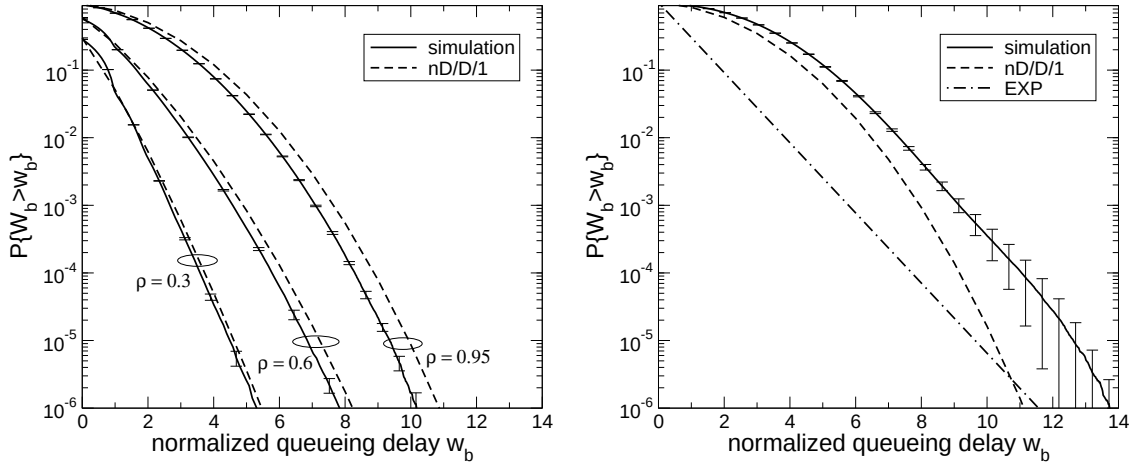
6.4.4.1 Poisson Process

With the Poisson process as the traffic model, the impact of the number of FECs n as well as the size threshold s_{th} is first looked at. By comparing the results from the simulation and the analysis, the accuracy of the analytical solution is evaluated. Especially, the parameter β in the large-queue approximation is to be determined experimentally.

Fig. 6.10 illuminates the evolution of the queueing delay distribution in dependence on the number of FEC flows n. The total offered load is fixed to 0.9. Pure size-based assembly is used with $s_{\text{th}} = 64$ KBytes. The queueing delay is normalized by the frame transmission duration and is denoted by W_{b}. The small queue approximations with $nD/D/1$ model are limited in the range of $0 \leq W_{\text{b}} < n$ since n is the maximal queue length in a $nD/D/1$ system. The large queue approximation with Eq. (6.16) for the Poisson process is denoted by EXP in the graph. The asymptotic constant β is set to 1 for the illustration.

It shows that the delay performance gets worse with larger n because the higher degree of superposition increases the variability of the frame arrival process on small time scales, which is similar to the cell level congestion in an ATM switch [RMV96]. In comparison with the simulation results, the $nD/D/1$ serves as a tight approximation in the region of small queues. Especially, the concave form of the CCDF curves indicates the negative correlation in the frame arrival process on the relevant time scales, which is well captured by the $nD/D/1$ model.

The large-queue behavior is distinguishable in the case of $n = 2$. The CCDF curve tends to be linear in the logarithmic coordinate, and becomes parallel to the analytical result from the large-queue approximation. This verifies the solution of the exponent part in Eq. (6.16). The gap between the simulation result and the analytical result comes from the presumption of $\beta = 1$, which underestimates the tail probability of large queue. For $n = 5, 10, 20$, the small-queue behavior becomes more apparent. The large-queue behavior is correspondingly shifted to the unobservable region of very small occurrence probability (below 10^{-6}), which is very difficult to be captured by common simulation technologies. However, the shapes of the curves indicate that the actual value of β for the large-queue behavior increases with growing n.



(a) $\rho = 0.3, 0.6, 0.95$ (b) $\rho = 0.99$

Figure 6.12: Influence of the system load on the queueing delay for the Poisson process

The normalized queueing delay under the setting of smaller frame size is quite similar to Fig. 6.10. To provide more insights into the impact of the frame size, in Fig. 6.11 the CCDF of the absolute queueing delay W is plotted for $s_{\text{th}} = 16$, 64 KBytes and $n = 2$, 5 respectively. The load remains 0.9. Only the simulation results are shown for a better readability of the graph. It is seen that the small-queue behavior becomes larger with the larger frame size, which leads to an obvious singular point between the small- and large-queue region. The small frame size, on the other hand, triggers an earlier emergence of the large-queue behavior. Irrespective of the setting of s_{th} and n, the CCDF curves for the large values of the delay are parallel to each other, as long as the large-queue behavior begins to show up there. This is consistent with the solution in Eq. (6.16) as it is noticeable that the exponent is independent of s_{th} and n. Finally, comparing the linear large-queue segments for $s_{\text{th}} = 16$ KBytes and 64 KBytes (typically $n = 2$), it is clear that the large frame size leads to a large β.

In Fig. 6.12, the analytical results are further compared with the simulations with respect to different system loads ρ. Since the degree of the underestimation of the large-queue CCDF by the presumption $\beta = 1$ increases with n and s_{th} according to the preceding discussion, relatively large n and s_{th} are chosen here to evaluate the accuracy of the analysis, i.e., $n = 20$ and $s_{\text{th}} = 64$ KBytes. In Fig. 6.12(a), it is seen that for the system load up to $\rho = 0.95$, the small-queue behavior dominates the queue distribution in the illustrated range of the probability and the small-queue approximation by the $nD/D/1$ model performs very well. The large-queue behavior occurs with very small probability and is not observable in the graph. On the other side, Fig. 6.12(b) shows the case of an extremely high load $\rho = 0.99$. Here, the $nD/D1$ approximation can not sufficiently keep up with the real tail probability due to the very high load. This is intuitive to be understood: an $nD/D/1$ system has the queue length always bounded by n even for $\rho = 1$, which is not true for the inspected edge node model. This limitation becomes critical in the heavy load situation. By contrast, the large-queue behavior begins to exhibit itself much earlier, the shape of which is well estimated by the large-queue analysis with however relatively large underestimation due to the fixed setting of $\beta = 1$ in the approximation.

Summarizing, the small-queue approximation by the $nD/D/1$ model performs generally well except for the extremely high load situation. The large-queue approximation with the asymp-

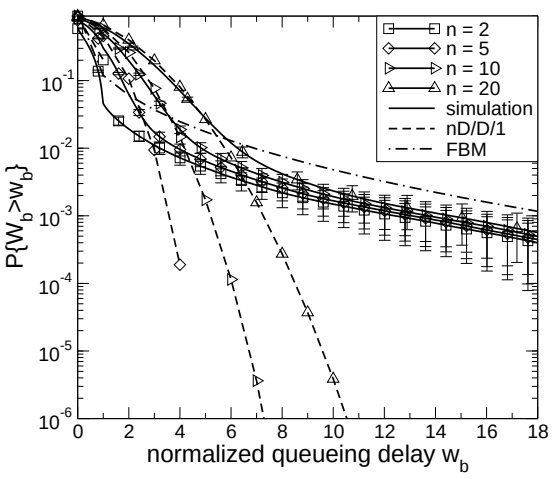


Figure 6.13: Normalized queueing delay wrt. n for the M/Pareto model

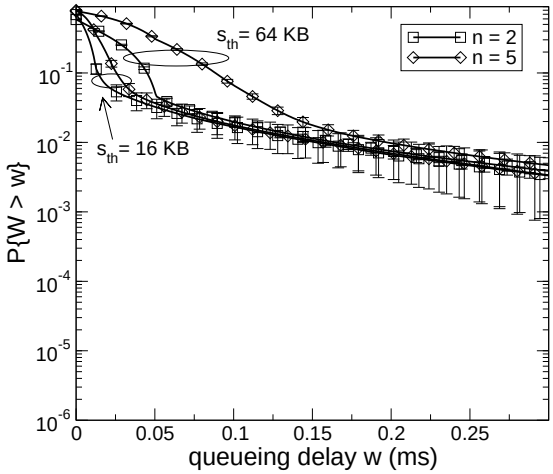


Figure 6.14: Influence of the frame size on the queueing delay for the M/Pareto model

totic constant $\beta = 1$ causes a certain underestimation. Nevertheless, the underestimation is small when n and s_{th} are not large. On the other hand, for large n and s_{th}, as long as the system load is not very high, the large-queue behavior is suppressed by the small-queue behavior in the performance scope of most practical interests. For example, to evaluate the delay jitter defined as the $1-10^{-3}$ quantile of the delay distribution [Y.1541], the small-queue approximation suffices in most cases, as illustrated in Fig. 6.12(a) and also in [Hu06]. At the extremely high load, which shall be seldom in normal practical systems, the large-queue approximation plays a more important role than the $nD/D/1$ model. Despite of the underestimation, the large-queue model well predicts the evolution tendency of the tail probability when the system load approaches the critical point (i.e., $\rho = 1$). This generally suffices in the qualitative evaluation of the heavy load performance.

6.4.4.2 M/Pareto Model

For the evaluation with the M/Pareto model, the same thread as that of the previous subsection is to be followed. Besides the system parameters of n, s_{th} and ρ, the access link rate c_{a} of the M/Pareto model has also a significant influence on the system performance and is specially inspected.

In Fig. 6.13, the CCDFs of the queueing delay are plotted with respect to different values of n. Here, $\rho = 0.9$ and $s_{\text{th}} = 64$ KBytes. The access link rate c_{a} of the M/Pareto model is set to 50 Mbps. The solution of the $nD/D/1$ model in Eq. (6.12) is used as the small-queue approximation and the solution of the FBM model in Eq. (5.3) with $\beta = 1$ is for the large-queue approximation.

The different behaviors in the two ranges of the queueing delay are quite obvious. In the range of the small queue, the curves differentiate from each other and follow a concave from, which is precisely estimated by the $nD/D/1$ model. In the large-queue range, the CCDF curves turn to be heavy-tailed and overlap with each other due to the LRD traffic property on large time scales.

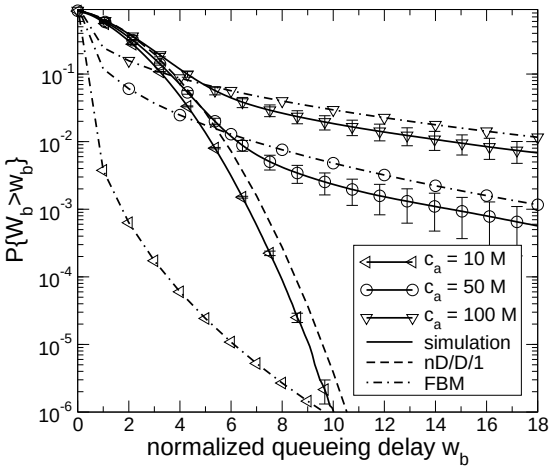


Figure 6.15: Normalized queueing delay wrt. c_a for the M/Pareto model

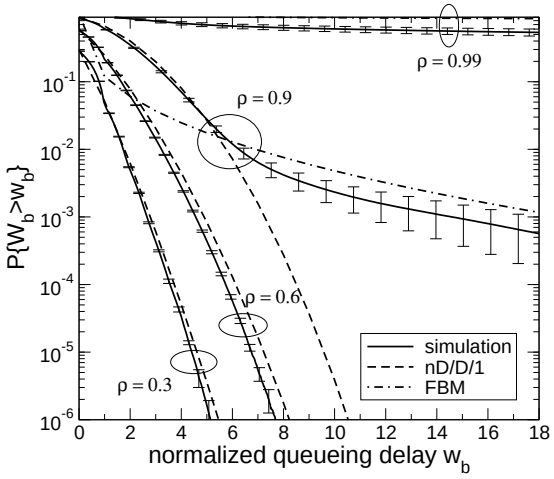


Figure 6.16: Normalized queueing delay wrt. ρ for the M/Pareto model

The FBM approximation well estimates the tendency of the tail behavior. The overlapping of the tails indicates that the parameter β in the large-queue approximation is insensitive to n.

