Mechanisms for Quality of Service Differentiation in Optical Burst Switched Networks

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

vorgelegt von

Klaus Dolzer

geboren in Stuttgart

Hauptberichter:	Prof. Dr.-Ing. Dr. h. c. mult. Paul J. Kühn
Mitberichter:	Prof. Dr. Ulrich Killat
Tag der mündlichen Prüfung:	10. Februar 2004

Institut für Kommunikationsnetze und Rechnersysteme
der Universität Stuttgart
2004

To Iris

Perfection is reached, not when there is no longer anything to add, but when there is no longer anything to take away.

Antoine de Saint-Exupery

Complexity is the primary mechanism which impedes efficient scaling, and as a result is the primary driver of increases in both capital expenditures (CAPEX) and operational expenditures (OPEX).

Mike O’Dell, former Chief architect at UUNET

Abstract

Whereas it is widely believed that the Network Layer of the OSI-ISO reference model is the convergence layer with respect to interoperability and service provisioning, an outstanding question remains which layer is the convergence layer with respect to Quality of Service, QoS. Despite an enormous amount of research work on IP-based QoS architectures like, e. g., IntServ and DiffServ, none of these architectures is broadly implemented in today's networks. Furthermore, even in the IP centric community, the comprehension arises that a lower layer may by the convergence layer with respect to QoS.

This trend towards an IP-over-WDM-based transport network architecture is driven by (i) the evolution of the photonic layer, (ii) the need for a higher dynamic in provisioning of bandwidth and (iii) the need for cost reduction in the core. Hereby, enhanced photonic components as well as the ability to transmit photonic signals for thousands of kilometers without regeneration allows to think about photonic networking.

In this thesis, the above mentioned evolution towards an IP-over-WDM-based transport network architecture is discussed and standardization efforts are highlighted. Optical burst switching, OBS, is one promising candidate of such a transport network architecture which realizes a hybrid approach of out of band signalling while data remains in the photonic domain all the time. By doing so, processing of header information can be carried out electronically which performs a decoupling of header processing and data forwarding. However, as the use of buffers is not mandatory for OBS and OBS does not comprise QoS functionality, an OBS-QoS mechanism is required which allows to simply and efficiently differentiate between a number of classes as service to the higher Network Layer. Therefore, a comprehensive overview of the functionality of OBS including burst assembly mechanisms as well as reservation mechanisms is presented and approaches reported in literature to realize OBS-QoS are shortly discussed.

Afterwards, the well-established theory of loss systems is introduced which is the basis for OBS-QoS mechanisms as buffers are not mandatory in OBS networks and wavelength channels can be modelled as servers. The outcome of this review is twofold: it can be stated that (i) reasonable QoS with respect to loss probabilities can only be obtained if the normalized offered traffic is controlled well below 1 and (ii) trunk reservation admission control is a very promising candidate to efficiently realize service differentiation.

Additionally to these requirements for a new OBS-QoS mechanism, further requirements are derived from shortcomings which are the result of an approximative performance analysis of the offset-based OBS-QoS mechanism reported in literature, namely the dependence of the loss probability on traffic characteristic of different classes as well as on the actual length of a burst.

These two sources of requirements lead to Assured Horizon which is a new combined framework for a reservation mechanism and a burst assembly mechanism as well as the communication between them. The framework is introduced here and an approximative performance analysis combined with a simulation study proof the functionality of this new approach which allows to draw the conclusion that QoS can be provided by transport networks as service to the above IP layer and thus the convergence layer with respect to QoS can be lower than the Network Layer.

Contents

Abbreviations

AAL	ATM Adaption Layer
ABR	Available Bitrate
ADM	Add/Drop Multiplexer
ANSI	American Standards Institute
AOTF	Acousto-Optic-Tunable-Filter
ASON	Automatically Switched Optical Network
ATM	Asynchronous Transfer Mode
BA	Behavior Aggregate
BAC	Burst Admission Control
BDP	Burst Drop Priority
BHP	Burst Header Packet
B-ISDN	Broadband - Integrated Services Digital Network
C	Compliant
CapEx	Capital Expenditure
CCITT	Comité Consultatif Internationale de Télégraphie et Téléphonie
CDM	Code Division Multiplexing
CLIP	Classical IP over ATM
CLP	Cell Loss Priority
CoV	Coefficient of Variance
CR-LDP	Constrained-based Routing using Label Distribution Protocol
CWDM	Coarse Wavelength Division Multiplexing
DFDP	Deflect First and Drop Policy
DFSDP	Deflect First, Segment and Drop Policy
DiffServ	Differentiated Services
DXC	Digital Cross Connect
DWDM	Dense Wavelength Division Multiplexing
FCFS	First Come, First Serve
FDL	Fiber Delay Line
FDM	Frequency Division Multiplexing
FEC	Forwarding Equivalent Class
FRP/DT	Fast Reservation Protocol/Delayed Transmission
FRP/IT	Fast Reservation Protocol/Immediate Transmission
Gbps	Gigabit Per Second
GFP	Generic Frameing Procedure
GMPLS	Generalized Multiprotocol Label Switching
GMPLS-LDP	GMPLS-Lable Distribution Protocol
GMPLS-TE	GMPLS-Traffic Engineering

GPRS	General Packet Radio Service
HDLC	High-Level Data Link Control
HSCSD	High-Speed Circuit-Switched Data
IBT	Inband Terminator
IEEE	Institute of Electrical and Electronics Engineers, Inc.
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IBT	Inband Terminator
I/O	Input/Output
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JET	Just Enough Time
JIT	Just In Time
LAN	Local Area Network
LAUC	Latest Available Unscheduled Channel
LAUC-VF	Latest Available Unscheduled Channel With Void Filling
LLC	Link Layer Control
LSP	Label Switched Path
LSR	Label Switched Router
MAC	Media Access Control
MAN	Metropolitan Area Network
MEMS	Micro-Electro-Mechanical Systems
Mbps	Megabit Per Second
MPLS	Multiprotocol Label Switching
MPλS	Multiprotocol Lambda Switching
NC	Non-Compliant
OBS	Optical Burst Switching
OC	Optical Container
OCh	Optical Channel Layer Network
OCS	Optical Circuit Switching
O-E-O	Opto-Electronic-Opto (Conversion)
OFS	Optical Flow Switching
OIF	Optical Internetworking Forum
OLS	Optical Lable Switching
OMS	Optical Multiplex Section Layer Network
OpEx	Operation Expenditure
OPS	Optical Packet Switching
OSI	Open System Interconnection
OTN	Optical Transport Network
OTS	Optical Transmission Section Layer Network
OXC	Optical Cross Connect

PASTA	Poisson Arrivals See Time Averages
PBP	Poisson Burst Process
PDH	Plesiochronous Digital Hierarchy
PNNI	Private Network-Node Interface or Private Network-Network Interface
PPP	Point-to-Point Protocol
QoS	Quality Of Service
RAM	Random Access Memory
RED	Random Early Discard
RFD	Reserve-A-Fixed Duration
RLD	Reserve-A-Limited Duration
RPR	Resilient Packet Rings
RSVP	Resource Reservation Protocol
SAP	Service Access Point
SCDT	Separate Control Delayed Transmission
SCFQ	Self-Clocked Fair Queueing
SDH	Synchronous Digital Hierarchy
SDL	Specification and Description Language
SDM	Space Division Multiplexing
SFDP	Select First and Deflect Policy
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical Network
STM	Synchronous Transport Module
TAG	Tell-And-Go
Tbps	Terabit Per Second
TCP	Transport Control Protocol
TDM	Time Division Multiplexing
TG	Tell & Go
ToS	Type of Service
TSpec	Traffic Specification
TM	Terminal Multiplexer
TR	Trunk Reservation Admission Control
TW	Tell and Wait
UNI	User Network Interface
UMTS	Universal Mobile Telecommunication System
VCC	Virtual Channel Connection (see: ATM)
VC	Virtual Container (see: SDH)
VPN	Virtual Private Network
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
WFQ	Weighted Fair Queueing
WL	Wavelength
WLAN	Wireless LAN
xDSL	x Digital Subscriber Line (e. g. Asymmetric)

Symbols

A_i	offered traffic of class i
$B(A, n) = B$	blocking probability (obtained by $E_{1, n}$)
B_{all}	overall blocking probability
B_i	blocking probability of class i
b_w	wasted bandwidth
c	speed of light
C_i	required number of servers per request of class i
$E_{1, n}$	Erlang loss formula/Erlang B formula
$F_i(u)$	distribution function of class i
F_i^f	distribution function of the forward recurrence time of class i
f_i	allocation factor of FEC i
h	mean transmission time
h_i	mean transmission time of class i
$h_{i, j}$	ratio of mean burst length of class i and j
i, j, k, m	index of, e. g., a service class
l_i	mean burst length of FEC i
$l_i^{(j)}$	estimate of l_i in the j^{th} time interval
$l_{i, C}$	mean length of compliant bursts of FEC i
$l_{i, NC}$	mean length of non-compliant bursts of FEC i
m	overall number of occupied servers
m_i	mean bandwidth of FEC i
N	random variable describing the number of arrivals in a time interval
n	number of service classes (in a loss system)
n_i	number of requests per service class
n_{wl}	number of wavelengths
p	branch probability
p_i	probability of state

$p(y\|n)$	probability of the number of Bytes contained in a file conditioned on number of arrivals n
$p(y, \tau)$	probability of the number of Bytes arriving in a time interval of lengths τ
$\tilde{p}_i(m)$	unnormalized distribution of probabilities of state
$p_i(m)$	normalized distribution of probabilities of state
$\tilde{p}_i^*(m)$	approximation of unnormalized distribution of probabilities of state
$P\{\ \}$	probability
$P_{\text{Drop, C}}$	drop probability of compliant bursts
$P_{\text{Drop, i}}$	drop probability of class i bursts
$P_{\text{Drop, NC}}$	drop probability of non-compliant bursts
$P_{\text{excess}}(\tau)$	probability that a threshold is exceeded within time interval (0, τ)
$P_{\text{Loss, all}}$	overall loss probability
$P_{\text{Loss, C}}$	loss probability of C bursts
$P_{\text{Loss, i}}$	burst loss probability of class i
$P_{\text{Loss, wl}}$	loss probability caused by unsuccessful wavelengths reservation process
$P^{(j)}{}_{\text{Loss, i}}$	j^{th}-order estimate of burst loss probability of class i
$P_{\text{Loss, \{JET, JIT, Horizon\}}}$	loss probability of reservation mechanism {JET, JIT, Horizon}
$P_{\text{Loss, NC}}$	loss probability of NC bursts
$P_{\text{NC}}(\tau)$	probability that an NC burst is generated within time interval (0, τ)
q_i	QoS metric
q_i	class dependent threshold of a trunk reservation mechanism
$q_i(n)$	class- and state dependent threshold of a trunk reservation mechanism
q_{NC}	threshold of a trunk reservation mechanism where NC bursts are dropped
R	overall reward
R_i	reward of class i
R_{max}	maximum overall reward
R_{norm}	normalized overall reward
r	link rate
r_{access}	rate of the access link
r_i	reserved bandwidth envelope of FEC i
S_C	share of C bursts
$S_{C,\,i}$	share of C bursts of class i

S_i	set of system states where losses can occur (if a new request arrives)
S_{NC}	share of NC bursts
$S_{NC, all}$	share of NC bursts of all classes
$S_{NC, i}$	share of NC bursts of class i
s_k	differentiation factor of class k
$t_{e2e,i}$	end to end delay of class i
w	number of wavelengths
w_a	number of currently allocated wavelengths
w_c	number of allocated wavelengths where the congestion state starts
$w_{max, i}$	maximum waiting time in the assembly buffer of class i
x	one-way distance
X	random variable describing the number of Bytes contained in a file
X_i	random variable describing the number of Bytes contained in file i
y	number of Bytes
Y_i	carried traffic of class i
$Y(\tau)$	random variable describing the number of arriving Bytes to the assembly buffer within time interval τ

α	parameter of the Pareto distribution
$\Delta_{i, j}$	effective offset difference between class i and class j
δ, Δ	offset between header and data
δ_{basic}	basic offset
δ_{QoS}	QoS offset
δ_i	offset between header and data of class i
λ	arrival rate
λ_i	arrival rate of class i
$\lambda_{max, i}$	maximum arrival rate
μ	termination rate
μ_i	termination rate of class i
θ_i	highest occupancy in a systen with trunk reservation access control where a request of class i can be accepted

θ_{NC}	highest occupancy in a systen with trunk reservation access control where an NC request can be accepted
σ_i	threshold in an assembly buffer of class i whose excess causes the generation of an NC burst
τ	observation time
τ_{const}	normalization constant
τ_i	timeout interval of class i
τ_{iat}	interarrival time

Chapter 1

Introduction

The evolution of our information society is heading towards a new era where it is usual to assume that any information is available in any place at any time. The major two drivers for this new information era are (i) ubiquitous broadband access to the information and (ii) new sophisticated, bandwidth-hungry services. One of the major difficulties of the final step towards this new era is that these two drivers mutually rely on each other. Without new bandwidth demanding services, it is not economically meaningful to invest in broadband access networks whereas without broadband access, new sophisticated services cannot be realized. Currently, we are about to enter the new era which is indicated by the evolutions identified in the following.

1.1 Ubiquitous Broadband Access

An evolution towards broadband access is visible in fixed access networks as well as in mobile access networks.

Fixed broadband access to the transport network infrastructure becomes reality for a steadily growing number of Internet users who are connected via xDSL, Digital Subscriber Line, e. g., Asymmetric. Such an access network provides a user with up to 768 kbps in the downlink in case of ADSL and newer products even offer twice the link speed. In commercial environments, the link rate in local area networks, LANs, is quickly increased driven by the progress of, e. g. Ethernet technology. Such technologies reach access speeds of 1 Gbps and an increase of the factor of ten is already worked on.

Broadband access in a mobile scenario starts today in the 2.5th generation of mobile systems with High-Speed Circuit-Switched Data, HSCSD, where up to four channels can be bundled reaching an access speed of 57,6 kbps or General Packet Radio Service, GPRS, where an access speed of up to 171.2 kbps is reached realized by an overlaid packet based air interface.

Real mobile broadband access starts with the third generation of mobile networks like universal mobile telecommunications system, UMTS, where an access link speed up to 2 Mbps can be reached by an individual user. As a next step, research in the context of fourth generation mobile networks focuses on heterogeneous access technologies also including wireless LANs, WLANs. Such access networks may provide users at certain so-called hot-spots like airports, train stations and hotels but also in department stores or in cafes with broadband access with multiples of the link speed of UMTS.

1.2 New Sophisticated Services

For the past decade(s), video-on-demand has been considered as the major ‘killer application‘ requiring enough bandwidth to justify the deployment of broadband access. Currently, a variety of new services is evolving in all areas of life. Hereunto, not only extensive data base queries and updates can be considered. Instead, many fields beyond the usual telecommunication applications explore the added value of broadband transmission to their products or to their daily working-day where employees can work remotely at a customers place or from their home offices.

One example for the added value of a product is the deployment of telematics in the automobile which can turn a car into a moving office or entertainment place. Furthermore, services like dynamic navigation systems which do not only tell the driver the best way but also consider, e. g., road congestions or weather conditions in their proposed route.

Another example is from the field of medicine where a remote specialist is electronically connected to a patient who is in a treatment or even in a surgery. As the remote specialist has the same information as his local colleague, he can advice him or even carry out some part of the surgery. Even more stringent requirements but also more benefit yields a scenario where the patient is moving, e. g., in an ambulance.

Last but not least, the vast majority of sophisticated new bandwidth-hungry services belong to the sector of (home-) entertainment. Applications like video telephony, online gaming or sending of photos and short videos are only the tip of the iceberg and unforeseeable new services may arise within a short period of time.

1.3 Consequences for the Transport Network Architecture

The consequences of the above discussed evolutions on the transport network architecture are three-fold. Firstly, a very great amount of bandwidth has to be provided by the transport network to applications in order to also satisfy tomorrow’s needs. This requirement leads directly to photonic technology which – driven by the break-through in wavelength division multiplexing – can transport an enormous amount of information over a single fiber.

Secondly, the requirement of ubiquitous broadband access prohibits a static allocation of bandwidth or a configuration which cannot be changed on demand within a very short interval in time. In this context, the evolution of photonic components and the milestones in ultra-long haul transmission allow to think about photonic networking and hence, a more dynamic photonic layer. Thirdly, the diversity of applications and thus also their manifold requirements to the transport network architecture calls for QoS differentiation capabilities.

From an economic perspective, a further requirement is a simple and cost-efficient solution with respect to capital expenditure, CapEx, as well as operation expenditure, OpEx, which gets intensified by the facts that revenue is growing slower than the costs for deploying increased bandwidth, and also because the willingness of users to pay much money for such services is limited.

All these consequences for the transport network architecture have to be seen under the boundary condition that the Internet and thus the Internet Protocol IP are currently the basis for most of the information exchange all over the world. Therefore, IP is widely seen as the convergence layer with respect to transmission, also for a future Internet architecture which is controlled by IP and makes use of the enormous amount of bandwidth provided by photonic networks. However, in anticipation of the following discussion, it should be already mentioned here that IP may not be the convergence layer with respect to QoS, instead lower layers should have QoS differentiation capabilities.

Summarizing, a new IP-controlled transport network architecture is required which provides a great amount of bandwidth, is highly dynamic and provides QoS support. Furthermore, this architecture should be simple and cost-efficient.

1.4 Organization of this Dissertation

The remainder of this dissertation is organized as follows: Chapter 2 provides a comprehensive introduction to transport network architectures including their classification in reference models. Furthermore, the evolution towards an IP-over-WDM-based transport network architecture is revealed. In this context, current trends in standardization from different organizations are discussed. Finally, in the last part of Chapter 2, currently discussed transport network architectures like optical circuit switching, optical packet switching and optical burst switching are briefly presented.

Chapter 3 focuses on optical burst switching, OBS. Hereby, a detailed overview of OBS is presented and burst assembly mechanisms as well as reservation mechanisms are classified and discussed. Finally, mechanisms for QoS support in OBS networks reported in literature are complied.

As the use of buffers in the core is not mandatory in OBS, the theory of loss systems is important to get a deeper understanding of OBS-QoS mechanisms. Therefore, teletraffic fundamentals on loss systems are presented in Chapter 4. In the first part, the M/G/n loss system is briefly discussed as representative of a one-class system whereas the second part focuses on multi-class loss systems.

Chapter 5 presents the modelling and performance evaluation of OBS-QoS mechanisms. The first part compares one-class reservation mechanisms with respect to their performance. The second part focuses on the performance evaluation of the offset-based OBS-QoS mechanism as it is the first and most important OBS-QoS mechanism reported in literature. Herefore, an approximative analysis of the burst loss probability is presented and a performance evaluation of different performance metrics by analysis as well as simulation is carried out dependent on various system and traffic parameters.

Based on shortcomings of the offset-based OBS-QoS mechanism revealed in Chapter 5, in Chapter 6 a new OBS-QoS framework called Assured Horizon is introduced which is based on the theory of multi-class loss systems. Chapter 6 starts with an overview on Assured Horizon including design goals and major new contributions. Then, building blocks of Assured Horizon are introduced in detail, namely its bandwidth reservation mechanism, its burst assembly mechanism and its burst reservation mechanism.

Finally, in Chapter 7, a performance evaluation of Assured Horizon is presented. Herefore, an approximative analysis of the burst loss probability is derived followed by detailed studies dependent on system and traffic parameters which are carried out based on the analysis as well as simulations.

In Appendix A, the view of the one-node case is extended to networks. Here, unpleasant effects of the offset-based OBS-QoS mechanism are revealed and it is shown how the results obtained by the performance evaluation of the Assured Horizon framework in a one-node scenario can be extended to a networking scenario.

Chapter 2

Towards an IP-over-WDM Transport Network Architecture

This chapter starts with an overview of the two reference models for network layering that are standardized from the telecommunications world and the data world, respectively, in order to classify the following approaches for IP-over-photonic network architectures. The focus of this thesis is hereby on the adaption of the Internet layer which is described in Section 2.2 and photonic transport network architectures which are discussed in Section 2.3.

In Section 2.3 current standards of optical transport networks as well as the implemented reality are discussed. Because of disadvantages of current implementations with respect to flexibility, dynamics and capital expenditure, CapEx, as well as operation expenditure, OpEx, a variety of approaches to improve and simplify current transport network architectures are introduced and discussed. This evolution towards an IP-over-photonic transport network architecture and related standardization efforts are described in Section 2.4. Section 2.5 introduces currently discussed IP-over-photonic network architectures and classifies them with respect to the reserved granularity of information and holding time of the reservation.

2.1 Reference Models

2.1.1 The OSI Reference Model

The Open System Interconnection, OSI, reference model – which is depicted in Figure 2.1a – was standardized from the International Organization for Standardization, ISO [72], and later also accepted by the International Telecommunication Union - Telecommunication Standardization Sector, ITU-T, recommendation X.200 [82]. Its standardization was carried out with the main goal as basis for the development of system architectures and to simplify the develop-

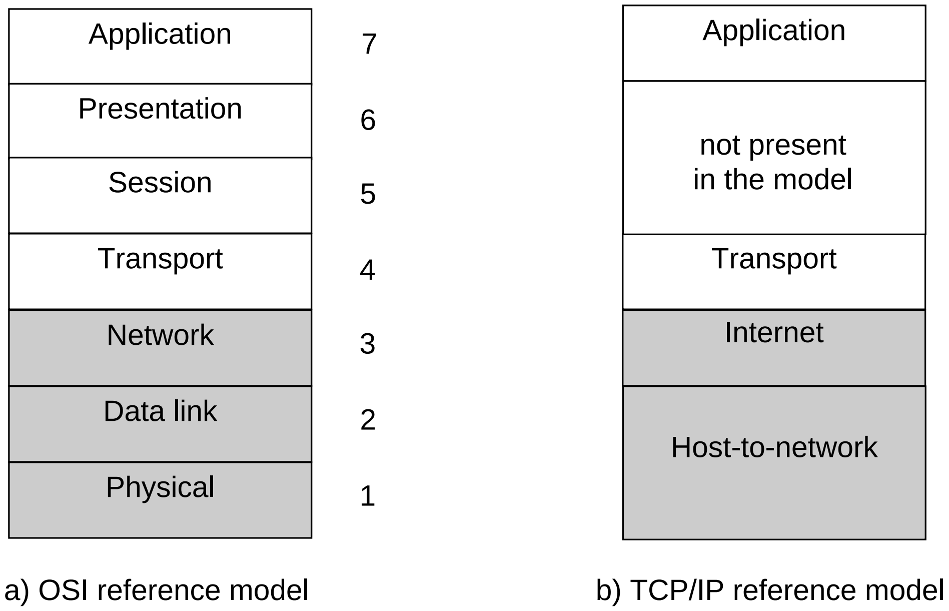


Figure 2.1: OSI and TCP/IP reference models

ment of protocols for telecommunication systems. One of its main achievements is the distinction between services, interfaces and protocols [131] [90]. A service describes the functionality of a lower layer which is offered to an adjacent higher layer via the interface (service access point, SAP) between them. The way this service is realized is defined in protocols which are transparent for higher layers. The communication through an interface between two adjacent layers was designed to be low in order to simplify the development. This concept of separation of functionality and thin interfaces allows for easy interchangeability of protocols within a layer without affecting higher layer protocols.

Figure 2.1a only depicts the data plane of the OSI reference model. The control plane as well as the management plane are not depicted and will not be discussed as they are not in the focus of this thesis. Instead, the focus of this thesis is limited to the lower three layers of the data plane, namely: physical layer, data link layer and network layer. The service of the physical layer is the transmission of raw bits. Hereby, it assures that a bit is received correctly. The data link layer enhances this service by offering integrity of data which comprises sequence integrity, data flow control, detection of transmission errors, error control as well as acknowledgement. This is achieved by introducing data frames. The service provided by the network layer can be summarized by interconnecting subnetworks in order to establish connectivity between end systems. For this purpose, an addressing scheme and routing functionality is contained in this layer, see also [131].

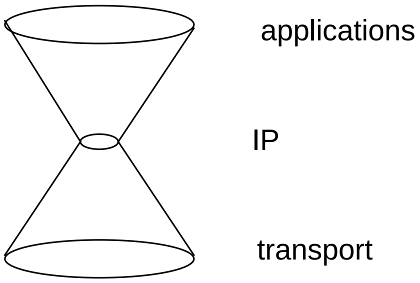


Figure 2.2: Network/Internet Layer as convergence layer

2.1.2 TCP/IP Reference Model

In contrast to the OSI reference model, the TCP/IP reference model does not thoroughly differentiate between services, interfaces and protocols. It was first described in [30] after the development of the protocols and is thus a description of a running system. Its origin is the ARPANET, the predecessor of today’s Internet. Accordingly, its roots and also its orientation come from the classical data world in contrast to the OSI reference model whose origin is the telecommunication world. Figure 2.1.b depicts this reference model which only consists of four layers. Comparable to the just discussed OSI reference model, the focus of this thesis lies on the lower two layers, namely Host-to-network and Internet. The Host-to-network layer is rarely defined and is just supposed to carry IP packets to its destination [131]. The Internet layer roughly corresponds to the Network layer of the OSI reference model and also defines routing and an addressing scheme. However, a special protocol, namely the Internet Protocol, IP, is defined as packet format.

2.2 The Network/Internet Layer

The Network/Internet Layer is widely seen as the convergence layer for all types of higher layer traffic as well as all types of transport networks. Figure 2.2 depicts this circumstance with a hour glass where the broad top corresponds to applications, the narrow middle to the IP layer and the broad bottom to transport networks. This view is mainly driven by the great success of the Internet, the vast distribution of IP to end systems throughout the world and the resulting intention to carry everything over IP. As this evolution is very important also for the rest of this thesis, the main characteristics of the IP layer are shortly reviewed in the following subchapters. For a comprehensive overview, see also [131] and [18].

2.2.1 Functionality

The main functionality of the Network/Internet Layer is the interconnection of subnets in order to provide connectivity between end systems. Therefore, an addressing scheme and routing functionality are contained in this layer.

This functionality does not differentiate between packets originated from applications with different requirements with respect to packet delay, jitter as well as loss probability. Therefore, research effort was focused on changing the Network/Internet Layer accordingly, as described in the following section.

2.2.2 QoS Architectures

Although every IP header contains a field which is denoted with type of service, ToS, this field is hardly used for QoS differentiation. In the core network, all devices operate according to the best effort principle and consequently do not differentiate between classes. With the need for a service integrating network to offer differentiated QoS support for different applications, the research community started work on architectures which allow for service differentiation. The major approaches are listed in the following.

All architectures have in common that packets of different service classes are separated either on different paths or in different queues. In case of different queues, a scheduler is responsible to merge traffic which is isolated from each other. Dependent on the complexity of a scheduling algorithm, a different grade of service differentiation can be obtained. For a comprehensive overview of scheduling algorithms, see [155].

IntServ

Integrated Services, IntServ [23], is an architecture which provides flow-based service guarantees based on reservation of bandwidth following the ATM architecture (Section 2.3.2). A resource reservation protocol, RSVP [24], is defined which reserves bandwidth for a flow according to a traffic specification, TSpec. Each router between source and destination remembers a state per flow in order to be able to guarantee the reserved resources. For robustness, RSVP is a soft-state protocol, i. e., states age out after a certain time if they are not refreshed intermediately. As the number of states may become very large in the core router and (refreshed) per-flow signalling requires a great amount of processing power, it became common sense in the research community that this approach does not scale in large backbone networks.