In Fig. 6.14, further results from the cases with $s_{\text{th}} = 16$ KBytes are included to show the influence of the frame size. The absolute delay W is looked at. From the overlapping in the tails of the CCDF curves, it is concluded that the large-queue performance with the M/Pareto input traffic is not affected by the frame size very much.

Summarizing Fig. 6.13 and Fig. 6.14, it is noticeable that the large-queue performance with the M/Pareto traffic is hardly changed by the assembly procedure and the superposition of FEC flows at all, in contrast to the case of the Poissonian traffic. This justifies the assumption of $\beta = 1$ in Eq. (5.3) for the large-queue approximation.

In Fig. 6.15, the influence of the access link rate c_a is inspected with $\rho = 0.9$, $n = 20$ and $s_{\text{th}} =$ 64 KBytes. Note that the small-queue approximation by the $nD/D/1$ model is independent of c_a, which fits the simulation results precisely. With the small value of c_a, the transition to heavy-tailed large-queue behavior is significantly deferred. For example, in the case of $c_a = 10$ Mbps, the large-queue behavior is completely suppressed in the region of interests. This effect is sufficiently captured by the analytical results as well.

Fig. 6.16 investigates the performance at different system loads. Here, $c_a = 50$ Mbps, $n = 20$ and $s_{\text{th}} = 64$ KBytes. It verifies that the heavy-tailedness caused by the LRD is only relevant in the high load situation ($\rho = 0.90, 0.99$). At the low and moderate load ($\rho = 0.3, 0.6$), the small-queue behavior dominates the performance on the concerned probability levels. The large-queue approximations result in very small tail probabilities below 10^{-6} and are therefore not shown in the graph. In this sense, the small-queue approximation suffices for the performance evaluation. In the extremely heavy load situation ($\rho = 0.99$), the small-queue approximation by the $nD/D/1$ model does not fit any more, similar to the scenario with the Poissonian traffic in the preceding section. The large-queue approximation by the FBM model, on the other hand, well captures the evolution of the CCDF curve.

According to the results in Fig. 6.15 and Fig. 6.16, the detrimental effect of the large-time-scale LRD characteristic can be sufficiently suppressed by appropriate link dimensioning and load control. A careful calculation on this point can bring significant performance benefits.

6.4.4.3 Discussion

Comparing the queueing performance between the Poisson process and the M/Pareto model, it is seen that in both cases the small-queue behavior is quite similar and can be uniformly modeled by the $nD/D/1$ system. Under certain circumstances, the small-queue behavior dominates the system performance in the range of practical interests. This also justifies the the application of the FIFO buffer here. In [Car04], it was discovered that more advanced scheduling disciplines like WRR and DRR do not bring much performance improvement as long as the system is not overloaded. In [Hu05], a deadline-based scheduling scheme was applied in the edge node and its performance was compared with the FIFO discipline. The study showed that the difference in the performance is marginal. These results can be well explained by the CBR-alike property of the assembled traffic. In ATM networks, the similar phenomenon was also found for the CBR services and the FIFO discipline is known to be sufficient for the scheduling of CBR flows [GK96, SGV99].

The particularities of the traffic models make differences only in the large-queue behaviors. Especially, the parameter β in the large-queue approximation depends on the assembly procedure when the client traffic follows the Poisson process. By contrast, when the M/Pareto traffic is concerned, the large-queue behavior does not show obvious dependence on the assembly parameters. The difference in the parameter β between the Poissonian traffic and M/Pareto model can be heuristically explained by the variance processes VAR$[V_t]$ of the corresponding frame flows as inspected in Section 6.4.2.3. VAR$[V_t]$ with the Poissonian incoming traffic grows slowly with t. So, even on large time scales, the increment in the variance process due to the assembly can be still apparent. According to Eq. (6.13), the absolute increment in the variance of the aggregated traffic due to the assembly depends on both the number of FEC flows n and the size threshold s_{th}, which reflects itself through the influence on β. On the contrary, VAR$[V_t]$ with LRD incoming traffic increases much faster. On large time scales, the incremental factor caused by the assembly is relatively small and can be neglected. For this reason, the tail behavior is insensitive to n and s_{th}. Furthermore, VAR$[V_t]$ in the case of the Poissonian traffic has a relatively long transition phase on the time scales between the small- and large-time-scale approximations, which can have an influence on the asymptotic constant β. This issue is still subject to advanced studies in the future.

6.5 Delay-Throughput Analysis

On the basis of the performance analysis for the frame assembly in Section 6.3 and for the transmission buffer in Section 6.4, a comprehensive study is to be carried out to outline the interrelation between the delay budget and the throughput.

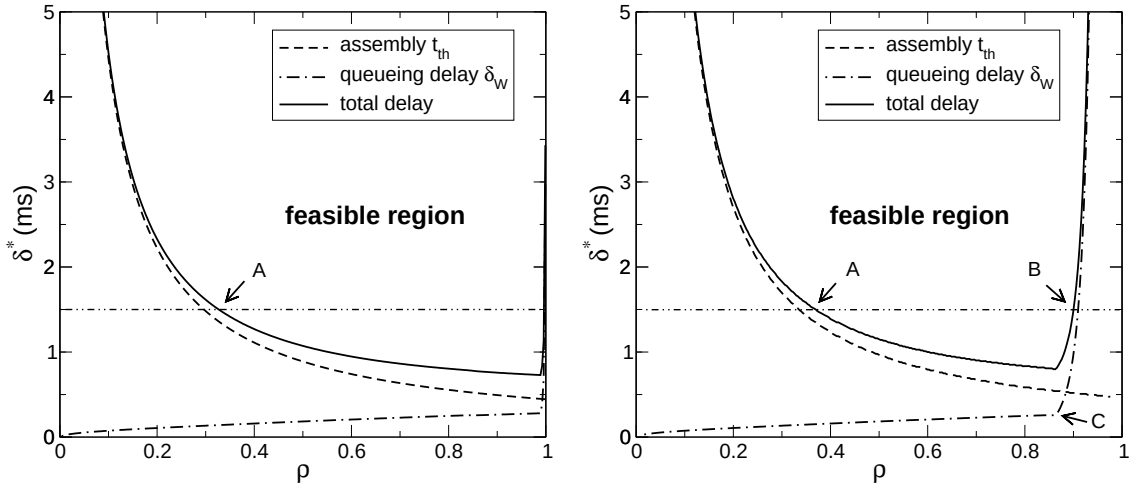
To concentrate on the role of the edge node in the QoS provisioning, the performance requirement on the frame blocking probability in core nodes is ignored. In other words, it is assumed

that the core network can tolerate any throughput that the edge node allows for. However, the ratio of the data traffic rate r_{dat} and the frame header rate r_{sig} is required to be kept above a minimal level for any FEC flow, in order to retain an efficient utilization of the resources on both of the data path and signaling path in the core network. For example, this minimal ratio can correspond to the ratio of the total bandwidth of the data channels and the maximal processing throughput of the SCU in the switching node. According to Eq. (6.1), this is equivalent to set a lower bound for the mean frame size $E[S_{\mathrm{b}}] = r_{\mathrm{dat}}/r_{\mathrm{sig}}$ in the frame assembly. Consequently, the timeout period t_{th} should be large enough to assure the lower bound of $E[S_{\mathrm{b}}]$. The *least necessary timeout period* is determined from Eq. (6.3) numerically.

The upper bound of the frame queueing delay in the transmission buffer is denoted by δ_{W} and defined as the $1-10^{-3}$ quantile of the delay distribution following the definition in [Y.1541]. The CCDF of the queueing delay $P\{W > w\}$ is derived according to the multi-scale analysis in Section 6.4. Note that this is actually the worst case queueing delay because it is based on the assumption that all frames have the maximal frame size s_{th}. Despite of this limitation, it serves as a simple and practical estimation of the delay contribution from the transmission queue. To determine $P\{W > w\}$, both small-queue and large-queue approximations are computed for each value of the delay w. Between them, the larger value of the tail probability is selected as the solution.

Summing up the least necessary timeout period t_{th} and the statistical upper bound δ_{W} of the queueing delay, it results in a total delay bound in the edge node. Approximately, this delay bound corresponds to the minimal delay budget δ^* that can be satisfied in the edge node at a specific system load. In Fig. 6.17, the relation between δ^* and the system load is illustrated through example scenarios. For the evaluation, it is specified that $E[S_{\mathrm{b}}] \geq 55$ KBytes. 10 FECs with the same setting of the assembly parameters are looked at. The size threshold is fixed to $s_{\mathrm{th}} = 64$ KBytes. The incoming traffic is equally distributed to individual FECs. The transmission rate of the wavelength channel is 10 Gbps.

Fig. 6.17(a) shows the case in which the incoming client traffic is modeled by the Poisson process. The least necessary timeout period t_{th}, the bound of the queueing delay and the total delay bound are plotted with respect to the total system load. In a wide range of the system load, the queueing delay is much smaller than the necessary timeout period for the assembly, which is due to the small queue behavior that is similar to the $nD/D/1$ system. At the same time, the very high transmission rate of the wavelength channel assures a fast frame transmission that reduces the absolute queueing time considerably. As a result, the curve of the total delay bound is to a large extent directed by the evolution of the assembly t_{th} until a very high load is reached. In the extremely high load situation, the queueing delay bound begins to be dominated by the large-queue behavior and increases dramatically. The effect exhibits itself in the total delay bound. Note that the region above the curve of the total delay bound represents the feasible combination of the delay budget specification and the throughput that can be satisfied by the edge node. In this sense, it is marked as the *feasible region* in the graph. Since the extremely high loads are in the normal operation of a system always avoided, the main restriction on the feasible region turns out to be the least necessary t_{th} in order to achieve an efficient traffic assembly. Obviously, the traffic at a high aggregation level so that with a adequately large data rate is favourite here, which can be determined from the feasible region quantitatively. Supposing that the delay budget amounts to 1.5 ms, a horizontal line is drawn at $\delta^* = 1.5$ ms and the intersection



(a) Poisson process (b) M/Pareto model ($c_a = 50$ Mbps)

Figure 6.17: Necessary delay budget with respect to the system load

point *A* with the curve of the total delay bound is obtained. Point *A* indicates here the level of the throughput above which the desirable assembly degree can be reached. This serves as an important reference in the traffic/network planning and can be also used as a subsidiary criterion in the admission control.

Fig. 6.17(b) depicts the delay-throughput relation for the case with the M/Pareto traffic. The basic system behaviors are similar to those in Fig. 6.17(a). The major difference appears in the range of the high system load. Due to the LRD of the M/Pareto traffic, the large-queue behavior is accompanied with the heavy-tailed CCDF at high loads, which results in a much faster increase of the queueing delay. This finally imposes a constraint on the feasible region in the high load range. Again, looking at the horizontal line of $\delta^* = 1.5$ ms, the intersection point *A* is similar to that in Fig. 6.17(a). Point *B* represents the maximal admissible throughput in consideration of the delay budget in an edge node.