DiffServ

As an answer to scaling problems of IntServ, Differentiated Services, DiffServ [14], was standardized which allows to differentiate between applications. In order to stay scalable, the concept of flow-based QoS is abandoned and substituted by traffic engineering for traffic aggregates. Hereby, all packets with the same behavior aggregate, BA, are forwarded to the next hop without differentiation. Thus, traffic forwarding works on a hop-by-hop basis. The obtained scaling capabilities come at the price of lack of guarantees for flows. Instead, only relative behavior between BAs are provided.

MPLS

In the context considered here, the main characteristic of MPLS is the separation of forwarding and routing which is carried out by the same device in IP. This separation allows to carry out classification and hence a routing decision once at the network edge and forward all packets along (pre-) calculated paths towards the destination according to an additional label in the header. Routing can be constrained-based and hence be the basis for traffic engineering. Thus, in this context, QoS-support is realized by separation of traffic, which is a different approach compared to IntServ and DiffServ. A more detailed description of MPLS is presented in Section 2.4.2.

2.2.3 Open Questions

Despite an enormous amount of research work during the last years, none of these IP-based QoS architectures are broadly applied in today’s networks. Hence, QoS differentiation is not realized in the Network/Internet layer. Instead, more and more people – even in the IP centric community IETF – believe that IP may not be the convergence layer with respect to QoS, see, e. g., [27]. Consequently, the task of a lower layer has to be extended to additionally provide service differentiation to higher layers. With respect to the transport of IP, the questions need to be answered:

- How can an efficient transport and control of IP be realized?
- Can lower layer(s) provide QoS differentiation as a service to the Network Layer?

Answers to these questions can be found in Chapter 3 and Chapter 6 which are aiming to suggest an environment for efficient QoS support.

2.3 Transport Network Architectures

Whereas the viewpoint of the previous section is on the overall functionality of a network, this section focuses on the transport functionality and therefore describes a network from the viewpoint of information transfer capabilities. This is especially important as the focus of this thesis is to efficiently carry IP packets over a photonic broadband transport network. Section 2.3.1 starts with a general, topology independent description of transport networks. Section 2.3.2 - Section 2.3.4 introduce transport network architectures on different layers in the network and, finally, Section 2.3.5 presents a snapshot on the implemented reality in today's transport networks.

2.3.1 General Transport Network Functionality

Transport network models describe the network functionality from the viewpoint of the information transfer capabilities. In ITU-T recommendation G.805 [78], a description of a generic – i. e., technology independent – transport network is defined. Its functionality can be decomposed in different so-called layer networks which are independent of each other. Each layer network offers a service to its adjacent higher layer which is provided via the interface between these layers. Nevertheless, these layers should neither be confused with layers of the OSI reference model nor the TCP/IP reference model introduced in Section 2.1.1 and Section 2.1.2, respectively. An OSI layer (TCP/IP layer) offers a specific service using one protocol among different protocols. On the contrary, each layer network offers the same service using a specific protocol [78].

2.3.2 Asynchronous Transfer Mode

The ITU-T (formerly Comité Consultatif Internationale de Télégraphie et Téléphonie, CCITT) started its work on standardizing the broadband integrated systems digital network, B-ISDN, in the 1980's with the standard I.121 [81] which was followed by a large number of further standards, also by different standardization bodies like, e. g., the ATMForum [158].

Asynchronous transfer mode, ATM, was designed as transport mechanism for future broadband networks including transmission, multiplexing as well as switching techniques. Its transport is based on fixed-sized cells with the length of 53 Bytes containing a 5 Byte cell header. As interface to higher layers, ATM specifies an adaption layer (ATM adaption layer, AAL) which offers 5 different services with respect to QoS. Connection-oriented communication is realized by a hierarchical concept of virtual connectivity, namely virtual connections at the lower level hierarchy and virtual path for higher level connectivity. The QoS support is based on asynchronous time division multiplexing and is achieved by reservation of some amount of bandwidth between mean and peak bandwidth, also called effective bandwidth [87] to virtual channel connections which offer protection from background traffic while allowing for multiplex gain. By doing so, ATM provides statistical bounds for cell loss probability, cell transfer delay as well as jitter and thus offers flexible QoS-support by a virtual link bandwidth management mechanism. For a comprehensive overview of traffic control in ATM, see, e. g. [89].

Besides the described functionality in the data plane, ATM also specifies rich control and management plane functionality [84] whose description is out of the scope of this thesis. Dependent on the functionality which is considered, ATM can be classified to be a layer 2 or a layer 3 transport network architecture in the context of the OSI reference model.

2.3.3 Synchronous Digital Hierarchy/ Synchronous Optical Network

In contrast to ATM, the synchronous digital hierarchy, SDH, and synchronous optical network, SONET, [9] are based on synchronous transport. SONET was standardized by the American National Standards Institute, ANSI [6], and is mainly applied in North America. SDH, which follows SONET, was standardized by ITU-T G.803 [77] and is mainly applied in Europe and Japan. As the two standards are quite similar, the notations of SDH will be used throughout the rest of this thesis.

SDH standardizes hierarchical time multiplexing functionality as well as network availability enhancement techniques like protection and restoration for optical transport networks and replaces the previous standard plesiocronous digital hierarchy, PDH, [73]. SDH offers transport modules to carry signals of different speed. Therefore, SDH introduces a hierarchy of so-called virtual containers, VCs, which can be multiplexed into synchronous transport modules, STMs. Its basic bitrate of 155 Mbps is carried by the synchronous transport module STM-1. Higher transmission speeds are achieved by byte-wise multiplexing of 4 bytes in order to obtain a fourfold of the STM-1. Thus, the transport modules in Table 2.1 are also defined. In this table, OC denotes optical container and is the abbreviation applied in SONET.

link speed	SDH	SONET
34 Mbps	-	OC 1
155 Mbps	STM 1	OC 3
622 Mbps	STM 4	OC 12
2.5 Gbps	STM 16	OC 48
10 Gbps	STM 64	OC 192
40 Gbps	STM 256	OC 768

Table 2.1: Container for SDH and SONET

In order to also provide networking functionality, the elements terminal multiplexer, TM, add/drop multiplexer, ADM, and digital cross connect, DXC, are defined. The task of a TM is to multiplex low speed signals into an STM-1 frame. The ADM defines the functionality of adding and dropping channels, VCs, and STMs of different speed. Finally, the DXC determines switching of channels between input and output ports of an optical node.

Besides the above discussed functionality concerning transmission and networking, the strengths of SDH are its rich network availability enhancement techniques like protection and restoration. The discussion of those is beyond the scope of this thesis.

With respect to lower layers, SDH offers synchronous transport of different granularities over one wavelength channel. If the underlying optical transport network, OTN, provides more than one wavelength per fiber (see Section 2.3.4.1), one SDH interface per wavelength channel is required. This also implies that wavelength division multiplexing, WDM, channels are operated as separated (wavelength) channels and not as shared resources, see also Chapter 3. Furthermore, with SDH, a system is operated in the so-called opaque mode where an opto-electronic-opto, O-E-O, conversion has to be carried out at every hop. Like ATM, SDH cannot be classified clearly to a layer of the OSI reference model. However, most of its functionality corresponds to classical layer 1.

2.3.4 Optical Transport Network

Prior to the discussion of the layering within an OTN in Section 2.3.4.2, an overview of multiplexing techniques for OTNs is presented in Section 2.3.4.1 as these techniques have gained an increasing impact on the deployment of transport network architectures.

2.3.4.1 Multiplexing Techniques

Multiplexing techniques that can be applied in optical transport networks are space division multiplexing, SDM, time division multiplexing, TDM, code division multiplexing, CDM, and wavelength division multiplexing, WDM. Additionally, a duplex mode defines the simultaneous transmission in both directions. These multiplexing techniques can also be combined in order to further increase the data rate.

SDM denotes a multiplexing scheme, where signals are carried on different fibers. The higher the desired bandwidth, the greater the number of required fibers. With TDM, the link rate is subdivided in order to interleave a greater number of slots which carry signals of lower data rate. Hereby, one can distinguish between synchronous and asynchronous TDM depending on the allocation of slots. CDM uses orthogonal codes for the coding of signals which can be transmitted at the same time. However, this multiplexing scheme is not (yet) applied in (optical) transport networks.

Finally, WDM is a multiplexing technique which is a subset of frequency division multiplex, FDM, which has been successfully applied in radio systems for many years. At WDM, carrier frequencies are in the optical domain. Both, FDM and WDM, are based on the principle that waves with different wavelength do not interfere (in first order). This can be exploited by simultaneously transmitting several signals with different wavelengths over the same fiber. If a

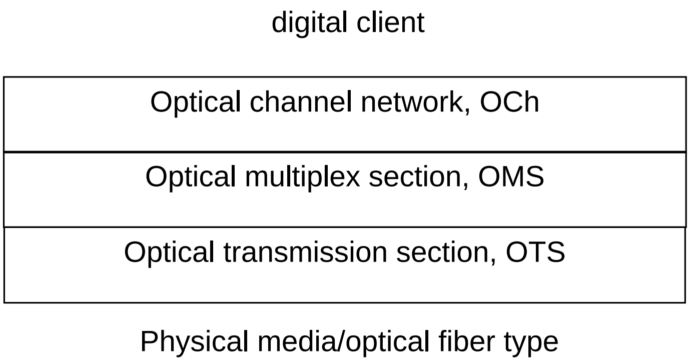

digital client
Optical channel network, OCh
Optical multiplex section, OMS
Optical transmission section, OTS
Physical media/optical fiber type

Figure 2.3: Optical transport network layer structure according to G.872

dense optical carrier frequency is used, WDM is called dense wavelength division multiplexing, DWDM [22], and if the carrier frequency is coarse, it is called coarse WDM, CWDM. Throughout the rest of this thesis, only the term WDM will be used, without respect to the spacing of the carrier frequency.

Often, the optical layer applying WDM is called WDM layer, especially in the context of IP-over-WDM. Although this is confusing – WDM is introduced here as multiplexing technique – the expression WDM and photonic will be used concurrently throughout the rest of this thesis as it is also common in the research community.

In real networks, only SDM, TDM and WDM play a role. The advantage of WDM compared to SDM and TDM is its cost effectiveness: SDM requires more fibers and thus more resources whereas TDM entails more expensive technology which is required to compensate disturbing effects caused by higher transmission speed.

Today’s equipment offers up to 160 wavelength channels which are operated at a speed of 10 Gbps. However, announcements already talk of 300 and more wavelengths which are operated up to 40 Gbps each, resulting in an overall rate per fiber of 12 Tbps [159]. Experiments indicate, that the increase in bandwidth through a higher number of wavelength seems to continue without limits. In [61], an experiment of a transmission of 1022 wavelength channels using a single laser source is referenced.

Caused by the just described increase of bandwidth, WDM-based transmission gains increasing importance in OTNs and is a major driver for changes to the network architecture which will be discussed in Section 2.4.

2.3.4.2 Layering of Optical Transport Network

In ITU-T recommendation G.872 [80], a model for an optical transport network is defined. This standard is technology dependent and focuses on optical networks which apply wavelength division multiplexing, WDM, see also Section 2.3.4.1, as multiplexing technology.

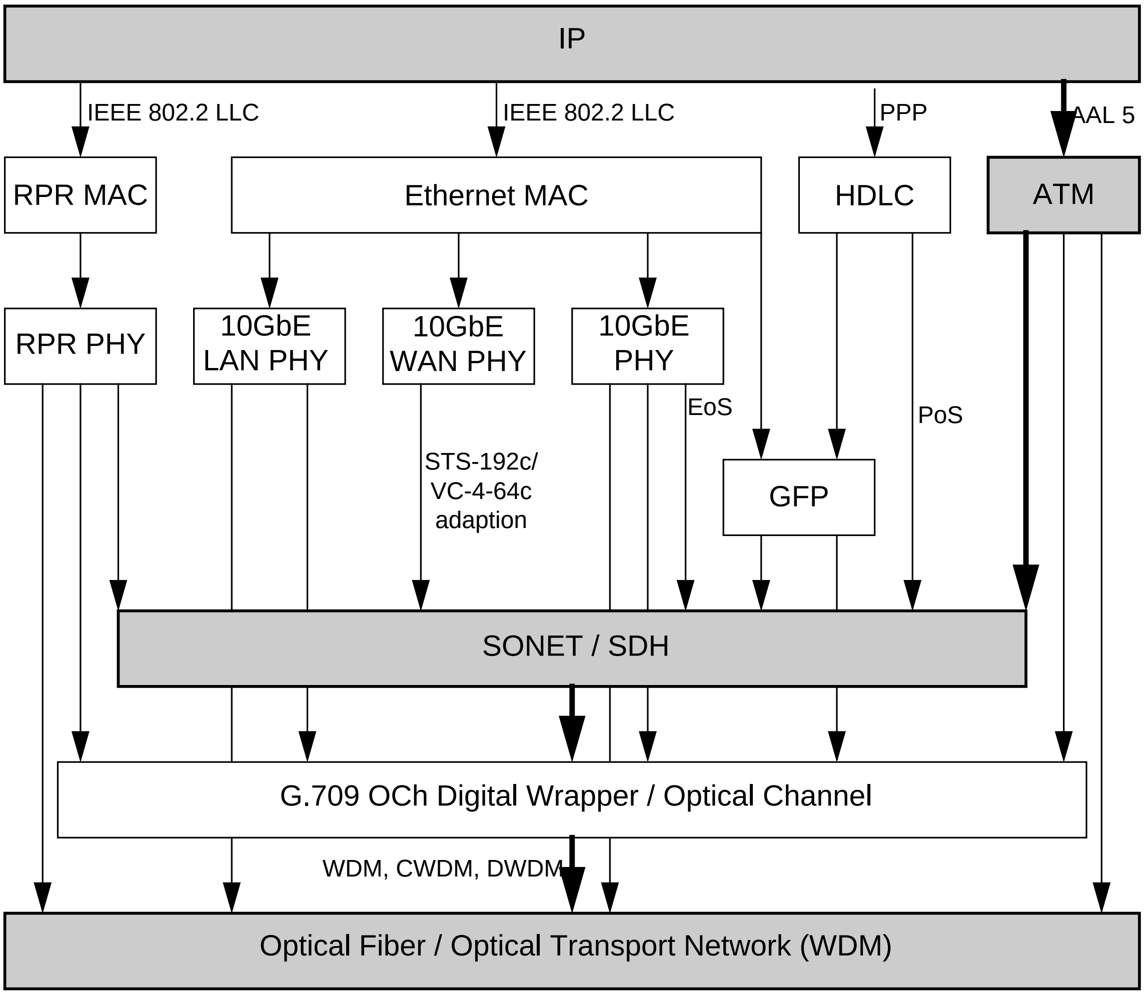


Figure 2.4: Possible transport of IP packet over an OTN [19]

G.872 defines three layer networks which are also depicted in Figure 2.3. Below these layer networks, a defined optical fiber type – which is not further described in this recommendation – carries optical signals. Above, a digital client like, e. g., SDH or ATM uses services provided by the optical transport network.

The functionality of the three layered networks is as follows, see also Figure 2.3.

- **The lowest layer is the optical transmission section layer network, OTS**. Its functionality deals mainly with transmission of optical signals on fibers. More precise, this layered network ensures integrity of optical transmission section as well as section level operation and management functionality.
- **The second layer is called optical multiplex section layer network, OMS**. Its functionality focuses on insurance of integrity as well as operation and management of a multi-wavelengths signal. Hereby, the special case that just one optical channel is contained in the multi-wavelength signal is also considered.

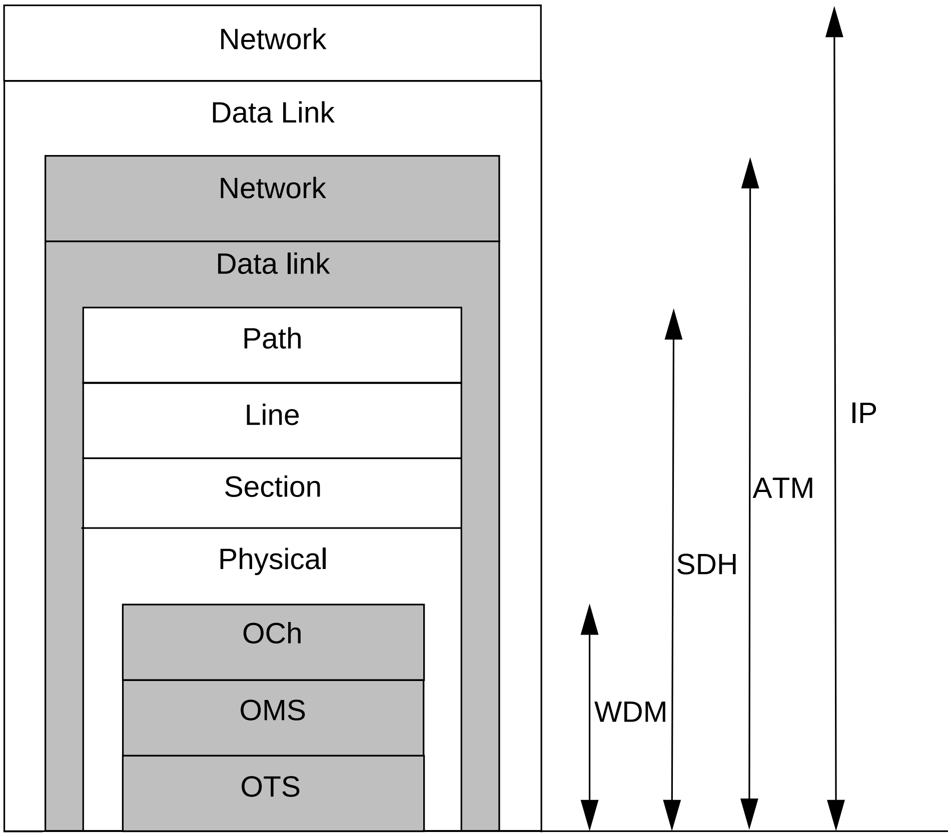


Figure 2.5: Layered view of IP/ATM/SDH/WDM network model

- **The highest layer is called optical channel network layer network, OCh**. Its functionality comprises network-wide insurance of integrity as well as operation and management of a single wavelength channel. Furthermore, functionalities for flexible switching of single wavelength channels are included. Routing, monitoring, grooming, protection and restoration [7] are also carried out in this layer network and make it therefore the most important one.

Concluding the above description, functionalities of the optical transport network does not only comprise definitions made for the physical and data link layer of the OSI reference model, but also some networking functionality. Besides this definition [121] and [127] give other definitions which do not exactly match the layering proposed in [80].

2.3.5 The Implemented Reality

2.3.5.1 Protocol Stacks

The OSI recommendation for layering of transport networks presented in Section 2.3.1 is only partly conforming with today's reality where the evolution of standardized functionality of different protocols led to multiple layered protocol stacks. As IP packets cannot directly be trans-

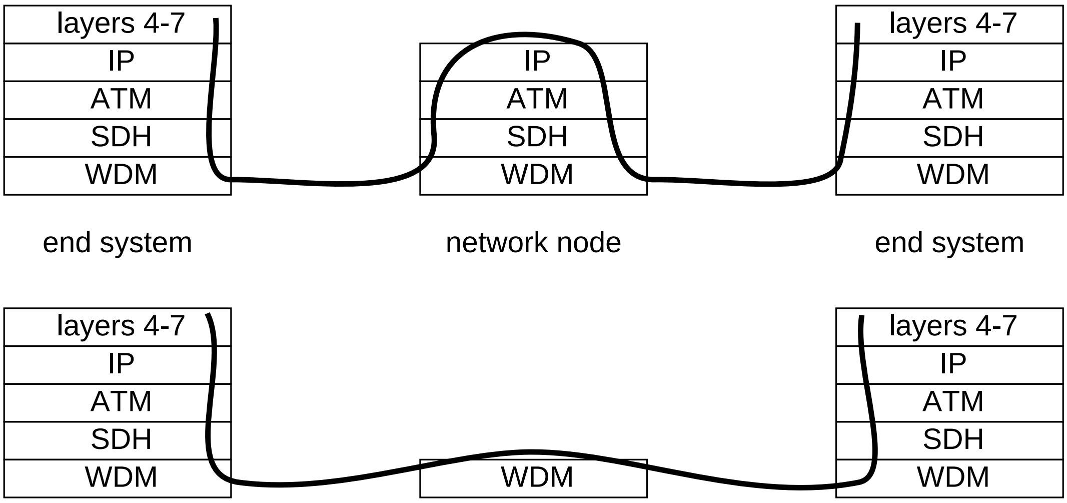


Figure 2.6: Networking modes. top: opaque, bottom: transparent

ported over an OTN, a variety of different protocols are required to carry out functionalities like framing and synchronization between IP and the OTN. Figure 2.4 [19] shows a variety of possibilities to perform IP over a fiber. A comprehensive discussion of Figure 2.4 can be found in [19] or, e. g., in [146]. In the following, only the emphasized parts of Figure 2.4 which are indicated as darker blocks and thicker arrows are discussed as this layering is typically applied in today's WANs. Figure 2.5 depicts these protocol stacks and splits each layer in its sublayers [131].

As can be seen from Figure 2.5, the protocol stacks of IP, ATM, SDH and WDM are layered on top of each other. Hereby, IP, ATM and SDH treat their adjacent lower layer as link layer. Thus, from IP point of view, ATM is its link layer. However, ATM considers SDH as its link layer and SDH treats WDM as its physical layer. From this figure, it can be immediately seen that this layering results in a high complexity which will be further discussed in Section 2.4.

2.3.5.2 Networking Modes

With respect to networking, fibers are only used as static point-to-point connections between electronic routers. In every node, the optical signal is converted to an electronic signal where functionalities like routing decisions are carried out in the Network Layer by IP. Finally, the electrical signal is again converted to an optical signal which is transported to the next hop.

Thus, at every node, a O-E-O conversion has to be carried out which is not only costly with respect to required resources but also with respect to transfer time [109]. This networking mode is called opaque.

In contrast to opaque networking, transparent networking denotes a mode where data can stay in the optical domain and no O-E-O conversion is required at every hop. However, this requires enhanced functionality of the OTN like, e. g., switching capability. In Figure 2.6, the opaque mode is depicted in the top whereas the transparent mode is depicted at the bottom. From this figure, it can be seen that the opaque mode requires all protocol stacks up to the IP layer in all intermediate network nodes whereas transparent node does not. Also based on this discussion, this thesis intends to yield an architecture for an OTN which works in the transparent mode.

2.4 Evolution and Standardization Towards IP-over-WDM

2.4.1 Data Plane

Some advantages of layering of protocol stacks which have evolved over the years are that these solutions greatly fulfill today's traffic requirements. Furthermore, the slowdown of economy entails a growth of network traffic which is much slower than forecasted previously. Hence, today's installed networks will also be able to satisfy tomorrow's needs. This is especially true as a great number of fibers in the field are not used (also called dark fibers) or are operated at a very low carried load (below 10%).

However, data traffic volume still grows at high rate [160] and broadband access like digital subscriber line, xDSL or wireless LAN, WLAN, (driven by evolutions towards 4th generation mobile networks) facilitate new applications with greater bandwidth demand. The below listed drawbacks of current transport networks motivate the effort to make the optical layer more dynamic and thereby enhance its functionality.

- **Increased bandwidth provided by the WDM layer**
  One of the main drivers towards transparent all-optical networking is the enormous growth of bandwidth provided by the optical layer. With high data rates on one wavelength combined with a large number of wavelengths which can be transported by one fiber, bitrates far beyond 1 Tbps can be achieved on one fiber. This is a link speed where increasing effort is required for an electronic router, especially if the transported packets are very short [83].
  Besides hardware (routers) limitations, ATM was not designed for carrying very high bitrates of 40 Gbps and beyond.

- **High costs**
  A non-technical but probably most important reason for the evolution towards IP-over-WDM is cost reduction of the transport system. This includes CapEx as well as OpEx. The opaque solution, where the optical layer only provides a large amount of bandwidth between routers requires a costly O-E-O conversion in every router. Hereunto, all wavelengths have to be electronically terminated by an SDH, an ATM as well as an IP interface each which becomes very costly for a large number of wavelengths.
  A further cause of high costs is the overhead which is introduced by the ATM and SDH pro-

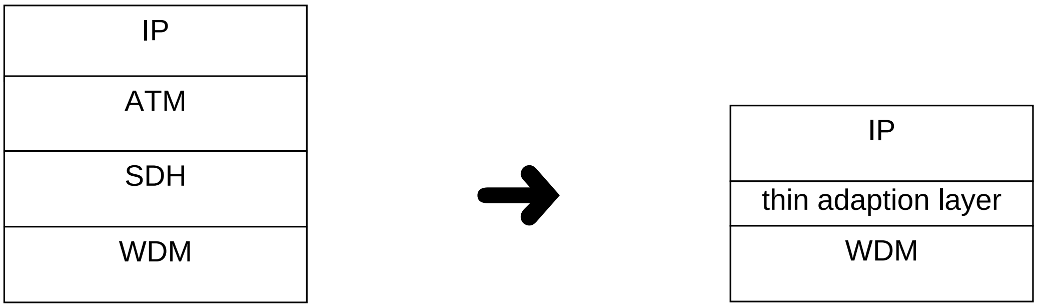


Figure 2.7: Possible evolution towards IP-over-WDM

tocols which are layered between IP and WDM. According to [5], ATM causes 10% ‘cell tax’ through the ratio of bytes in the header and in payload. By also considering measured lengths of IP packets which are mapped into ATM cells and the commonly applied mapping protocol classical IP over ATM, CLIP, the overhead caused by ATM increases to approximately 25% [5]. In addition, another 4% of overhead is added from SDH which results in roughly 30% capacity which is consumed by ATM and SDH.

Furthermore, transparent optical transport of data allows for carrying a maximum bitrate on any wavelength regardless of the protocol or framing structure. Thus, also different interface technologies could be carried by a single transport infrastructure.

Lastly, another factor for high costs in transport network which can be significantly reduced by eliminating ATM and SDH is the power consumption as well as the need for large footprints.

Based on the just discussed disadvantages of today’s layering principle, the aim is an evolution from opaque point-to-point optical pipes to transparent optical networking, see, e. g., [59], [60] [110] and [107]. This is especially important as about 70% of traffic in today’s backbone nodes is transit traffic and thus would be transported much cheaper, more efficient and faster without O-E-O conversion at every hop. In this context, the conjunction of an OTN and the data networking protocol IP remains a major challenge. Figure 2.7 depicts the current IP-over-ATM-over-SDH-over-WDM stack and the aimed IP-over-WDM stack with a thin adaption layer between the IP layer and WDM. This adaption layer could be, e. g., realized by optical burst switching which is introduced in Chapter 3. Nonetheless, it should be emphasized here, that this is an idealized architectural view on the protocol stacks. In reality, existing SDH equipment will not be thrown away just for the sake of a simplified protocol layering, especially not if the margins which can be earned in the sector of transport networks are small. Instead, an evolution towards such an IP-over-WDM architecture might be realized where the functionality of existing equipment is reduced step by step to necessary tasks which can be performed cheaper than with new equipment.

The main enablers of this transition towards an IP-over-WDM architecture are [62] (i) the progress in optical components which allows to start thinking about optical switching as well as (ii) ultra long haul transmission of optical signals which allow to transmit a signal for 1000’s of kilometers without regeneration.

2.4.2 Control Plane

Besides data plane issues, a further important driver for the evolution towards IP-over-WDM is the attempt of having a single control plane which controls IP on the network layer as well as the OTN, see [13], [140] and [145]. In current IP-over-ATM-over-SDH-over-WDM networks (applying classical IP over ATM), two separate control planes exist for the IP and the SDH protocol stack. Hereby, ATM is controlled by IP and WDM is operated statically. These two control planes are overlaid and fully separated which entails a set of disadvantages.