To mitigate the performance degradation caused by the LRD, a practical measure is to limit the system load under a certain level so as to prevent the emergence of the heavy-tailed large-queue behavior. This load bound is identified by the critical point *C* in Fig. 6.17(b). Since the evolution between the small-queue and large-queue behavior with the M/Pareto traffic is very much influenced by the access link rate c_a, the dependence of the load bound on c_a is evaluated in a numerical approach based on the analysis in Section 6.4. In Fig. 6.18, the solution for the current system scenario is depicted. The x-axis shows the ratio of the access link rate c_a and the wavelength channel rate c in percent. It can be seen that a larger degree of overdimensioning is necessary when the link rate in access networks increases. This quantitative relationship between the load bound and the scaling of link rates in the network hierarchy is very important for the practical network dimensioning.

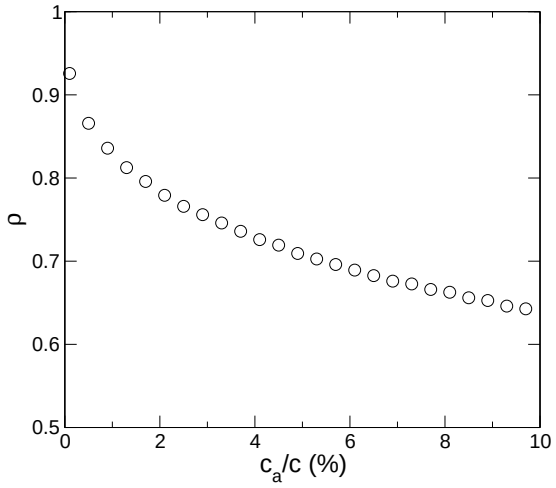


Figure 6.18: Load bound to avoid the impact from LRD

6.6 Admission Control

In the previous section, the delay-throughput relationship is studied without considering the constraints from the sustainable data rate r^*_{dat} and the sustainable header rate r^*_{sig} directly. Only the degree of the assembly is specified by setting a lower bound on the mean frame size. In this section, the attention is extended to the complete admission control problem formulated in Section 6.1.3. Note that the QoS requirements in the core network are satisfied as long as the admitted data rate and frame rate do not exceed r^*_{dat} and r^*_{sig}, respectively. In contrast to Section 6.5, the ratio of the actual data rate and the frame rate, which is equivalent to the mean frame size, does not turn up explicitly as an admission condition in order to allow for the flexibility in the admission control.

6.6.1 Algorithm

The admission control is concerned when a new connection request is to be handled or an existing connection needs to change its traffic parameters as well as the service requirements. Without loss of generality, a new connection request is treated here. In the edge node, this corresponds to setup up a new FEC indexed by i for a certain delay-sensitive service. In Fig. 6.19, the flow chart of a general admission control algorithm is illustrated.

Besides the common contents of a request like the destination address and specific routing policy, the traffic profile and the QoS requirements are of special interests here. For the performance model proposed in this thesis, the traffic profile should include the requested data throughput $r_{\text{req},i}$ and other traffic characteristics like the Hurst parameter in case of the LRD traffic. On the other hand, the QoS requirements include the specification of the E2E loss probability and delay budget. According to the required performance on the E2E loss, the QoS guarantee scheme introduced in Section 2.5.3 can be applied to check the resource availability on the selected path through the core network, in combination with the routing decision. In this way, the sustainable data rate $r^*_{\text{dat},i}$ and $r^*_{\text{sig},i}$ are determined for this FEC. At the same time, the delay budget δ^*_i in the edge node is derived by subtracting the total delay (i.e., the propagation

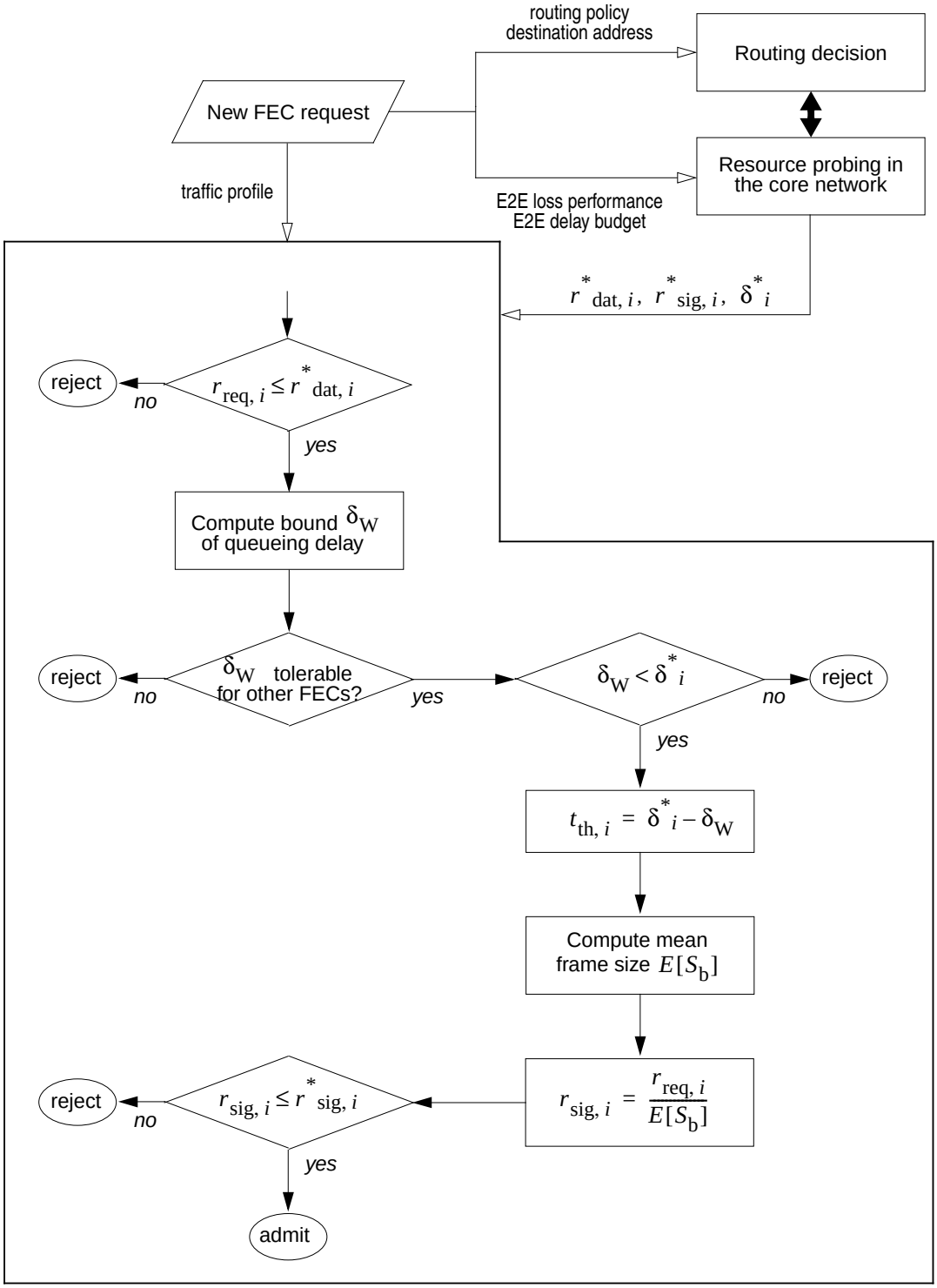


Figure 6.19: Algorithm of the admission control in the edge node

delay, maximal FDL delay, etc.) in the core network from the E2E delay budget. The traffic profile, $r^*_{\mathrm{dat},i}$, $r^*_{\mathrm{sig},i}$ and δ^*_i serve as the input parameters for the admission control algorithm further on, as denoted with the hollow arrow line in Fig. 6.19.

The first admission test is to assure that the requested data rate $r_{\mathrm{req},i}$ does not exceed the sustainable data rate $r^*_{\mathrm{dat},i}$. If this is verified, the statistical bound δ_W for the queueing delay in the transmission buffer is evaluated according to the traffic characteristics, the current system load plus the load from the new request, the total number of FECs and the maximal frame size s_{th}. It is necessary to point out that the analytical method introduced in Section 6.4 assumes that the frame size is fixed to s_{th} and the incoming traffic is equally distributed between all FECs, which is not always the case in the realistic operation. This approach serves here as a worst case analysis. In other words, a certain overestimation is to be tolerated in the evaluation of the queueing delay.

The obtained queueing delay is first applied to check whether all the existing FEC flows can bear the increment in the queueing delay if the new request is accepted. Note that the different FECs can have different delay budgets. If there is no performance violation with the existing FEC flows and the obtained queueing delay is smaller than the delay budget δ^*_i of the new FEC i, the bound of the assembly delay, i.e., the timeout period $t_{\mathrm{th},i}$ is determined from δ^*_i after the deduction of the queueing delay δ_W. With the timeout period, the mean frame size $\mathrm{E}[S_b]$ is derived according to the analysis in Section 6.3.1. The corresponding header rate $r_{\mathrm{sig},i}$ can be further calculated from the requested data rate $r_{\mathrm{req},i}$ and $\mathrm{E}[S_b]$. As known from Section 6.3.2, with the time/size-based assembly a higher data rate always leads to a higher header rate. Therefore, the computed $r_{\mathrm{sig},i}$ represents the maximal header rate of the new request. As the final test, if the obtained header rate is smaller than the sustainable header rate $r^*_{\mathrm{sig},i}$, all admission conditions are satisfied and the new request is accepted.

6.6.2 Practical Issues

As the estimation of the statistical bound δ_W of the queueing delay amounts to a significant computational overhead, measures can be taken to simplify it by setting an upper bound on the system load of the edge node as implied in Section 6.5. In the execution of the algorithm, the delay bound δ_W is treated as a constant that is calculated from the maximal system load. The resulting overestimation is negligible typically when δ^*_i is large.

The algorithm in Fig. 6.19 does not consider any lower bound for the mean frame size in order to allow the flexibility in the admission decision. However, if too many requests with small data rates are admitted, the overall system utilization can be finally limited by the processing overhead in the SCUs of core switches. In practice, supplementary admission policy can be applied to require that the requested data rate $r_{\mathrm{req},i}$ should be above a minimal value (cf. Section 6.5). Alternatively, this issue can be incorporated in the design of the pricing model for the service provisioning.

The parameters for traffic characteristics must be specified for both the delay analysis and the header rate calculation. In the application, the Poisson process is generally used as a worst-case traffic model in the absence of the LRD. In this case, relatively small number of parameters are concerned. Besides the traffic intensity, the statistics of the packet length can be obtained by traffic measurements. In case the M/Pareto model is to be used for LRD traffic flows, the model parameters can be determined from the structures of the client networks (e.g., the access link rate) as well as by the model matching on the basis of traffic measurements.

It is worth to emphasize that the Poisson process and the M/Pareto model are here used only as the example traffic models. The analytical procedure developed in this thesis is applicable for general backbone traffic that can be treated as fluid flows and has the marginal distribution similar to the Gaussian distribution. Without a presumption of the traffic model, the measurements of the mean rate and variance process of the flows are sufficient for the QoS analysis. Especially in the multi-scale queueing analysis for the transmission buffer, there are opportunities to construct simple submodels on the different time scales. For example, the so called parsimonious model [WTSW97] can be used as the large-time-scale submodel to characterize the LRD behaviors. It needs only three parameters: the Hurst parameter, the mean traffic rate and the variance of the rate.

Because of the high degree of traffic aggregation in the transport network, most of the parameters characterizing the traffic for a specific type of service or for a specific client network (e.g., a campus network) have generally stable values in the long term and are suitable to be determined off-line. Especially, on the basis of traffic measurements and statistical data analysis, traffic patterns from all relevant client networks can be summarized and classified into a small number of basic profiles. The admission control procedure can be much simplified for each of the basic profiles. For a new request, the best matching profile is selected from these basic profiles and used for a fast admission control decision.