In addition to a significantly greater implementation effort and the risk of unwanted interaction, layering of protocol stacks also entails the existence of several management systems as each protocol needs its own management functionality. Hence, a change in policy or an topology upgrade requires to change several management systems and thus results in high complexity. As a consequence thereof, the time to switch bandwidth or a wavelength is also great which opposes the goal to make the optical layer more dynamic.

In order to overcome these drawbacks, the standardization bodies IETF, OIF and ITU-T started work to standardize a common control plane for IP and the optical transport layer, which is described in the following subchapters, see also [148] for general issues.

2.4.2.1 Standardization Efforts of the IETF

MPLS

MPLS [117] has evolved from a variety of approaches from different companies which wanted to combine layer 2 switching by ATM and layer 3 routing by IP. The approaches are IP Switching [101] from Ipsilon, Tag Switching [111] from Cisco, ARIS [162] from IBM and CSR [85] from Toshiba. In this context, MPLS was standardized in order to merge these vendor specific approaches which all eliminate the control plane of ATM and run IP and ATM under a common control.

The combination of layer 2 switching and layer 3 routing is obtained by separating the classification of an IP packet (to a longest prefix match) and forwarding functionality to the next hop which is usually carried out by each IP router. In the context of MPLS, such a router is called label switched router, LSR. With MPLS, classification is only carried out once at the network ingress and results in mapping of a packet to a forwarding equivalent class, FEC. Comparable to switching in ATM networks, a label (or stack of labels) is associated to an IP packet and all succeeding LSRs along the path towards the destination and an IP packet is forwarded based on a label. Besides label-based table look-up, also operations on a label stack like push, pop and swap of label are standardized, see, e. g., [117].

Two signalling protocols (resource reservation protocol, RSVP, and constraint-based routing using label distribution protocol, CR-LDP) are defined to setup label switched paths, LSPs, between ingress and egress nodes and to distribute routing information. Thus, MPLS realizes a distributed control plane.

In addition to speed-up of routing, this separation of routing and forwarding has a number of benefits. One of those is the possibility to perform traffic engineering. According to, e. g., an assigned port number or source/destination identifier, a packet can be classified to belong to a different FEC and hence, packets can be routed on different ways towards the destination.

Summarizing, the MPLS traffic engineering control plane comprises resource discovery, state information dissemination, path selection and path management which are independent of each other [7].

MPλS

MPλS [7], [58] evolved based on the idea to also control the OTN with concepts for a common control plane developed within the MPLS framework. Consequently, most of the aforementioned protocols and mechanisms can be reused. So far, the OTN has no control plane and all changes are made by network management. As already mentioned earlier, this also entails long processing time for new fibers or increased switching capacity.

As LSRs and OXCs have similar requirements with respect to control, [7] proposes to make an OXC an IP-addressable device. Instead of using an electronic label, an OXC uses a wavelength per LSP and one additional wavelength to carry the overall signalling. In contrast to MPLS, label merging and label stacking are not possible in this approach. Furthermore, the granularity of a label is a wavelength and thus, the number of labels is identical with the number of wavelengths. In order to allow for a smaller granularity and to better utilize wavelengths, traffic aggregation [45], [96] – which is also called traffic grooming in that context – can be carried out in the electronic domain at the edges. Such an ability of real-time provisioning of optical channels is a first step towards a more dynamic OTN and consequently a preliminary, circuit-switched version of a next-generation optical Internet.

MPλS is not standardized, however, its ideas are picked up and enhanced by GMPLS whereby the transition from MPλS to GMPLS are fluent. Hence, MPλS can be denominated as intermediate step between MPLS and GMPLS.

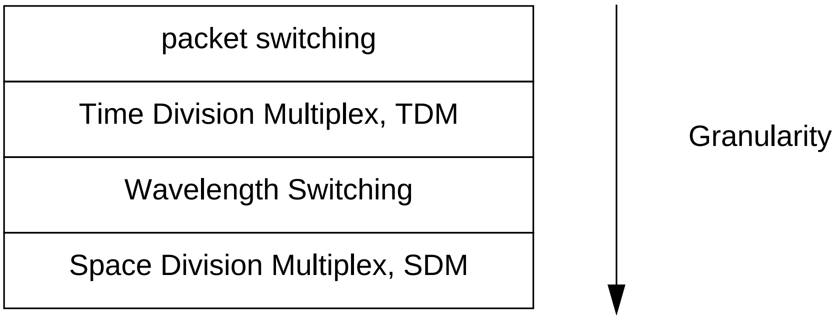


Figure 2.8: Multiple switching layers for GMPLS

GMPLS

GMPLS [156] extends MPLS and MPλS in that way that it introduces multiple switching layers, see Figure 2.8, which define the granularity of information which is switched. The finest granularity is the packet switching layer, which is applied in the present MPLS. The next coarser switching layer applies TDM. This layer was newly added in order to consider the switching granularity of SDH/SONET. Still coarser is the wavelength switching layer whose origin is MPλS. In this layer, whole wavelengths are switched. Finally, the SDM switching layer provides granularity to switch fibers.

The focus of GMPLS is to adapt existing protocols and mechanisms developed for MPLS to also support the just introduced switching layers. Especially signalling and routing have to be extended in order to carry also information about OXCs.

For the deployment of GMPLS, the overlay and the peer model are proposed. The overlay model has different instances of the control planes in OXC and LSR domain and therefore separates the OTN from IP. On the contrary, the peer model uses a single instance control plane and thus allows for integrated bandwidth-on-demand networking [4], [156]. Design and implementation issues for a MPλS/GMPLS control plane can be found in [148]. [10] and [11] discuss routing and signalling enhancements for GMPLS, respectively.

2.4.2.2 Standardization Efforts of the OIF

The Optical Internetworking Forum, OIF, [163] standardized a user-network interface, UNI, which allows a client (e. g., IP, ATM, SDH) to dynamically request/establish an optical channel. Hereby, signalling protocols are based on the ones standardized in GMPLS. OIF UNI 1.0 defines establishment and deletion of a connection, status exchange as well as auto detection of neighbors and services [103], [12] and is consequently a large step towards a more dynamic OTN.

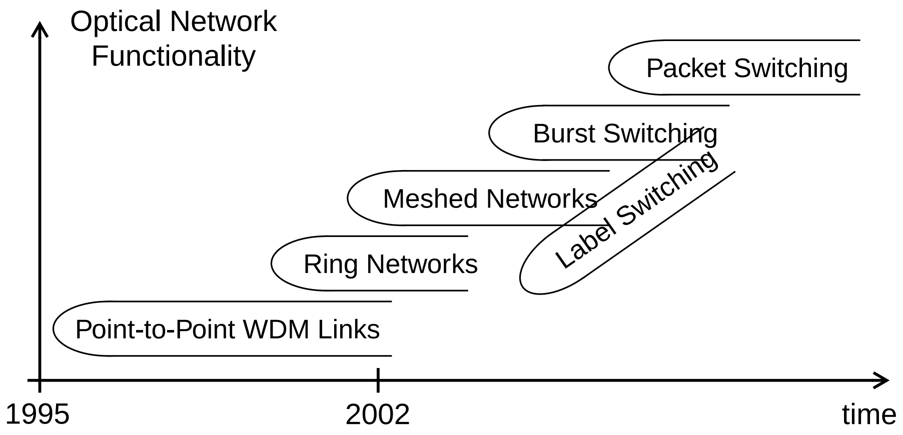


Figure 2.9: Evolution of photonic transport networks [43]

2.4.2.3 Standardization Efforts of the ITU-T

ITU-T standardized in G.8080 [79] an architecture for the automatically switched optical network, ASON, which can be applied on SDH transport networks according to G.803, see also Section 2.3.3 as well as to OTNs according to G.872, see also Section 2.3.1. The focus of G.8080 is hereby on a set of control plane components which allow to dynamically setup, maintain and release connections which are requested by clients (e. g., IP, ATM, SDH/ SONET) or by network management. For this functionality, a (recursive) hierarchical model exists where one node co-ordinates the remaining nodes. A distributed model is optionally defined. For call and connection management, the ASON recommendation G.8080 does not define specific protocols. Therefore, G.7713.1 [74], G.7713.2 [75] and G.7713.3 [76] define specific protocol recommendations for distributed call and connection control management. Whereas G.7713.1 recommends to use PNNI, G.7713.2 and G.7713.3 specify GMPLS-TE and GMPLS-LDP, respectively, (see Section 2.4.2.1) as signalling protocol [157]. As the UNI specified from OIF (see Section 2.4.2.2) is also considered by the ASON architecture, ITU-T sets up an umbrella with recommendation G.8080 to integrate all current standardization approaches for a dynamic OTN.

2.5 IP-over-WDM Network Architectures

In literature, a variety of different approaches can be found which propose to transport IP packets with a thin adaption layer over the photonic layer. These approaches can be distinguished in the grade of dynamic of the photonic layer.

Figure 2.9 [43] depicts a possible evolution scenario of optical network functionality against time. In 1995, the deployment of WDM technology started with static point-to-point links.

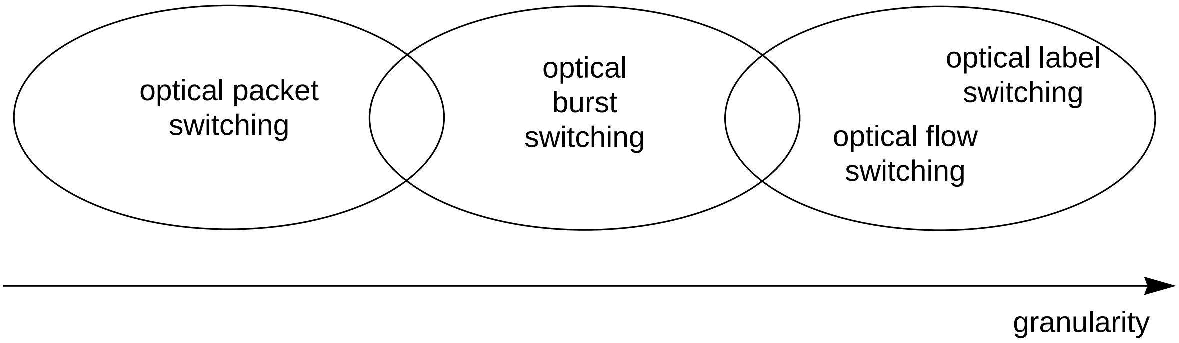


Figure 2.10: Granularity of IP-over-photonic architectures

Later, ring and meshed networks followed which increase the functionality to static circuit-switched networks. Higher dynamic is reached with GMPLS-based approaches which can be denoted as optical label switching, OLS, burst switching, OBS, or packet switching, OPS, which all are not yet applied today.

Depending on the granularity of switching, the functionality of OLS can be increased. In Figure 2.9, it is assumed that OPS applies fully optical switching and thus the highest optical networking functionality can be reached. As will be discussed later, OBS is based on a hybrid switching technology and thus is drawn below OPS. If or when such technologies will be applied mainly depends on whether the achievable revenue justifies high costs for dynamic photonic components. Additionally, technology evaluation is crucial as some technologies like, e. g., dynamic optical amplifiers are very expensive or even not realizable with today's technology.

In Figure 2.10, the three major IP-over-WDM architectures are distinguished by the switching granularity. Optical packet switching has the finest granularity whereas optical label switching as well as optical flow switching belong to circuit switched approaches and consequently have the greatest granularity. In between, optical burst switching fills the gap. As can be seen from Figure 2.10 and read in, e. g., Chapter 3, the limits between those architectures are not sharp and – depending on the applied definition – they even overlap.

In the following subchapters, the three major architectures depicted in Figure 2.10 are discussed in more detail. Section 2.5.1 gives an overview of circuit switched approaches for the optical layer, Section 2.5.2 discusses packet switched approaches and, finally, Section 2.5.3 introduces a hybrid approach called burst switched whose granularity is between circuit and packet switched.

2.5.1 Optical Circuit Switching

The architectures classified as (optical) circuit switched have the coarsest granularity with respect to allocated bandwidth as well as time a connection between source and destination is established. Herein, optical label switching, OLS, and optical flow switching, OFS, can be found in literature. For the setup and release of wavelengths, it is assumed here, that signalling functionality, e. g., by the ASON framework which has been discussed in Section 2.4.2, can be used. At the moment, OCS approaches with dynamic setup start to be implemented.

2.5.1.1 Optical Label Switching

Optical label switching is an approach with the coarsest granularity in allocated bandwidth and time. It was first denoted as MPλS and later as GMPLS (see Section 2.4.2). In OLS, a wavelength is at the same time the label and thus the smallest granularity in the network. In order to obtain a smaller granularity, several connections with smaller bandwidth requirement can be aggregated/groomed to one wavelength at the network edge.

The setup and release of wavelengths between ingress and egress router is rather static and thus is either topology-driven, or by management using signalling functionality comprised in the ASON framework, see Section 2.4.2. The time a wavelength path exists unchanged is generally much greater than the lifetime of an average flow and many (successing) flows with the same source and destination edge router share a common wavelength. Thus, the resulting network topology arising from routing of wavelengths has to be optimized with respect to minimal wavelength usage and load distribution in order to efficiently utilize the available resources. Concluding, the intelligence and hence the majority of work is contained in the (constrained-based) routing of wavelengths. For wavelengths routing, many studies as well as tools are available in literature, see, e. g., [126] for a comprehensive overview.

The advantages of the OLS approach are that losses cannot occur after setup of a connection. Like in all circuit-switched approaches, the only service degradation is the (call) blocking probability at setup or by grooming to a wavelength. A further advantage is the simplicity of that approach as switching does not have to be very fast and can be based on a much cheaper technology. Thus, add-drop multiplexer can be used which are cheaper than complex optical cross connects. Furthermore, (re-) computation of wavelength routing can be carried out offline.

On the other hand, OLS has also a variety of disadvantages. The first disadvantage is the dependency of the required number of wavelengths from the number of edge routers. In order to have full connectivity in a network with k edge routers and n service classes, at least $(k-1) \cdot n$ wavelengths are required.

A second major disadvantage is the difference between the quasi-static wavelength channels and the bursty IP traffic [36], [32], [46] which is transported by them. On every channel, a large amount of bandwidth has to be over-allocated in order to be also able to transport traffic peaks [87]. Thus, most of the time this amount of bandwidth is not used. This waste of bandwidth even gets worse as wavelengths are not used as common shared resource. Consequently, the

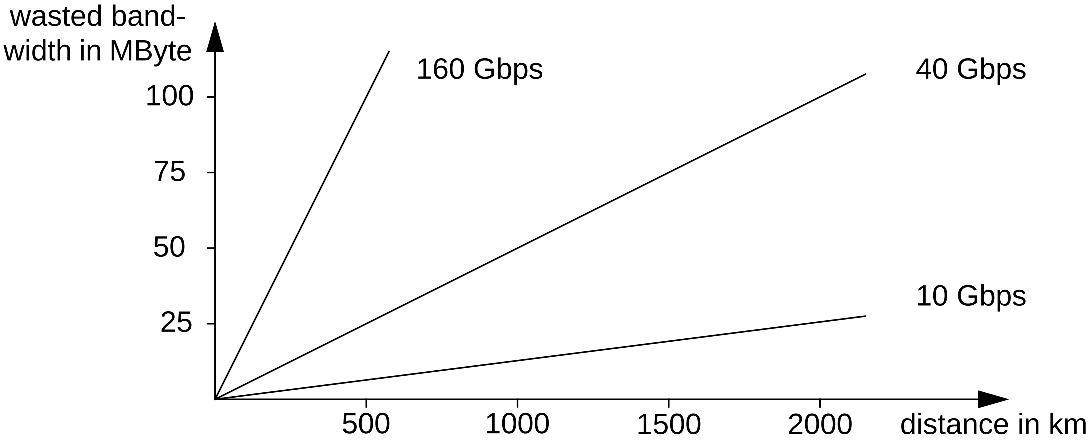


Figure 2.11: Bandwidth- delay product

spare bandwidth has to be reserved on every wavelength and a multiplexing gain approximated by the M/G/n loss system cannot be exploited (see also Section 4.1).

2.5.1.2 Optical Flow Switching

Approaches which are classified as optical flow switching, OFS, [143] have a granularity in time which is much finer than the one of OLS. With OFS, a wavelength is only allocated for the duration of a flow and released afterwards. This allows to reuse resources (and thus to achieve multiplexing gain) which is especially important in larger networks. In contrast to the more or less topology-driven OLS, OFS is traffic-driven. Hence, routing of wavelength cannot be carried out offline and the switching of wavelength has to be much faster in order not to cause too much switching overhead.

Most critical for the performance of OFS is the detection of the begin and the end of a flow. In [95] and [99] simulation studies – in a slightly different context of predecessor protocols of MPLS – show the dependencies of the percentage of switched flows on a variety of parameters as well as the required number of labels (wavelength) depending on the time for recognition of the end of a flow.

In contrast to electrical flow switching where IP packets can be forwarded before a flow is detected, OFS requires the setup of a wavelength before the transmission of the first byte starts. If, at network ingress, a packet cannot be classified to belong to an existing flow, an acknowledged setup has to be carried out. The end of a flow has to be detected with a timer, e. g., if no packet belonging to that respective flow has arrived during a certain period of time.

The advantages of OFS are the ability to use all wavelengths as one common shared resource and thus obtain a multiplexing gain. Furthermore, OFS better supports the bursty nature of packet traffic.

Besides these advantages compared to OLS, OFS has also some disadvantages. One of them is caused by the large bandwidth delay product in photonic networks, i. e., the amount of Bytes which can be transmitted in such a broadband network during an end-to-end transmission time increases with increasing bandwidth. As a link remains idle during that time (end-to-end transmission and processing at every node) a great amount of bandwidth is wasted in broadband networks, see also Figure 2.11. Herein, the wasted bandwidth measured by the amount of wasted capacity dependent on the one way distance x can be calculated according to

$$b_w = \frac{2 \cdot x}{c} \cdot r \qquad (2.1)$$

with the link bit rate r and the signal propagation speed $c \approx 200000 \mathrm{km/s}$. In this formula, only the two-way transmission time is considered whereas the processing time (which can be of the same order of magnitude) is omitted.

It can be seen that already for a one-way distance of 1000 km, e. g., Stuttgart - Berlin, a considerable amount of bandwidth cannot be used due to waiting for the setup acknowledge. In case of, e. g., 10 Gbps 12.5 MByte are wasted. This waste of bandwidth might lead to a poor utilization if the mean flow size is not significantly greater. For greater link bandwidth, the amount of wasted bandwidth increases considerably. Hereby, the related issue that large buffers are required at the network ingress to buffer packets which wait for an setup acknowledge has to be addressed.

Additionally, another waste of bandwidth is caused by the time to detect the end of a flow in order to release the allocated resources.

2.5.1.3 Fast Circuit Switching in ATM

For completion, concepts for fast circuit switching which were developed at the end of the 80ies are also mentioned in this context. A comprehensive overview can be found in [25].

Besides reservation of bandwidth or buffer for a connection, protocols for fast resource management are available in the context of ATM. They were fist described in [68]. Approaches of this category can be classified in approaches which do not adapt to current network load and categories which do. The latter approaches have their origin in data communication [108]. A representative is Available Bit Rate, ABR, with the flavors rate-based [15] and window-based [92].

In the category of approaches which do not adapt to current network load are fast bitrate reservation, and fast buffer reservation. The latter one is not interesting in the context of photonic networks, as the availability of (optical) buffers is mandatory. For the first category, a fast reservation protocol with delayed transmission FRP/DT, as well as with immediate transmission, FRP/IT, is available in literature, see, e. g., [21]. The main motivation to prefer FRP/IT to FRP/

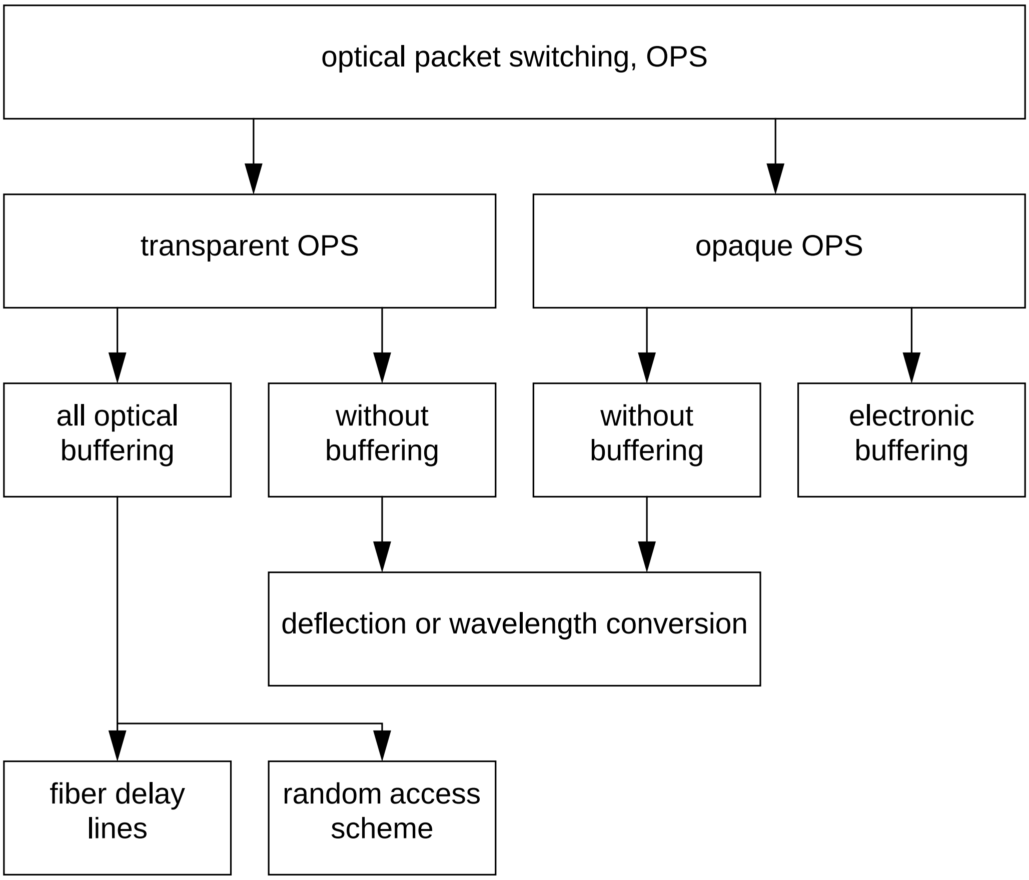


Figure 2.12: Classification of available OPS realizations [139]

DT is a scenario where the burst duration is small compared to the round trip delay, which is also true for broadband optical networks.

The FRP/IT protocol is a smooth transition between circuit switching and burst switching (see also Figure 2.10). Its main characteristics are bandwidth reservation on burst level, i. e., for a set of ATM cells at the same time, reservation ‘on the fly’, and no acknowledgement whether the burst reservation is successful. In case of not enough resources, the burst can be either buffered in intermediate nodes or it has to be discarded. This principle is promising and thus also the basis for bandwidth reservation in optical burst switched networks, see Chapter 3. In [21] it is stated, that burst admission control has to be carried out in a distributed way. This principle is also realized in the framework Assured Horizon in Chapter 6.

2.5.2 Optical Packet Switching

In optical packet switching, OPS [139], [133], [149], [113], which is sometimes also called photonic packet switching, the granularity in time is further reduced compared to OFS in order to also allow multiplexing between flows. Major work was performed in projects like ATMOS [97], KEOPS [52], WASPNET [71], and HORNET [142]. The major tasks of a switch is to perform switching, buffering and header translation [70]. Compared to electronic packet switching (e. g., ATM, see Section 2.3.2), significant differences arise due to the boundary conditions of the optical layer. Hereby, the way to place and process the packet header and the way to perform contention resolution cannot be applied like in electronic packet switching. Figure 2.12 [139] presents a classification of OPS with respect to these two issues.

- **Realization of the packet header (header translation)**

  In OPS, the information about the destination contained in the packet header is mostly valid locally and thus has to be substituted in every node. As shown in Figure 2.12, this can be performed transparently in the optical domain or opaque requiring an O-E and an E-O conversion in every node.

  Also, in contrast to electronic packet switching, the packet header cannot only be sent serially prior to the data, but also on a different wavelength or in parallel to the data at a lower frequency by subcarrier multiplexing [71]. However, subcarrier multiplexing has the disadvantage that routing decisions can only be carried out after the entire packet is received. This may cause longer delays in case packets are very long. If the packet header is sent on a separate wavelength, effort is required in order to ensure realignment between header and data. However, it is easy to demultiplex the header at every node. Finally, serial transmission requires more bandwidth as penalty for simplified header extraction.

- **Mechanisms for contention resolution (buffering)**

  The performance of packet switching heavily depends on the ability to perform contention resolution of packets which are destined at the same time for the same output. As a simple optical random access memory is not available, different schemes can be applied [70], [139], [28], see also Figure 2.12. The most simple optical buffer is called fiber delay line, FDL. It is simply a long fiber which delays the signal by a constant signal propagation time. As simple optical buffer, a constant delay in time by fiber delay lines, FDLs can be applied. A cascade of FDLs of different lengths allows to perform more sophisticated buffering. Additional concepts are deflection to other ports or conversion to other wavelengths (at the same port). Whereas deflection routing may cause global contention originated by local contention, wavelength conversion is an important contention resolution scheme in optical networks. However, (costly) wavelength converters are required therefore.

- **Technology for very fast switching (switching)**

  For the complexity of switching, the mean packet size and whether packets have equal or variable length is crucial. For reasons of simplicity, most approaches assume equal packet length. The mean packet length determines how fast switching has to be carried out, as the time for switching has to be significant shorter than the time to transmit a packet of mean length. The switching time also determines the required switching technology and thus directly the cost of a system.

Despite much research work on OPS, its realization is still far in the future. This is mainly because of the costly realization of contention resolution with sophisticated optical buffers and the fact that header processing still cannot be performed all-optically at reasonable costs.

2.5.3 Optical Burst Switching

Optical burst switching, OBS [105], [134], [141], is a hybrid approach between OCS and OPS which tries to combine the advantages of both approaches and avoiding most of their drawbacks. The main characteristics of OBS are [43]:

- OBS granularity is between circuit and packet switching.
- There is a separation between control information (header) and user information (data). Header and data are usually carried on different channels with a strong separation in time.
- Resources are allocated without explicit two-way end-to-end signalling, instead so-called one-pass reservation is applied.
- Bursts may have variable lengths.
- Burst switching does not require buffering.

Note that not all of these features must be satisfied and ‘smooth’ transitions to packet and to (fast) circuit switching are possible, which is also indicated in Figure 2.10. Although the concept of burst switching has been already known since the 1980s [2], [3], [21], it has never been a big success in electrical networks. The main reason is that its complexity and realization requirements are comparable to that of more flexible electronic packet switching techniques (like, e.g., ATM).

However, with the introduction of very high speed optical transmission techniques this has changed. Now, there is an even increasing discrepancy between optical transmission speed and electronic switching capability. Moreover, due to cost and complexity aspects, it is advantageous to keep data in the optical domain and to avoid O-E conversion. On the other hand, as already mentioned in Section 2.5.2, all-optical packet switching is still too complex to perform all processing in the optical domain.

Therefore, a hybrid approach like burst switching seems very promising: it keeps data in the optical domain but separates control information which allows sophisticated electronic processing of this control data.