6.7 Summary

In the admission control in an OPS/OBS edge node, QoS requirements in the core network and in the edge node have to be jointly considered. To assure the E2E loss performance through the core network, for each FEC both the traffic rate on the data path and the header rate on the signaling path must be limited, which can be quantitatively determined from the resource availability in the core switches. Furthermore, the delay-sensitive services require a bounded E2E transmission delay. Since FDL-buffering capacity in OPS/OBS core switches is very limited, the bound of the delay within the core OPS/OBS network can be obtained by the calculation of the propagation delay and the maximal FDL delay on the path. In the edge node, the traffic is subject to the assembly delay and transmission queueing delay.

The traffic assembler in the edge node plays a central role in relating all these different QoS metrics together. The configuration of the assembly parameters directly decides the header rate. The timeout parameter of the assembly itself serves as the bound for the assembly delay. Finally, the assembly procedure changes the traffic characteristics and hence has an impact on the transmission queueing delay of the frames in the edge node.

The contributions of this chapter are summarized into three aspects. First, the quantitative relationships between the traffic assembly and the QoS metrics are analyzed. In order to determine the header rate from the data rate, a closed-form solution to the mean frame size is derived. According to this solution, it is also verified that with fixed assembly parameters the header rate is non-decreasing with the increasing data rate. This means that the load on the data channel and the load on the control channel always reach the maxima at the same time. Therefore, they can be treated uniformly in the admission control. To evaluate the queueing delay in the transmission buffer, it is first figured out that the worst queueing performance emerges when all

frames have the maximal frame size, which is equivalent to a pure size-based assembly scheme. With respect to this worst-case scenario, the variance process of the assembled traffic is analyzed, from which the CBR-alike small-time-scale traffic behavior is identified. According to the multi-scale queueing analysis proposed in Chapter 5, the $nD/D/1$ model is suggested for the approximation of the small-queue behavior of the transmission buffer. On large time scales, the assembly does not change the variance process except causing a constant increment in the variance. Correspondingly, the large-queue behavior is to a great extent determined by the model of the client traffic. In combination with simulations, approximate solutions to the large-queue behavior are obtained for the Poisson traffic model and the M/Pareto model. Integrating the small-queue and large-queue approximation, the CCDF of the queueing delay is derived. The statistical delay bound is further determined at a certain quantile of the delay distribution. These results provide an in-depth understanding of the system behavior of the edge node and serve as the foundation for a system-wide performance evaluation and the design of the admission control.

Second, a comprehensive performance analysis is carried out for the edge node in consideration of the delay budget and the degree of the traffic assembly. The constraint of the delay budget on the throughput is specially inspected. With a given delay budget, it is shown that a minimal throughput is desired in order to keep the assembly degree above a certain level. On the other hand, the upper bound of the throughput is decided by the saturation behavior of the transmission queue at high system loads. Especially with the M/Pareto model, the heavy-tailed large-queue behavior due to the LRD property only has a considerable negative impact on the delay-throughput relation when the system load exceeds a certain level. This threshold load level, referred to as the critical point in the context, is derived depending on the access link rate of the aggregated traffic. The study shows that a larger access link rate shifts the critical point to a lower load level. This critical point is significant for the specification of the maximal system load in the practical network design and operation.

Finally, an admission control algorithm is proposed on the basis of the aforementioned performance analysis. Shortly speaking, the algorithm follows the thread to test the statistical bound of the worst-case queueing delay, determine the timeout period, calculate the mean frame size and test the header rate, sequentially. In practice, the procedure can be simplified. Also, further supplementary admission conditions can be included. The traffic profile of the request, which is necessary for the admission control, is determined according to off-line traffic measurements. Pattern-based profile specification is suggested.

7 Conclusions and Outlook

This dissertation studies the QoS provisioning in the edge node of optical packet switched (OPS) and optical burst switched (OBS) networks. It models, analyzes and evaluates the link-layer performance of the edge node and provides the solution to the admission control for services with guaranteed QoS. At the same time, in order to tackle the complex traffic characteristics in the superposition of traffic flows after the assembly, a new method for multi-scale queueing analysis is proposed and applied.

Chapter 2 gives an overview on the general OPS/OBS network architectures. Since the full OPS node with the optical switching control unit (SCU) is difficult to be implemented in the foreseeable future, the attention on the OPS is confined to the opto-electronic solution. OPS and OBS thus share common performance issues like the bottleneck of the header processing and the data loss due to the on-the-fly switching. Correspondingly, they can apply similar solutions of the traffic assembly in edge nodes as well as the channel management, contention resolution and scheduling in core switching nodes. The schemes for QoS provisioning in core OPS/OBS networks in literature are surveyed and classified. QoS mechanisms for service differentiation in single switching node are presented first, which are prerequisite in the QoS provisioning. Then, QoS models for the absolute E2E guarantee of the loss performance are introduced.

Chapter 3 concentrates on the edge node of OPS/OBS networks in which the traffic assembler and transmission scheduler are recognized as the central functional units in terms of the link layer. Different assembly schemes are introduced. The timeout period and frame size threshold are shown to be the two most important assembly parameters. The relevant work in the traffic characterization and performance evaluation for the traffic assembly is reviewed. Most of the work relied on the method of point process analysis and considered only a single assembly queue. This limits the applicability of the results when a system-wide scenario is to be evaluated and more complex input traffic must be considered. There was relatively less work done with regard to the traffic scheduling in the edge node. The available work in literature is briefly surveyed. In connection to the introduction in Chapter 2, the role of the edge node in the E2E QoS guarantee is further clarified. It is particularly highlighted that (a) the edge node must take care of the generated header rate to avoid congestion in the SCUs of switching nodes; (b) the delay in the edge node should be bounded by the allocated delay budget. Taking these aspects into consideration, the admission problem is elaborated for an assured service provisioning.

An overview on the characteristics of the client traffic, typically IP traffic, is provided in Chapter 4. Traffic measurements in recent years discovered that the backbone IP traffic exhibits different characteristics on different ranges of time scales. Among them, the uncorrelated property in small time scales and long range dependence (LRD) on large time scales are most prominent.

The M/Pareto model is introduced as a popular traffic model that well seizes this kind of feature. To deal with such complex traffic patterns in the queueing analysis, the methods of time scale decomposition and integrated analysis are introduced. The time scale decomposition approach was developed for the queueing analysis in ATM networks. It formulates separate cell-scale and burst-scale subproblems, each of which only considers the specific traffic characteristic on the respective range of time scales. A closed-form solution is obtained by integrating the solutions of the subproblems. However, this approach does not explicitly include the time scale as a model parameter. In a generalized application, it is difficult to determine how to decompose the time scales to build up the subproblems. The integrated approach, on the other hand, explicitly models the time-scale-dependent traffic characteristics by a function of the time scale, e.g., the effective bandwidth or the variance process. The solution is derived by an optimization procedure with respect to this function. This makes the analysis quite general for multi-scaling queueing problem. However, the computational complexity is relatively high and not suitable for an on-line application, for example, in the performance estimation for admission control.

To avoid the disadvantages of the original time scale decomposition and the integrated analytical approaches, Chapter 5 proposes a new method with a combined application of the principles of both sides. This method employs the effective bandwidth or the variance process to characterize the traffic. The changes in the characteristics along the time scale can thus be precisely identified, according to which the time scales are decomposed into multiple segments. For each segment, a submodel can be constructed with regard to the traffic characteristic on the respective time scales. In case the submodels are instances of standard queueing models, the overall queueing performance can be conveniently obtained by integrating the known solutions of the standard queueing models. The method is verified by a comparative study with simulations and the analytical results show a satisfying accuracy.

Chapter 6 focuses on the performance analysis and the admission problem in the edge node. In connection to the edge node's tasks in the QoS provisioning illuminated in Chapter 3, the admission problem is formally formulated, in which the frame header rate and the transmission queueing delay are recognized as the two most important performance measures. In the first part of the chapter, by a fluid-flow modeling of the incoming client traffic, the frame header rate is derived from the assembly parameters of the timeout period and the frame size threshold. It is also verified that the resulting header rate is non-decreasing with the increase in the traffic rate under fixed assembly parameters. This allows for a solely attention on the maximal data rate of a service request in the assessment by the admission control. Furthermore, the transmission queueing delay is analyzed by the method proposed in Chapter 5. The variance process of the assembled traffic is derived to identify the traffic characteristic on small time scales and large time scales, respectively. Correspondent submodels are constructed and their solutions are provided. It is shown that the overall queueing performance is well estimated by integrating the solutions of individual submodels.

In the second part of Chapter 6, the total necessary delay budget for the edge node is studied with respect to the node throughput under the condition that the assembly degree, i.e., the mean frame size, is fixed. The results show the significance of a sufficient aggregation level of the incoming traffic to achieve an efficient traffic assembly. The constraint of the queueing delay on the throughput is clearly depicted. On this point, the influence of the LRD is especially inspected.

In the last part of the chapter, an admission control algorithm is proposed to integrate the derived QoS model for the edge node and those known solutions contributed to the core networks. Thus, it provides a complete approach for the E2E QoS guarantee in OPS/OBS networks. In this algorithm, the E2E loss performance on the data path is assured by limiting the traffic rate of the concerned FEC flow. This corresponds to the sustainable data rate that can be derived by those performance models introduced in Chapter 2. To assure the timely header processing in core switches, a sustainable frame header rate is also specified. For a new request, the algorithm analyzes the frame queueing delay in the edge node, determines the assembly timeout in consideration of the delay budget and judges whether the constraints of the sustainable data/header rate can be held. The algorithm is applicable for general backbone traffic. In practice, the traffic profile of the request needs to be determined for the admission control. This is realized either by traffic measurements or by the assignment of a predefined basic traffic profile to the request.

Further work could be the application of this QoS model in the traffic engineering, network planning and optimization in combination with those schemes assuring the E2E loss performance in the core network. The work could be further extended to include the economic study of the network. Especially, it would be very interesting to compare these results with those obtained for circuit-switched OTNs in order to illuminate the advantages and disadvantages of the different switching technologies in the provisioning of guaranteed QoS. This would serve as a significant reference in aligning the research activities for future OTNs.

A Queueing Delay with Superposition of Heterogeneous FEC Flows

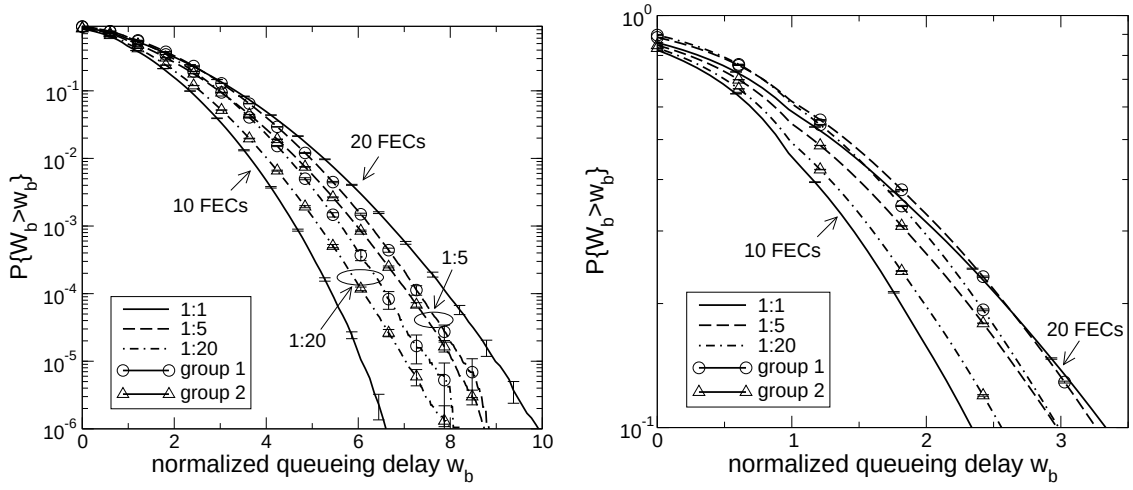
In the system model in Chapter 6, only one individual service class is considered. All FECs in Fig. 6.2 are of the same guaranteed service class. Therefore, the incoming client traffic of the FECs is supposed to have the similar traffic characteristics. Consequently, the same traffic model can be used for all FECs. For most of the traffic parameters like the packet length distribution, the Hurst parameter, etc., it is also reasonable to apply the same parameter setting for all FECs. This leads to the adoption of the homogeneous FEC flows in the evaluation of the queueing delay in Chapter 6.