Thus, the major difference to OCS is the one-pass reservation and its granularity which is usually shorter than a flow or even a connection. The major difference to OPS is the separation of header and data in space and in time (which is an option in OPS), its granularity which is usually coarser than the one of OPS and finally the fact that buffers are not mandatory. Furthermore, the coarser granularity of OBS compared to OPS allows to apply switching technologies which are much slower and hence much cheaper than technologies required for OPS. Therefore, the rest of this thesis focuses on an IP-over-WDM-based architecture with OBS as adaption layer between IP and the (dynamic) OTN. A detailed discussion of OBS follows in Chapter 3.

Chapter 3

Optical Burst Switching

In this chapter, a comprehensive discussion of optical burst switching, OBS, is presented from the viewpoint of network architectures, (reservation) protocols, QoS support and qualitative performance evaluation. Further issues like node architectures including optical components, see, e. g., [26], are beyond the scope of this thesis.

Section 3.1 starts with an overview of the functionality as well as aspects related to control and switching technology and introduces design parameters. In Section 3.2, burst assembly mechanisms – which are very important for the network performance – are introduced, discussed and classified. Section 3.3 presents an overview of reservation mechanisms. Herein, reservation mechanisms reported in literature which do not differentiate between service classes are described and classified. Finally, Section 3.4 discusses challenges which arise when QoS differentiation is supported in an OBS network and approaches which can be found in literature.

3.1 Overview of OBS

3.1.1 Functionality

At the moment, OBS is still at its definition phase which is indicated by a strongly increasing number of publications on new reservation mechanisms, [141], [134], [144], [150], [105], [38], [130], [43], assembly mechanisms [57], [144], [33], [65], [42], [41], [138] prototypes [48], [8] and architectures [29], [147], [135], [41], [40], [153]. Although there is still no unique definition of OBS in literature, it is widely agreed that the following list describes its main characteristics [43]:

- OBS granularity is between circuit and packet switching.

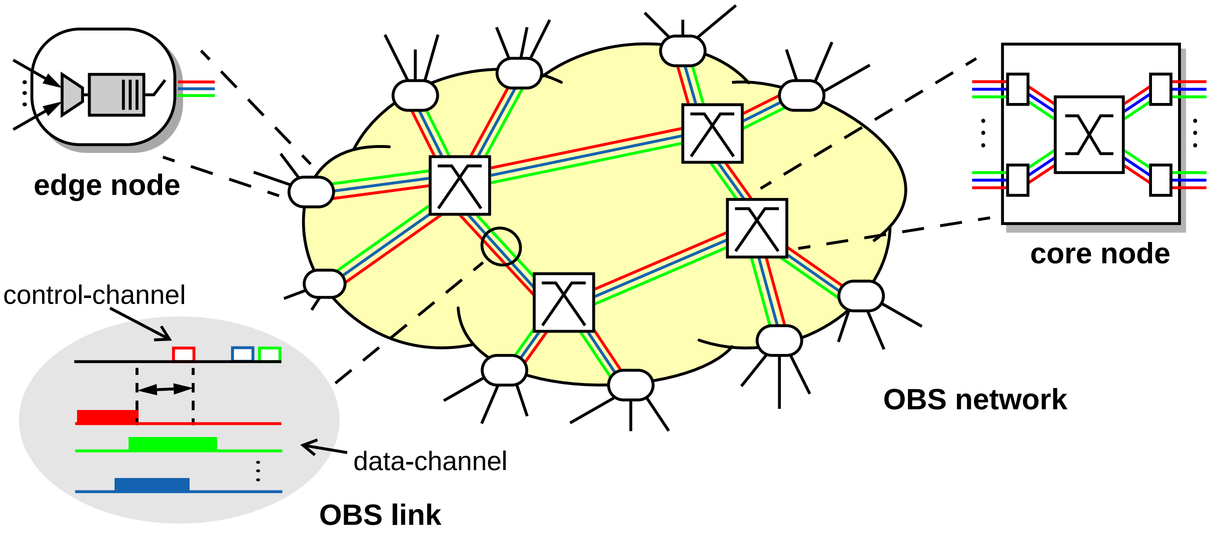


Figure 3.1: Functionality of optical burst switching

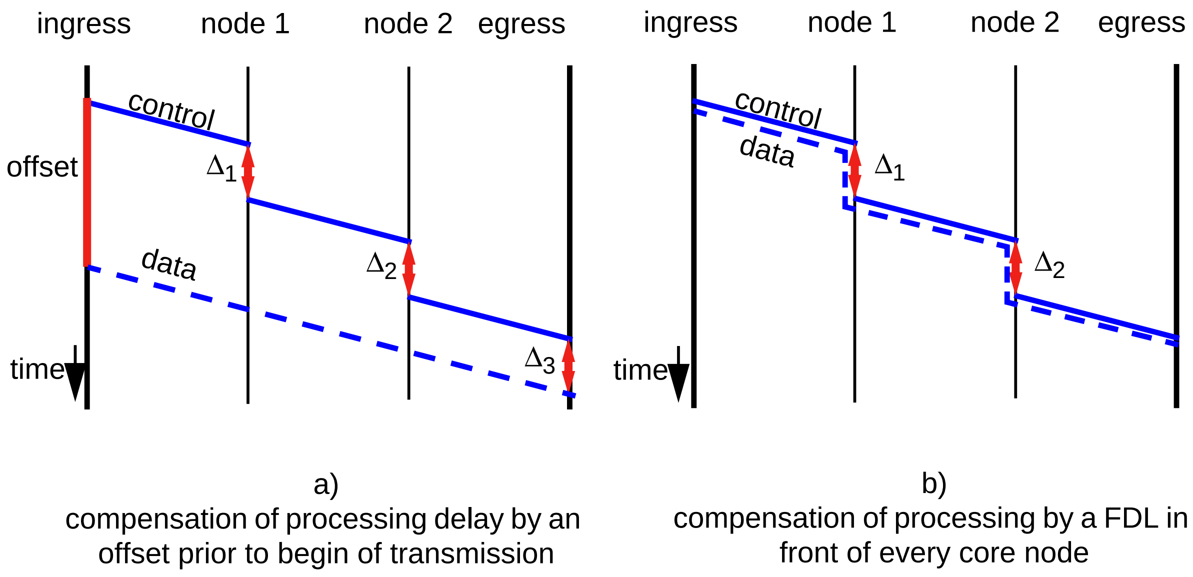
- There is a separation between control information (header) and data. Header and data are usually carried on different channels with a strong separation in time (see example OBS network link in Figure 3.1).
- Resources are allocated without explicit two-way end-to-end signalling, instead so-called one-pass reservation is applied.
- Bursts may have variable lengths.
- Burst switching does not require buffering inside the core network.

Note that not all of these features must be satisfied and ‘smooth’ transitions to packet and to (fast) circuit switching are possible, see Figure 2.10.

Although the concept of burst switching has been already known since the 1980s ([2], [3], [21]) it has never been a big success in electrical networks. The main reason is that its complexity and realization requirements are comparable to that of more flexible electronic packet switching techniques (like, e.g., ATM).

However, with the introduction of very high speed optical transmission techniques this has changed. Now, there is an even increasing discrepancy between optical transmission speed and electronic switching capability. Moreover, due to cost and complexity aspects, it is advantageous to keep data in the optical domain and to avoid opto-electronic conversion. On the other hand, all-optical packet switching is still too complex to perform all processing in the optical domain.

Therefore, a hybrid approach like burst switching seems very promising: it keeps data in the optical domain but separates control information which allows sophisticated electronic pro-



a)
compensation of processing delay by an offset prior to begin of transmission

b)
compensation of processing by a FDL in front of every core node

Figure 3.2: Functionality of one-pass reservation

cessing of this control information. By doing so, decoupling of speed of header processing and data forwarding is achieved.

Figure 3.1 shows some of the main characteristics of an OBS network. There are two types of nodes. In edge nodes, traffic is collected from access networks and assembled into larger data units, so-called bursts. Core nodes serve as transit nodes in the core network. Their main task is switching bursts optically without any processing of the data part. To achieve this, some control information containing reservation requests is necessary ahead of every burst's transmission time.

There are several possibilities how to perform reservation of data channel bandwidth. This thesis concentrates on the evaluation of so-called separate control, delayed transmission, SCDT, schemes. These reservation concepts are based on a strong separation of control information and data. A reservation request is sent in a separate control packet on a different channel while the actual transmission of the data burst is delayed by a certain basic offset (see Figure 3.1 and Figure 3.2). This basic offset enables intermediate nodes to process control information and set up the switching matrix. In contrast to systems with immediate transmission, which send control information together with the burst, the network can perform header processing without buffering data in each node along the path. This principle is depicted in Figure 3.2a. SCDT, however, requires higher complexity in edge nodes and introduces additional delay to bursts. The basic offset has to compensate for the sum of processing times in all intermediate nodes. Therefore, some upper limit of the number of intermediate nodes has to be known prior to reservation which requires some kind of source routing. In each core node, offset information in the header has to be reduced by the actual processing delay. An alternative approach (see Figure 3.2b) is to send control and data at the same time on different wavelengths and compen-

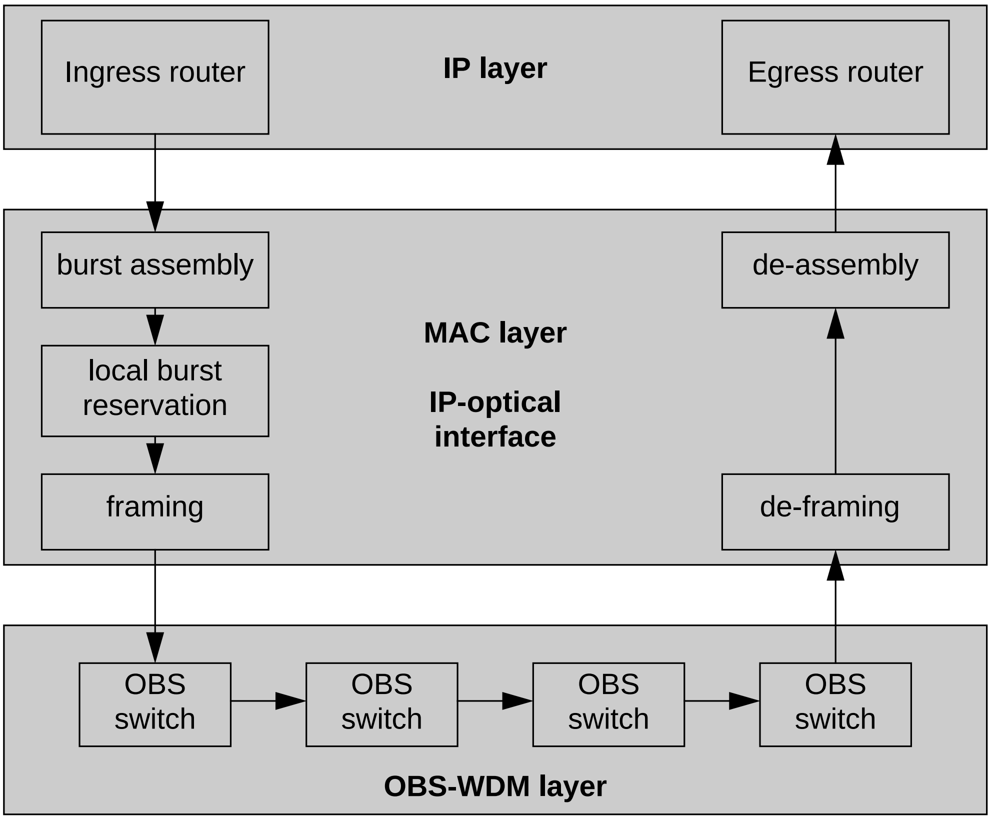


Figure 3.3: IP-optical interface between IP layer and OBS-WDM layer [33], [34]

sate the processing time of the control information by an FDL for data in front of every core node. This has some advantages which will be discussed in Section 5.1.

Comparable to the reservation mechanism FRP/IT introduced in Section 2.5.1.3, SCDT schemes use one-pass reservation, i.e., the sender of a burst does not wait for an acknowledgement of its reservation request, see Figure 3.2. This approach stands in contrast to two-pass reservation as typically applied during connection setup in circuit switched optical networks. The advantage of a one-pass reservation is higher efficiency with respect to throughput as there is no overhead caused by propagation delay, see also Section 2.5.1.2. An example may illustrate this. The transmission time of a 100 kByte burst on a 10 Gbps link is 80 µs while the propagation delay over a distance of 1000 km (which is the distance between Stuttgart and Berlin) is about 5 ms. Hence, during 10 ms of waiting for the acknowledgement to arrive, 125 bursts could have been transmitted. By including processing times or assuming higher bitrates, the number of transmitted bursts even increases.

In Figure 3.3 the functionality of OBS is mapped to network layers. The thin adaption layer between IP and the optical layer introduced in Figure 2.7 is called MAC layer or IP-optical interface in [33] and [34]. Burst assembly, local reservation of bursts as well as framing is carried out at network ingress while de-framing and de-assembly is the functionality of this layer at network egress. In the OBS-WDM layer, a reservation mechanism (see Section 3.3) and possibly also OBS-QoS mechanisms (see Section 3.4) is performed. Hereby, bursts are kept in this OBS-WDM layer.

The area of application of OBS is a scenario where bandwidth becomes a scarce resource. In this case, neither the static allocation of wavelengths nor the wavelength routed approach by ASON/GMPLS, Section 2.5.1, allows for enough multiplexing in order to gain enough bandwidth. Additionally, the bursty nature of IP traffic [36] requires a high degree of over-provisioning in order to also be able to transport traffic peaks. On the other hand, the application of OBS is seen in a technological environment where optical buffering of packets is not widely available as optical RAM and consequently OPS is too complex and too costly. Accordingly, OBS might be a technology which bridges the gap between OLS and OPS, see also a possible evolution scenario in Figure 2.9.