Nevertheless, the traffic intensity between E2E network nodes generally differs from each other. So, it is quite normal that the FEC flows have different traffic rates, which should be taken into account in the performance evaluation in the edge node. This appendix focuses on this special case of heterogeneous FEC flows and inspects the influence of the traffic distribution among the FECs on the queueing performance by means of simulations.

A.1 System Scenario

The concerned scenario is constructed on the basis of the system model illustrated in Fig. 6.2. Totally 20 FECs are looked at. The pure size-based assembly is applied for each FEC and the size threshold $s_{\text{th}} = 64$ KBytes. The assembled frames are scheduled according to the FIFO discipline in the unbounded transmission buffer. The transmission rate of the channel is 10 Gbps and the total system load is set to 0.9. Similar to Chapter 6, the incoming client traffic is synthesized by the Poisson process and the M/Pareto model, respectively. The same setting of traffic parameters as described in Section 6.2 is adopted.

To study the impact of the heterogeneous traffic rates of the FECs, the 20 FECs are classified into two groups. Each group consists of 10 FECs and is denoted by Group 1 and Group 2, respectively. The total offered traffic is distributed to Group 1 and Group 2 according to a certain ratio. Within each group, the offered traffic is equally distributed to individual FECs. The ratio of the offered traffic between Group 1 and Group 2 is tuned for different simulation runs. In the following presentation of the simulation results, the cases with the ratio of $1:1$, $1:5$ and $1:20$ are shown. Note that the case with the ratio equal to $1:1$ is equivalent to the scenario of the homogeneous FEC flows.



(a) tail behavior of the CCDF (b) CCDF in the domain of the small queue

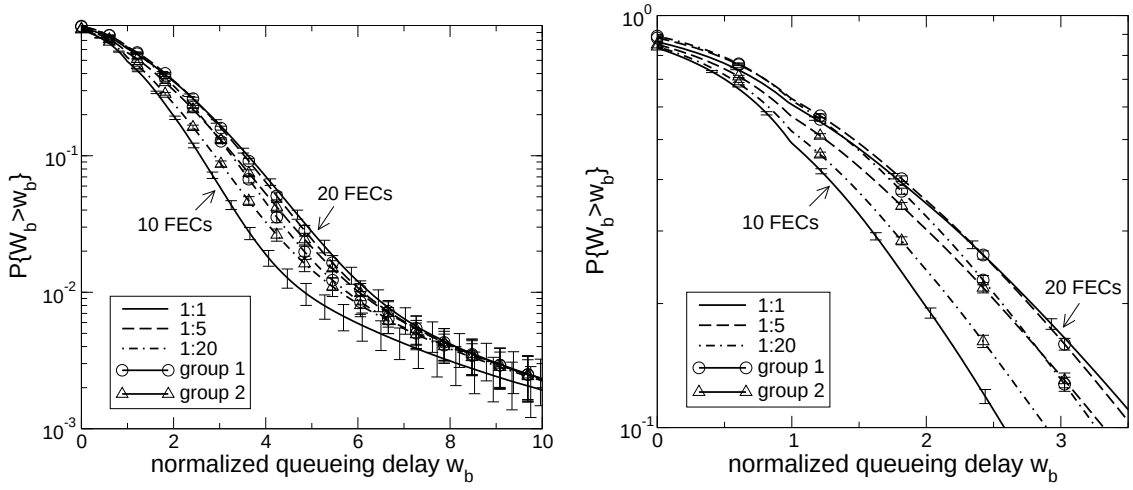
Figure A.1: CCDF of the queueing delay with the Poissonian traffic

A.2 Simulation Results

The CCDF of the queueing delay is first studied with the Poissonian traffic and the simulation results are shown in Fig. A.1. In the case of the homogeneous FEC flows (i.e., ratio 1 : 1), Group 1 and Group 2 receive the same queueing performance and therefore there is no need to distinguish between them. Whereas with the heterogeneous flow rates (i.e., ratio 1 : 5 and 1 : 20), the delay CCDF is measured and plotted for the frames from Group A and Group B separately. In Fig. A.1(a), the tail behavior of the CCDF is outlined. As shown, when the difference in the traffic intensity gets large, the delay performance in terms of the tail behavior turns better for both groups. It can be imagined that when the diversity further grows, the extreme case will be that the traffic intensity of Group 1 becomes zero and the entire offered traffic is distributed to Group 2. This degrades to the case with 10 homogeneous FEC flows, the delay performance of which is also depicted. It is seen that the CCDFs of the 10 homogeneous FECs and 20 homogeneous FECs serve as the lower bound and the upper bound respectively for the tail behavior of the transmission queue. Furthermore, comparing the delay of Group 1 and Group 2 in the same simulation run, it is recognized that the group with the relatively small traffic rate (Group 1) has a worse tail performance than the group with the large traffic rate (Group 2). However, the difference in the CCDF becomes apparent only when there is a large unbalance in the traffic intensity among the FECs.

Fig. A.1(b) zooms out the delay CCDFs in the scope of the small queue. In the cases of heterogeneous traffic rates, although the CCDF of Group 1 can exceed the curve of the 20 homogeneous FEC flows in a limited range of the queue, the latter begins to envelope the CCDF curves soon as the queue grows.

When the incoming client traffic is synthesized by the M/Pareto model, the LRD property has an influence on the delay performance. However, the basic system behavior caused by the heterogeneous traffic rates is similar to that of the Poissonian traffic. Fig. A.2(a) shows the tail behavior of the CCDF of the transmission queue with the M/Pareto model. Starting from the curve of the 20 homogeneous FEC flows, the CCDF curves of both Group 1 and Group 2



(a) tail behavior of the CCDF (b) CCDF in the domain of the small queue

Figure A.2: CCDF of the queueing delay with the M/Pareto traffic model ($c_a = 50$ Mbps)

decline as the diversity in the traffic intensity increases. Finally, the performance with the 10 homogeneous FEC flows represents an lower bound. Especially, it is noticeable that all the CCDF curves go on to exhibit the heavy-tailed property and overlap with each other in the scope of the large queue. This means that the large-queue behavior is not influenced by the traffic distribution among the FECs, which is consistent with the analysis in Chapter 6. The gap between the curves of *20 FECs* and *10 FECs* corresponds to the maximal extent of the impact from the ratio of the traffic intensity between the two groups. Due to the overlapping in the domain of the large queue, the gap is very constrained, indicating that the possible influence from the ratio of the traffic intensity is limited.

To figure out the details in the scope of the small queue, Fig. A.2(b) zooms out the CCDFs from Fig. A.2(a). The behaviors of the CCDFs are analogous to those in Fig. A.1(b).

A.3 Conclusions

Through the simulation study on the transmission queueing delay, it is seen that the heterogeneous traffic rates among the FECs lead to a diversity in the delay performance among the FECs. The FEC flows with a small traffic rate have relatively large queueing delay and those flows with a large traffic rate tend to have small queueing delay. In comparison with a correspondent scenario of homogeneous FEC flows, heterogeneous traffic rates can result in a worse queueing performance only in the limited domain of the small queue. In the scope of the large queue, however, the delay CCDF of the homogeneous FEC flows serves as an upper bound. In this sense, the homogeneous traffic stands for an important reference scenario in the evaluation of the tail probability of the queueing delay in the OPS/OBS edge node.

Bibliography

[AAB+07] J. Aracil, N. Akar, S. Bjornstad, M. Casoni, K. Christodoulopoulos, D. Careglio, J. Fdez-Palacios, C. Gauger, O. G. de Dios, G. Hu, E. Karasan, M. Klinkowski, D. Moratt'o, R. Nejabati, H. Overby, C. Raffaelli, D. Simeonidou, N. Stol, G. Tosi-Beleffi, and K. Vlachos. Research in optical burst switching within the e-Photon/ONe Network of Excellence. *Optical Switching and Networking*, 4(1):1–19, February 2007.

[AK93] H. Akimaru and K. Kawahima. *Teletraffic - Theory and Applications*. Springer Verlag, 1993.

[APW05] R. Almeida, J. Pelegrini, and H. Waldman. A generic-traffic optical buffer modeling for asynchronous optical switching networks. *IEEE Communications Letters*, 9(2):175–177, 2005.

[BBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated services. *RFC 2475, IETF*, December 1998.

[BBPV03] M. Baresi, S. Bregni, A. Pattavina, and G. Vegetti. Deflection routing effectiveness in full-optical IP packet switching networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 2, pages 1360–1364 vol.2, 2003.

[BBWP03] D. J. Blumenthal, J. E. Bowers, L. W. Wang, and H. N. Poulsen. Optical signal processing for optical packet switching networks. *IEEE Communications Magazine*, 41(2):S23–S29, 2003.

[BC89] R. Ballart and Y.-C. Ching. SONET: now it's the standard optical network. *IEEE Communications Magazine*, 27(3):8–15, March 1989.

[BC00] S. Bodamer and J. Charzinski. Evaluation of effective bandwidth schemes for self-similar traffic. In *13th ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management*, 2000.

[BCS94] R. Braden, D. Clark, and S. Shenker. Integrated services in the Internet architecture: an overview. *RFC 1633, IETF*, June 1994.

[BD02] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag, 2nd edition, 2002.

[BFB93] F. Borgonovo, L. Fratta, and J. Bannister. Unslotted deflection routing in all-optical networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, pages 119–125, Houston, November 1993.

[BhH89] A. Bhargava, P. humblet, and M. G. Hluchyj. Queueing analysis of continuous bit-stream transport in packet networks. In *Proceedings of IEEE Global Telecommunications Conference (Globecom)*, volume 2, pages 903–907, November 1989.

[Blu01] D. J. Blumenthal. Photonic packet switching and optical label switching. *Optical Networks Magazine*, 2(6):1–12, 2001.

[Bod04] S. Bodamer. *Mechanisms for Relative Quality of Service Differentiation in IP Network Nodes - 88th Report on Studies in Congestion Theory.* PhD thesis, University of Stuttgart, 2004.

[Bou98] J.-Y. L. Boudec. Application of network calculus to guaranteed service networks. *IEEE Transactions on Information Theory*, 44(3):1087–1096, May 1998.

[BPR01] T. Bonald, A. Proutiere, and J. Roberts. Statistical performance guarantees for streaming flows using expedited forwarding. In *Proceedings of IEEE INFOCOM*, 2001.

[BPS94] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer. Photonic packet switches: architectures and experimental implementations. *Proceedings of the IEEE*, 82(11):1650–1667, 1994.

[BS05] N. Barakat and E. Sargent. Analytical modeling of offset-induced priority in multiclass OBS networks. *IEEE Transactions on Communications*, 53(8):1343–1352, 2005.

[BS06] N. Barakat and E. Sargent. Separating resource reservations from service requests to improve the performance of optical burst-switching networks. *IEEE Journal on Selected Areas in Communications*, 24(4):95–107, 2006.

[BT01] J.-Y. L. Boudec and P. Thiran. *Network Calculus – a Theory of Deterministic Queueing Systems for the Internet*. Springer-Verlag, LNCS 2050, 2001.

[Buc05] H. Buchta. *Analysis of Physical Constraints in an Optical Burst Switching Network*. PhD thesis, Technische Universität Berlin, 2005.

[Bur81] D. Burman. Insensitivity in queueing systems. *Advances in Applied Probability*, 13(4):846–859, 1981.

[BZ96] J. C. Bennett and H. Zhang. WF2Q: worst-case fair weighted fair queueing. In *Proceedings of IEEE INFOCOM*, pages 120–128, San Francisco, March 1996.

[Cal00] F. Callegati. Optical buffers for variable length packets. *IEEE Communications Letters*, 4(9):292–294, 2000.

[Car04] J. Cartagena. Scheduling of assembled traffic in IP-over-Photonics edge nodes. Student thesis, University of Stuttgart, September 2004.

[CB97] M. E. Crovella and A. Best. Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6), 1997.

[CC01] F. Callegati and W. Cerroni. Wavelength allocation algorithms in optical buffers. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 2, pages 499–503 vol.2, 2001.

[CCC+04] F. Callegati, W. Cerroni, G. Corazza, C. Develder, M. Pickavet, and P. Demeester. Scheduling algorithms for a slotted packet switch with either fixed or variable length packets. *Photonic Network Communications*, 8(2):163–176, September 2004.

[CCF01] T. Chich, J. Cohen, and P. Fraigniaud. Unslotted deflection routing: a practical and efficient protocol for multihop optical networks. *IEEE/ACM Transactions on Networking*, 9(1):47–59, February 2001.

[CCK05] J. Choi, J. Choi, and M. Kang. Dimensioning burst assembly process in optical burst switching networks. *IEICE - Transactions on Communications*, E88-B(10):3855–3863, 2005.

[CCRS06] F. Callegati, W. Cerroni, C. Raffaelli, and M. Savi. QoS differentiation in optical packet-switched networks. *Computer Communications*, 29(7):855–864, April 2006.

[CEBSC06] S. Charcranoon, T. El-Bawab, J.-D. Shin, and H. Cankaya. Group-scheduling for multi-service optical burst switching (OBS) networks. *Photonic Network Communications*, 11(1):99–110, January 2006.

[CHA+01] M. Chia, D. Hunter, I. Andonovic, P. Ball, I. Wright, S. Ferguson, K. Guild, and M. O'Mahony. Packet loss and delay performance of feedback and feed-forward arrayed-waveguide gratings-based optical packet switches with WDM inputs-outputs. *IEEE/OSA Journal of Lightwave Technology*, 19(9):1241–1254, 2001.

[CHT01] Y. Chen, M. Hamdi, and D. Tsang. Proportional QoS over OBS networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, San Antonio, November 2001.

[CLCQ02] X. Cao, J. Li, Y. Chen, and C. Qiao. Assembling TCP/IP packets in optical burst switched networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, Taipei, November 2002.

[CLW96] G. L. Choudhury, D. M. Lucantoni, and W. Whitt. Squeezing the most out of ATM. *IEEE Transactions on Communications*, 44(2):203–217, 1996.

[COR01] J. Chuzhoy, R. Ostrovsky, and Y. Rabani. Approximation algorithms for the job interval selection problem and related scheduling problems. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 348–356, October 2001.

[CS98] J. Choe and N. Shroff. A central-limit-theorem-based approach for analyzing queue behavior in high-speed networks. *IEEE/ACM Transactions on Networking*, 6(5), 1998.

[CSS99] C. Courcoubetis, V. Siris, and G. Stamoulis. Application of the many sources asympototic and effective bandwidths to traffic engineering. *Telecommunication Systems*, 12:167–191, 1999.

[CT95] C.-S. Chang and J. Thomas. Effective bandwidth in high-speed digital networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1091–1100, 1995.

[CTT99] G. Castanon, L. Tancevski, and L. Tamil. Routing in all-optical packet switched irregular mesh networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, volume 1B, pages 1017–1022, 1999.

[CWXQ03] Y. Chen, H. Wu, D. Xu, and C. Qiao. Performance analysis of optical burst switched node with deflection routing. In *Proceedings of the IEEE International Conference on Communication (ICC)*, Anchorage, May 2003.

[CZC+04] K. Cheng, Y. Zhu, X. Cao, D. H. K. Tsang, and D. T. K. Tong. Architecture for an optical burst switch with WDM shared output buffer. *Journal of Optical Networking*, 3(6):417–432, 2004.

[DB02] M. Dueser and P. Bayvel. Analysis of a dynamically wavelength-routed optical burst switched network architecture. *IEEE/OSA Journal of Lightwave Technology*, 20(4):574–585, April 2002.

[DDC+03] L. Dittmann, C. Develder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Sole-Pareta, W. Cerroni, N. Leligou, L. Dembeck, B. Mortensen, M. Pickavet, N. Le Sauze, M. Mahony, B. Berde, and G. Eilenberger. The European IST project DAVID: a viable approach toward optical packet switching. *IEEE Journal on Selected Areas in Communications*, 21(7):1026–1040, 2003.

[DG01] K. Dolzer and C. M. Gauger. On burst assembly in optical burst switching networks—a performance evaluation of just-enough-time. In *Proceedings of the 17th International Teletraffic Congress (ITC 17)*, 2001.

[DGSB01] K. Dolzer, C. M. Gauger, J. Späth, and S. Bodamer. Evaluation of reservation mechanisms for optical burst switching. *AEÜ International Journal of Electronics and Communications*, 55(1):18–26, January 2001.

[Dol04] K. Dolzer. *Mechanisms for Quality of Service Differentiation in Optical Burst Switched Networks - 86th Report on Studies in Congestion Theory.* PhD thesis, University of Stuttgart, 2004.

[dVRG04] M. de Vega Rodrigo and J. Goetz. An analytical study of optical burst switching aggregation strategies. In *Proceedings of the Third International Workshop on Optical Burst Switching (WOBS)*, 2004.

[EBS02] T. S. EI-Bawab and J.-D. Shin. Optical packet switching in core networks: between vision and reality. *IEEE Communications Magazine*, 40(9):60–65, 2002.

[EL00] V. Eramo and M. Listanti. Packet loss in a bufferless optical WDM switch employing shared tunable wavelength converters. *IEEE/OSA Journal of Lightwave Technology*, 18(12):1818–1833, December 2000.

[ELP03] V. Eramo, M. Listanti, and P. Pacifici. A comparison study on the number of wavelength converters needed in synchronous and asynchronous all-optical switching architectures. *IEEE Journal on Lighwave Technology*, 21(2):340–355, February 2003.

[ELS05] V. Eramo, M. Listanti, and M. Spaziani. Resources sharing in optical packet switches with limited-range wavelength converters. *IEEE/OSA Journal of Lightwave Technology*, 23(2):671–687, 2005.

[EM97] A. Elwalid and D. Mitra. Traffic shaping at a network node: theory, optimum design, admission control. In *Proceedings of IEEE INFOCOM*, 1997.

[ENNS00] A. Erramilli, O. Narayan, A. Neidhardt, and I. Saniee. Performance impacts of multi-scaling in wide area TCP/IP traffic. In *Proceedings of IEEE INFOCOM*, Tel Aviv, March 2000.

[ENW96] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2):209–223, April 1996.

[FBTZ02] V. Firoiu, J.-Y. L. Boudec, D. Towsley, and Z.-L. Zhang. Theories and models for Internet quality of service. *Proceedings of the IEEE*, 90(9):1565–1590, September 2002.

[Fel00] A. Feldmann. *Self-Similar Network Traffic and Performance Evaluation*, chapter Characteristics of TCP connection arrivals. John Wiley & Sons, 2000.

[FGHW99] A. Feldmann, A. C. Gilbert, P. Huang, and W. Willinger. Dynamics of IP traffic: a study of the role of variability and the impact of control. *ACM SIGCOMM Computer Communication Review*, 29(4):301–313, October 1999.

[FGW98] A. Feldmann, A. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. In *Proceedings of the ACM/SIGOMM*, Vancouver, September 1998.

[FJ95] S. Floyd and V. Jacobson. Link-sharing and resource management models for packet networks. *IEEE/ACM Transactions on Networking*, 3(4):365–386, 1995.

[FKW+01] T. Fjelde, A. Kloch, D. Wolfson, B. Dagens, A. Coquelin, I. Guillemot, F. Gaborit, F. Poingt, and M. Renaud. Novel scheme for simple label-swapping employing XOR logic in anintegrated interferometric wavelength converter. *IEEE Photonics Technology Letters*, 13(7):750–752, 2001.

[FLB05] D. Fiems, K. Laevens, and H. Bruneel. Performance analysis of an all-optical packet buffer. In *Proceedings of the Optical Network Design and Modelling Conference (ONDM)*, pages 221–226, 2005.

[FPS02] H. Feng, E. Patzak, and J. Saniter. Size and cascadability limits of SOA based burst switching nodes. In *Proceedings of the 28th European Conference on Optical Communication (ECOC 2002)*, Copenhagen, September 2002.

[FZJ05] F. Farahmand, Q. Zhang, and J. Jue. Dynamic traffic grooming in optical burst-switched networks. *IEEE/OSA Journal of Ligthwave Technology*, 23(10):3167–3177, 2005.

[Gau02] C. M. Gauger. Dimensioning of FDL buffers for optical burst switching nodes. In *Proceedings of the Optical Network Design and Modelling Conference (ONDM)*, Torino, February 2002.

[Gau03] C. M. Gauger. Trends in optical burst switching. In *Proceedings of SPIE ITCOM*, September 2003.

[Gau04] C. M. Gauger. Optimized combination of converter pools and FDL buffers for contention resolution in optical burst switching. *Photonic Network Communications*, 8(2):139–148, September 2004.

[Gau06] C. M. Gauger. *Novel Network Architecture for Optical Burst Transport - Communication Networks and Computer Engineering Report No. 92.* PhD thesis, University of Stuttgart, 2006.

[GCT00] A. Ge, F. Callegati, and L. S. Tamil. On optical burst switching and self-similar traffic. *IEEE Communications Letter*, 4(3), 2000.

[GDSB01] C. M. Gauger, K. Dolzer, J. Späth, and S. Bodamer. Service differentiation in optical burst switching networks. In *Proceedings of the 2. ITG Symposium on Photonic Networks*, pages 124–132, Dresden, March 2001.

[Gil01] A. Gilbert. Multiscale analysis and data networks. *Applied and Computational Harmonic Analysis*, 10:185–202, 2001.

[GK96] M. Grossglauser and S. Keshav. On CBR service. In *Proceedings of IEEE INFOCOM*, pages 129–137, San Francisco, March 1996.

[GKS04] C. M. Gauger, M. Köhn, and J. Scharf. Comparison of contention resolution strategies in OBS network scenarios. In *Proceedings of the 6th International Conference on Transparent Optical Networks (ICTON)*, volume 1, pages 18–21 vol.1, 2004.

[Gol94] S. Golestani. A self-clocked fair queueing scheme for broadband applications. In *Proceedings of IEEE INFOCOM*, pages 636–646, April 1994.

[GRG+98a] P. Gambini, M. Renaud, C. Guillemot, F. Callegati, I. Andonovic, B. Bostica, D. Chiaroni, G. Corazza, S. Danielsen, P. Gravey, P. Hansen, M. Henry, C. Janz,

A. Kloch, R. Krahenbuhl, C. Raffaelli, M. Schilling, A. Talneau, and L. Zucchelli. Transparent optical packet switching: network architecture and demonstrators in the KEOPS project. *IEEE Journal on Selected Areas in Communications*, 16(7):1245–1259, 1998.

[GRG+98b] C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clerot, S. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P. Hansen, D. Hunter, A. Kloch, R. Krahenbuhl, B. Lavigne, A. Le Corre, C. Raffaelli, M. Schilling, J.-C. Simon, and L. Zucchelli. Transparent optical packet switching: the European ACTS KEOPS project approach. *Journal of Lightwave Technology*, 16(12):2117–2134, 1998.

[GTJL05] X. Guan, I. Thng, Y. Jiang, and H. Li. Providing absolute QoS through virtual channel reservation in optical burst switching networks. *Computer Communications*, 28:967–986, 2005.

[HA00] D. Hunter and I. Andronovic. Approaches to optical Internet packet switching. *IEEE Communications Magazine*, 38(9):116–122, 2000.

[HCA98] D. Hunter, M. Chia, and I. Andonovic. Buffering in optical packet switches. *IEEE/OSA Journal of Ligthwave Technology*, 16(12):2081–2094, 1998.

[HDG03] G. Hu, K. Dolzer, and C. M. Gauger. Does burst assembly really reduce self-similarity? In *Proceedings of the Optical Fiber Communication Conference (OFC)*, Atlanta, March 2003.

[HHA07] J. A. Hernández, G. Hu, and J. Aracil. Analysis of IP delay variation in edge OBS nodes. In *Proceedings of the 12th European Conference on Networks and Optical Communications (NOC)*, Stockholm, June 2007.

[HK06] G. Hu and M. Koehn. Evaluation of packet delay in OBS edge nodes. In *Proceedings of the 8th International Conference on Transparent Optical Networks (ICTON)*, Nottingham, 2006.

[HMQ+05] G. Hu, G. Muretto, F. Querzola, C. Gauger, and C. Raffaelli. Traffic and performance analysis of optical packet/burst assembly with self similar traffic. In *Proceedings of the 7th International Conference on Transparent Optical Networks (ICTON)*, volume 1, 2005.

[Hu04] G. Hu. Impact of access bandwidth on aggregated traffic behavior and queueing performance. In *Proceedings of the 12th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems/3rd Polish-German Teletraffic Symposium*, 2004.

[Hu05] G. Hu. QoS guarantee for real time services in OBS edge nodes. Technical Report for European Network of Excellence E-Photon/ONe, University of Stuttgart, August 2005.

[Hu06] G. Hu. Performance model for a lossless edge node of OBS networks. In *Proceedings of IEEE Global Telecommunications Conference (Globecom)*, San Francisco, November 2006.

[HXL+05] A. Huang, L. Xie, Z. Li, D. Lu, and P.-H. Ho. Optical self-similar cluster switching (OSCS) - a novel optical switching scheme by detecting self-similar traffic. *Photonic Network Communications*, 10(3):297–308, November 2005.

[IA02] M. Izal and J. Aracil. On the influence of self-similarity on optical burst switching traffic. In *IEEE Global Telecommunications Conference (Globecom)*, 2002.

[IJAM06] M. Izal, D. M. J. Aracil, and E. Magana. Delay-throughput curves for timer-based OBS burstifiers with light load. *IEEE Journal on Lighwave Technology*, 24(1), 2006.

[ISNS02] M. Iizuka, M. Sakuta, Y. Nishino, and I. Sasase. A scheduling algorithm minimizing voids generated by arriving bursts in optical burst switched WDM network. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, volume 3, pages 2736–2740, Taipei, November 2002.

[JG03a] S. Junghans and C. M. Gauger. Architectures for resource reservation modules for optical burst switching core nodes. In *Proceedings of the 4. ITG Symposium on Photonic Networks*, Leipzig, May 2003.

[JG03b] S. Junghans and C. M. Gauger. Resource reservation in optical burst switching: Architectures and realizations for reservation modules. In *Proceedings of the Optical Networking and Communications Conference (OptiComm)*, Dallas, October 2003.

[Jun04] S. Junghans. A testbed for control systems of optical burst switching core nodes. In *Proceedings of the Third International Workshop on Optical Burst Switching (WOBS)*, San Jose/CA, October 2004.

[Jun05] S. Junghans. Pre-estimate burst scheduling (PEBS): an efficient architecture with low realization complexity for burst scheduling disciplines. In *Proceedings of the Fifth International Workshop on Optical Burst/Packet Switching (WOBS)*, 2005.

[KA04] A. Kaheel and H. Alnuweiri. Quantitative QoS guarantees in labeled optical burst switching networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, pages 1747–1753, Dallas, 2004.

[KA05] A. Kaheel and H. Alnuweiri. Batch scheduling algorithms: a class of wavelength schedulers in optical burst switching networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 3, pages 1713–1719, 2005.

[Kan06] B. Kanafin. Modeling and analysis of quality of service in OBS edge nodes. Master thesis, University of Stuttgart, January 2006.

[Kau02] F.-J. Kauffels. *Optische Netze*. mitp-Verlag, 2002.

[kcMT98] k claffy, G. Miller, and K. Thompson. The nature of the beast: recent traffic measurements from an Internet backbone. In *Proceedings of International Networking Conference (INET)*, 1998.

[Kel96] F. Kelly. *Stochastic Networks: Theory and Applications*, chapter Notes on effective bandwidths, pages 141–168. Number 4. Oxford University Press, 1996.

[Kle75] L. Kleinrock. *Queueing Systems: Volume 1/2*. John Wiley & Sons, 1975.

[KM03] K.-I. Kitayama and M. Murata. Versatile optical code-based MPLS for circuit, burst, and packet switchings. *Journal of Lightwave Technology*, 21(11):2753–2764, 2003.

[KMFB04] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido. A nonstationary poisson view of Internet traffic. In *Proceedings of IEEE INFOCOM*, 2004.

[KMK04] A. Kumar, D. Manjunath, and J. Kuri. *Communication Networking: an Analytical Approach*. Morgan Kaufmann, 2004.

[KN02] J. Kilpi and I. Norros. Testing the gaussian character of access network traffic. In *Proceedings of the 2nd ACM SIGCOMM Internet Measurement Workshop*, 2002.

[KS98] H. S. Kim and N. B. Shroff. Loss probability calculations and asymptotic analysis for finite buffer multiplexers. *IEEE/ACM Transactions on Networking*, 6(4):411–421, August 1998.

[KS99] E. Knightly and N. Shroff. Admission control for statistical QoS: theory and practice. *IEEE Network*, 13(2):20–29, March 1999.

[KSC91] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis. Weighted round-robin cell multiplexing in a general-purpose ATM switch chip. *IEEE Journal on Selected Areas in Communications*, 9(8):1265–1279, October 1991.

[Küh79] P. Kühn. Approximate analysis of general queuing networks by decomposition. *IEEE Transactions on Communications*, 27(1):113–126, January 1979.

[Küh06a] P. J. Kühn. *Lecture Notes: Communication Network I*. University of Stuttgart, 2006.

[Küh06b] P. J. Kühn. *Lecture Notes: Communication Network II*. University of Stuttgart, 2006.

[Küh06c] P. J. Kühn. *Lecture Notes: Teletraffic Theory and Engineering*. University of Stuttgart, 2006.

[KWS00] K.-I. Kitayama, N. Wada, and H. Sotobayashi. Architectural considerations for photonic IP router based upon optical code correlation. *Journal of Lightwave Technology*, 18(12):1834–1844, 2000.

[LA04] J. Liu and N. Ansari. The impact of the burst assembly interval on the OBS ingress traffic characteristics and system performance. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 3, pages 1559–1563 Vol.3, 2004.

[Lae02] K. Laevens. Traffic characteristics inside optical burst switched networks. In *Proceeding of the SPIE OptiCom*, 2002.

[LB03] K. Laevens and H. Bruneel. Analysis of a single-wavelength optical buffer. In *Proceedings of IEEE INFOCOM*, volume 3, pages 2262–2267, 2003.

[Lee06] S. Lee. Packet-based burst queue modeling at an edge in optical-burst switched networks. *Computer Communications*, 29(5):634 – 641, March 2006.

[LKA+06] H.-C. Leligou, K. Kanonakis, J. Angelopoulos, I. Pountourakis, and T. Orphanoudakis. Efficient burst aggregation for QoS-aware slotted OBS systems. *European Transactions on Telecommunications*, 17(1):93–98, 2006.

[LQXX04] J. Li, C. Qiao, J. Xu, and D. Xu. Maximizing throughput for optical burst switching networks. In *Proceedings of IEEE INFOCOM*, volume 3, pages 1853–1863, 2004.

[LSKS05] S. Lee, K. Sriram, H. Kim, and J. Song. Contention-based limited deflection routing protocol in optical burst-switched networks. *IEEE Journal on Selected Areas in Communications*, 23(8):1596–1611, 2005.

[LSX+03] Z. Liu, M. S. Squillante, C. H. Xia, S.-Z. Yu, and L. Zhang. Profile-based traffic characterization of commercial Web sites. In *Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pages 231–240, Berlin, 2003.

[LT07] H. Li and I. L.-J. Thng. Edge node buffer usage in optical burst switching networks. *Photonic Network Communications*, 13(1):31–51, January 2007.

[MMM+05] L. Muscariello, M. Mellia, M. Meo, M. A. Marsan, and R. L. Cigno. Markov models of Internet traffic and a new hierarchical MMPP model. *Computer Communications*, 28(16):1835–1851, October 2005.

[MV96] M. Montgomery and G. D. Veciana. On the relevance of time scales in performance oriented traffic characterizations. In *Proceedings of IEEE INFOCOM*, volume 2, pages 513–520, March 1996.

[MZA05] Y. Mingwu, L. Zengji, and W. Aijun. Accurate and approximate evaluations of asynchronous tunable-wavelength-converter sharing schemes in optical burst-switched networks. *IEEE/OSA Journal of Ligthwave Technology*, 23(10):2807–2815, 2005.

[Nor95] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal of Selected Areas in Communications*, 13(6), 1995.

[NR01] A. Neukermans and R. Ramaswami. MEMS technology for optical networking applications. *IEEE Communications Magazine*, 39(1):62–69, January 2001.

[NRSV91] I. Norros, J. Roberts, A. Simonian, and J. Virtamo. The superposition of variable bit rate sources in an ATM multiplexer. *IEEE Journal of Selected Areas in Communications*, 9(3), 1991.

[NW98] A. Neidhardt and J. Wang. The concept of relevant time scales and its application to queueing analysis of self-similar trafic. In *Proceedings of SIGMETRICS '98/PERFORMANCE '98*, 1998.

[NZA99] T. D. Neame, M. Zukerman, and R. G. Addie. Modeling broadband traffic streams. In *Proceedings of IEEE Global Telecommunications Conference (Globecom)*, 1999.

[OS05] H. Overby and N. Stol. Providing absolute QoS in asynchronous bufferless optical packet/burst switched networks with the adaptive preemptive drop policy. *Computer Communications*, 28(9):1038–1049, June 2005.

[OSHT01] M. J. O'Mahony, D. Simeonidou, D. K. Hunter, and A. Tzanakaki. The application of optical packet switching in future communication networks. *IEEE Communications Magazine*, 39(3):128–135, March 2001.

[OT05] N. Ogino and H. Tanaka. Deflection Routing for Optical Bursts Considering Possibility of Contention at Downstream Nodes. *IEICE Transactions on Communication*, E88-B(9):3660–3667, 2005.

[PCM+04] M. H. Phung, K. C. Chua, G. Mohan, M. Motani, and T. C. Wong. Absolute QoS signalling and reservation in optical burst-switched networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, Dallas/TX, November 2004.

[PCM+05] M. Phung, K. Chua, G. Mohan, M. Motani, T. Wong, and P. Kong. On ordered scheduling for optical burst switching. *Computer Networks*, 48(6):891–909, August 2005.

[PCM+07] M. Phung, K. Chua, G. Mohan, M. Motani, and D. Wong. An absolute QoS framework for loss guarantees in optical burst-switched networks. *IEEE Transactions on Communications*, 55(6):1191–1201, June 2007.

[Per06] A. Perin. Design and performance evaluation of scheduling mechanisms in the edge nodes of OBS networks. Student thesis, University of Stuttgart, October 2006.

[PF95] V. Paxson and S. Floyd. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3), 1995.

[PKC97] K. Park, G. Kim, and M. E. Crovella. On the effect of traffic self-similarity on network performance. In *Proceedings of SPIE International Conference on Performance and Control of Network Systems*, November 1997.

[PP06] V. S. Puttasubbappa and H. G. Perros. Performance analysis of limited-range wavelength conversion in an OBS switch. *Telecommunication Systems*, 31(2–3):227–246, March 2006.

[PTZD03] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot. Long-term forecasting of Internet backbone traffic: observations and initial models. In *Proceedings of IEEE INFOCOM*, volume 2, pages 1178–1188, 2003.

[PW00] K. Park and W. Willinger. *Self-Similar Network Traffic and Performance Evaluation*, chapter Self-similar network traffic: an overview. John Wiley & Sons, 2000.

[QY99] C. Qiao and M. Yoo. Optical burst switching (OBS)—a new paradigm for an optical Internet. *Journal of High Speed Networks*, 8(1):69–84, January 1999.

[QY00] C. Qiao and M. Yoo. Choices, features and issues in optical burst switching. *Optical Networking Magazine*, 1(2):36–44, April 2000.

[QYD01] C. Qiao, M. Yoo, and S. Dixit. Optical burst switching for service differentiation in the next-generation optical Internet. *IEEE Communications Magazine*, 39(2), 2001.

[Ram02] R. Ramaswami. Optical fiber communication: from transmission to networking. *IEEE Communications Magazine*, 40(5):138–147, May 2002.

[Ric95] J. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, 2nd edition, 1995.

[RLFB05] W. Rogiest, K. Laevens, D. Fiems, and H. Bruneel. A performance model for an asynchronous optical buffer. *Performance Evaluation*, 62(1-4):313–330, October 2005.

[RMV96] J. Roberts, U. Mocci, and J. Virtamo. *Broadband Network Teletraffic – Final Report of Action COST 242*. Springer-Verlag, 1996.

[ROB04] R. Rajaduray, S. Ovadia, and D. Blumenthal. Analysis of an edge router for span-constrained optical burst switched (OBS) networks. *IEEE Journal on Lighwave Technology*, 22(11), 2004.

[RS02] R. Ramaswami and K. N. Sivarajan. *Optical Networks*. Morgan Kaufmann, 2nd edition, 2002.

[RT02] J. Ramamirtham and J. Turner. Design of wavelength converting switches for optical burst switching. In *Proceedings of IEEE INFOCOM*, New York, June 2002.

[RVZW03] Z. Rosberg, H. L. Vu, M. Zukerman, and J. White. Blocking probabilities of optical burst switching networks based on reduced load fixed point approximation. In *Proceedings of IEEE INFOCOM*, volume 3, pages 2008–2018, March 2003.

[RZ03] C. Raffaelli and P. Zaffoni. Packet assembly at optical packet network access and its effects on TCP performance. In *Workshop on High Performance Switching and Routing (HPSR)*, pages 141–146, Torino, June 2003.

[RZVZ06] Z. Rosberg, A. Zalesky, H. L. Vu, and M. Zukerman. Analysis of OBS networks with limited wavelength conversion. *IEEE/ACM Transactions on Networking*, 14(5):1118–1127, October 2006.

[SC99] V. Sivaraman and F. Chiussi. Statistical analysis of delay bound violations at an earliest deadline first (EDF) scheduler. *Performance Evaluation*, 36-37:457–470, 1999.

[SC00] V. Sivaraman and F. Chiussi. Providing end-to-end statistical delay guarantees with earliest deadline first scheduling and per-hop traffic shaping. In *Proceedings of IEEE INFOCOM*, 2000.

[SGV99] J. Sahni, P. Goyal, and H. M. Vin. Scheduling CBR flows: FIFO or per-flow queuing? In *Proceedings of the Ninth IEEE International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Basking Ridge, June 1999.

[She05] Y. Shen. QoS support in edge node for optical burst switching networks. Student thesis, University of Stuttgart, October 2005.

[SNV02] P. Salvador, A. Nogueira, and R. Valadas. Modeling multifractal traffic with stochastic L-systems. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, volume 3, pages 2518–2522, November 2002.

[Sta02] W. Stallings. *High-Speed Networks and Internets: Performance and Quality of Service*. Prentice Hall, 2nd edition, 2002.

[SV96] M. Shreedhar and G. Varghese. Efficient fair queueing using deficit round robin. *IEEE/ACM Transactions on Networking*, 4(3):375–385, June 1996.

[TP00] T. Tuan and K. Park. *Self-Similar Network Traffic and Performance Evaluation*, chapter Congestion control for self-similar network traffic. John Wiley & Sons, 2000.

[Tur99] J. S. Turner. Terabit burst switching. *Journal of High Speed Networks*, 8(1):3–16, January 1999.

[TWJ02] X. Tian, J. Wu, and C. Ji. A unified framework for understanding network traffic using independent wavelet models. In *Proceedings of IEEE INFOCOM*, New York City, June 2002.

[TYC+00] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott. Optical routing of asynchronous, variable length packets. *IEEE Journal of Selected Areas in Communications*, 18(10):2084–2093, October 2000.

[UOS04] A. Undheim, H. Overby, and N. Stol. Absolute QoS in synchronous optical packet switched networks. In *Proceedings of Norsk Informatikk Konferanse (NIK)*, pages 137–148, Stavanger, November 2004.

[VJ03] V. M. Vokkarane and J. P. Jue. Prioritized burst segmentation and composite burst-assembly techniques for QoS support in optical burst-switched networks. *IEEE Journal on Selected Areas in Communications*, 21(7):1198–1209, September 2003.

[VZJC02] V. M. Vokkarane, Q. Zhang, J. P. Jue, and B. Chen. Generalized burst assembly and scheduling techniques for QoS support in optical burst-switched networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, pages 2747–2751, Taipei, November 2002.

[WMA02] X. Wang, H. Morikawa, and T. Aoyama. Burst optical deflection routing protocol for wavelength routing WDM networks. *Optical Networks Magazine*, pages 12–19, November/December 2002.

[WPRT02] W. Willinger, V. Paxson, R. H. Riedi, and M. S. Taqqu. *Long-Range Dependence: Theory and Applications*, chapter Long-range dependence and data network traffic. Birkhauser, 2002.

[WTSW97] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, February 1997.

[WZV02] J. White, M. Zukerman, and H. L. Vu. A framework for optical burst switching network design. *IEEE Communication Letters*, 6(6):268–270, June 2002.

[XQLX04] J. Xu, C. Qiao, J. Li, and G. Xu. Efficient burst scheduling algorithms in optical burst-switched networks using geometric techniques. *IEEE Journal on Selected Areas in Communications*, 22(9):1796–1811, 2004.

[XVC99] Y. Xiong, M. Vandenhoute, and H. Cankaya. Design and analysis of optical burst-switched networks. In *Proceedings of the SPIE Conference on All Optical Networking*, volume 3843, pages 112 – 119, September 1999.

[XVC00] Y. Xiong, M. Vanderhoute, and H. C. Cankaya. Control architecture in optical burst-switched WDM networks. *IEEE Journal of Selected Areas in Communications*, 18(10):1838–1851, October 2000.

[XY02] F. Xue and S. J. B. Yoo. Self-similar traffic shaping at the edge router in optical packet switched networks. In *Proceedings of IEEE International Conference on Communications (ICC)*, 2002.

[Y.1541] ITU. Rec. Y.1541: Network Performance Objectives for IP-Based Services, International Telecommunication Union, ITU-T, 2003.

[YCQ02a] X. Yu, Y. Chen, and C. Qiao. Performance evaluation of optical burst switching with assembled burst traffic input. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, 2002.

[YCQ02b] X. Yu, Y. Chen, and C. Qiao. A study of traffic statistics of assembled burst traffic in optical burst switched networks. In *Proceedings of SPIE OptiCom*, 2002.

[YLC+04] X. Yu, J. Li, X. Cao, Y. Chen, and C. Qiao. Traffic statistics and performance evaluation in optical burst switched networks. *Journal of Lightwave Technology*, 22(12), 2004.

[YLES96] J. Yates, J. Lacey, D. Everitt, and M. Summerfield. Limited-range wavelength translation in all-optical networks. In *Proceedings of IEEE INFOCOM*, pages 954–961, San Francisco, March 1996.

[YQ97] M. Yoo and C. Qiao. Just-enough-time (JET): A high speed protocol for bursty traffic in optical networks. In *Proceedings of the IEEE/LEOS Summer Topical Meetings*, pages 26–27, Montreal, Que. , Canada, August 1997.

[YQ00] M. Yoo and C. Qiao. QoS performance of optical burst switching in IP over WDM networks. *IEEE Journal of Selected Areas in Communications*, 18(10):2062–2071, October 2000.

[YR06] L. Yang and G. Rouskas. A framework for absolute QoS guarantees in optical burst switched networks. In *Proceedings of Broadnets 2006*, San Jose, October 2006.

[YTSC04] M. Yuang, P. Tien, J. Shih, and A. Chen. QoS scheduler/shaper for optical coarse packet switching IP-over-WDM networks. *IEEE Journal on Selected Areas in Communications*, 22(9):1766 1780, November 2004.

[YXM+02] S. Yao, F. Xue, B. Mukherjee, S. J. B. Yoo, and S. Dixit. Electrical ingress buffering and traffic aggregation for optical packet switching and their effect on TCP-level performance in optical mesh networks. *IEEE Communications Magazine*, 40(9):55–72, September 2002.

[YZV01] M. Yang, S. Q. Zheng, and D. Verchere. A QoS supporting scheduling algorithm for optical burst switching DWDM networks. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, San Antonio, November 2001.

[Zal06] A. Zalesky. Optimizing an OBS scheduler buffer. In *Proceedings of the 1st International Conference on Performance Evaluation Methodolgies and Tools*, Pisa, October 2006.

[Zha95] H. Zhang. Service disciplines for guaranteed performance service in packet-switching networks. *Proceedings of the IEEE*, 83(10):1374–1396, October 1995.

[ZLJ05] T. Zhang, K. Lu, and J. Jue. An analytical model for shared fiber-delay line buffers in asynchronous optical packet and burst switches. In *Proceedings of the IEEE International Conference Communications (ICC)*, volume 3, pages 1636–1640, 2005.

[ZRMD03] Z. Zhang, V. J. Ribeiro, S. Moon, and C. Diot. Small-time scaling behaviors of Internet backbone traffic: an empirical study. In *Proceedings of IEEE INFOCOM*, 2003.

[ZVJC04] Q. Zhang, V. Vokkarane, J. Jue, and B. Chen. Absolute QoS differentiation in optical burst-switched networks. *IEEE Journal on Selected Areas in Communications*, 22(9):1781–1795, November 2004.

[ZWZ+04] A. Zalesky, E. Wong, M. Zukerman, H. Vu, and R. Tucker. Performance analysis of an OBS edge router. *IEEE photonics technology letters*, 16(2):695–697, February 2004.