3.1.2 Control

In this thesis, especially for the framework Assured Horizon which is introduced in Chapter 6, the existence of GMPLS-based control (Section 2.4.2.1) is assumed. Hereby, bursts are classified to forwarding equivalent classes, FECs, at the ingress. (Constraint-based) routing is carried out by GMPLS resulting in an allocation of a label for every FEC which fixes a path per service class between ingress and egress. However, only fibers are determined whereas WLs are allocated dynamically for every burst by a reservation mechanism, see Section 3.3.

Such a framework including the GMPLS layer is discussed in [33] and also called labelled optical burst switching [105] or burst switching in virtual circuit mode [134].

3.1.3 Switching Technology

This Section shortly reviews and discusses [120], where possible switching technologies for burst switching are compared.

In order to operate a packet-switched system with efficient utilization, the switching time, which is visible as guard time between bursts, has to be negotiable compared to a mean burst transmission time. Therefore, the applied switching technology and the mean burst length in a system have to be chosen accordingly.

Figure 3.4, which is taken from [120], presents the context between burst length, burst transmission time and switching time against a logarithmic time scale. From this graph, it can be seen which switching technology is fast enough for OBS. Micro electro-mechanical systems, MEMS, [164] require switching times between 1-10 ms which is even greater than the transmission time of a very large burst (about 1 MByte at a link rate of 10 Gbps) and consequently cannot be applied in OBS networks. Acousto-Optic-Tunable-Filters, AOTFs, [119] have a

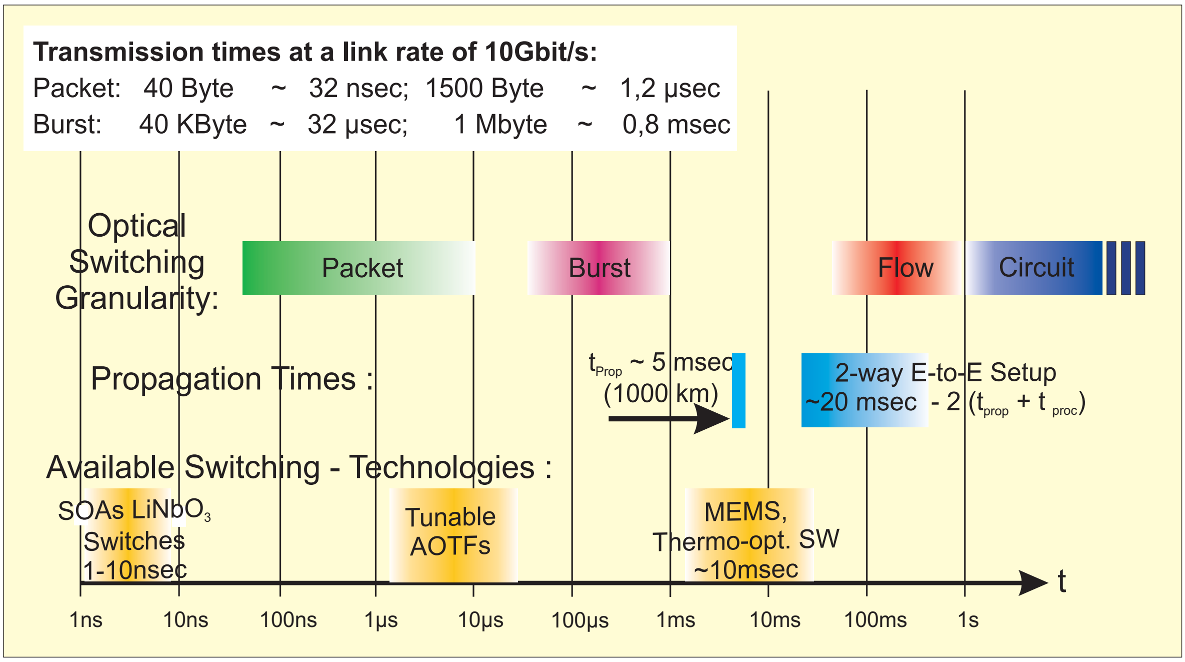

Transmission times at a link rate of 10Gbit/s:
Packet: 40 Byte ~ 32 nsec; 1500 Byte ~ 1,2 µsec
Burst: 40 KByte ~ 32 µsec; 1 Mbyte ~ 0,8 msec
Optical Switching Granularity:
Packet
Burst
Flow
Circuit
Propagation Times :
t_Prop ~ 5 msec (1000 km)
2-way E-to-E Setup ~20 msec - 2 (t_prop + t_proc)
Available Switching - Technologies :
SOAs LiNbO3 Switches 1-10nsec
Tunable AOTFs
MEMS, Thermo-opt. SW ~10msec
t
1ns
10ns
100ns
1µs
10µs
100µs
1ms
10ms
100ms
1s

Figure 3.4: Context between transmission time, propagation delay and switching time [120]

switching time just below 10 µs and can be applied if the mean burst duration is about 1 ms in order to consume about 1% of the capacity for switching. Finally, Semiconductor Optical Amplifiers, SoAs [112], satisfy the needs for efficient switching of bursts. Here, the switching time is below 10 ns and thus completely negligible compared to burst duration. For the rest of this thesis, it will be assumed that SoAs are used as components to switch optical bursts. As a consequence switching times will be neglected and all evaluations focus on networking performance.

3.1.4 Design Parameters of OBS

The following list describes the most important design parameters for OBS and includes examples from literature.

- **Contention resolution**. Like in any packet switched network which is based on multiplexing gain between flows, contention can also occur in OBS. Contention resolution for OBS can be in space, in time or in frequency.
  The vast majority of publications assumes that basic contention resolution is in frequency domain which requires WDM transmission technology with full wavelength conversion in a core node such that each burst can be switched to any wavelength of an output fiber channel. However, there is a trade-off between performance benefits due to higher number of wavelength channels and higher cost due to more wavelength converters [125], [124].
  Although buffers (contention resolution in time) are not mandatory in OBS, their (additional) use can further decrease the burst loss probability. Many proposals avoid buffers or

use only simple FDLs to keep the system significantly less complex than a packet switching system [104], [134], [141], other work includes sophisticated buffering concepts [151]. Finally, contention resolution can be also in space by deflection routing, which is considered for OBS in [137]. If a contention occurs on one output port, a burst is sent on another available output port which also leads towards the desired destination. However, this contention resolution mechanism may result in global congestion caused by local congestion.

- **Resource reservation mechanism**. Key system resources which have to be reserved are channels and possibly buffers. There are several proposals in literature ([105], [134], [141]) which are classified and compared in Section 3.3.
- **QoS support**. First proposals for burst switching only considered one class of bursts [134], [141]. Due to the increasing importance of QoS support, recent proposals extended the OBS concept to multiple service classes [151], [152], [136], [35], [41]. Section 3.4 will discuss and classify different available mechanisms to realize service differentiation.
- **Protocol aspects**. Designing a protocol for OBS strongly depends on the reservation mechanism and QoS support to be realized but still offers some degrees of freedom. Even for the one-pass reservation scheme the focus is on, ‘one-way’ [104] or ‘two-way’ [141] protocols are possible. In the latter case, blocking events or successful channel reservations are reported back. Note that even with two-way protocols in an SCDT scheme burst transmission starts before any confirmation message is received at the initiating node. However, this feedback from the network can be used to adapt to overload situations.
- **Node architecture and technology**. Depending on the design choices for the parameters listed above, there are many realization possibilities for a burst switching node. Basic building blocks are I/O interfaces, control information processing units such as a reservation manager, and switching systems for control and user data possibly including buffers (see Figure 3.1). [134] gives a very detailed description of an example node architecture, [151] describes various delay line concepts.

3.1.5 Flavours of OBS

Besides the definition presented at the begin of this chapter, the following flavours of OBS can be found in literature which will not be considered in the rest of this thesis:

- **All bursts have equal length**

This flavour of OBS assumes that all bursts have the same size, see, e. g., [93]. The major advantage is simplified switching. However, a disadvantage is the fact that bursts might be padded[1] in case not enough data is available in an assembly buffer. This concept is closely related to ATM where the transport unit is also fixed in size, see Section 2.3.2.

- **Acknowledged two-way setup**

In this flavour of OBS, a burst will be only transmitted after having received a positive acknowledgement of its setup request. The advantage of such a scheme is that no burst is lost in the network. However, blocking of the network has to be considered which leads to either additional delay or to buffer overflow (and thus losses) at the network ingress. Besides, acknowledged two-way setup entails long delays caused by the large bandwidth-delay product, which was discussed in an example in Section 3.1.1. Examples of such schemes are Tell and Wait, TW, [38] and the publications of the group from Bayvel, e. g., [37] and [47].

- **Mandatory buffers in all nodes**

Very recent publications [129] propose some sort of store-and-forward routing for optical bursts. Hereby, each burst is buffered for a certain time in a node in order to determine which burst can be forwarded to the next node and which burst should be dropped.

3.2 Burst Assembly

Burst assembly denotes the subsumption of a number of smaller transport units, e. g., IP packets, to one larger transport unit called burst. This results in a granularity (in Bytes) which is coarser than the original granularity and thus in greater mean size of transport units as well as a reduction in mean interarrival time between bursts.

Motivations for greater transport units are the possibility to apply slower and consequently cheaper switching technologies. This is especially important for systems with very high data rates. Furthermore, the overhead caused by signalling can be reduced in two ways. Firstly, the ratio between header and data gets smaller, and secondly, a switch has to handle less requests to switch the same amount of information. This is especially important as the number of processed requests per time unit is usually a limiting factor.

In addition, aggregation of flows is a general concept to reduce the complexity in the network while still allowing for service differentiation [44], [45]. This concept can be easily combined with burst assembly. Finally, burst assembly directly influences traffic characteristics, like burst interarrival time and burst length (distribution) which have major impact on the network performance. In order to improve traffic characteristics, traffic control mechanisms can also be integrated in the burst assembly process.

[1] In order to save resources, padding of higher priority burst could be done with bytes from an assembly queue of lower priority classes.

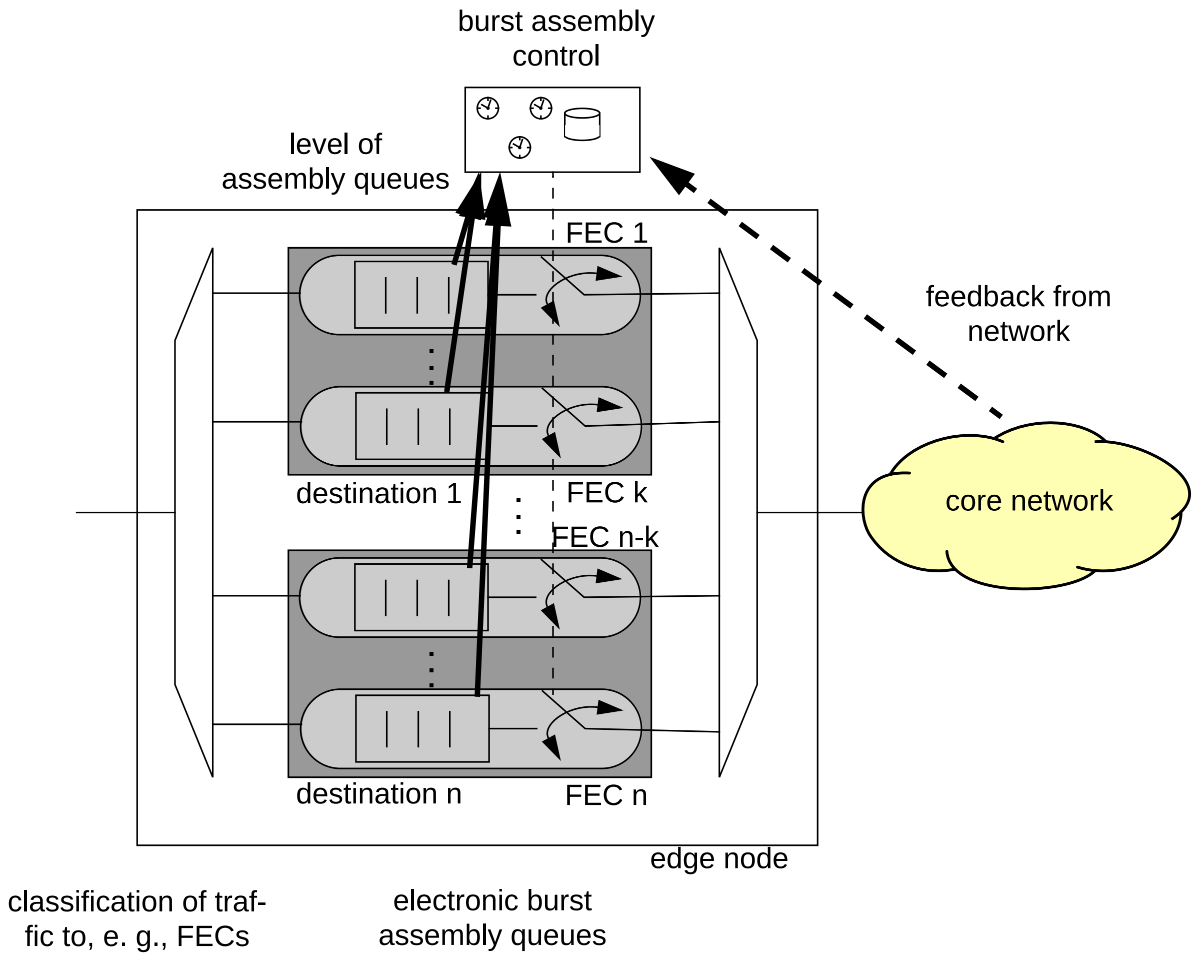


Figure 3.5: Model of a burst assembler at an edge node

3.2.1 Functionality

Figure 3.5 depicts the model of an edge node containing several (electronic) burst assembly queues per outgoing destination. By isolating traffic in different assembly queues, not only service class differentiation can be performed, but also different assembly strategies per service class can be applied. In order to decide in which queue an arriving packet should be written to, packets are classified. If GMPLS control is applied to control the network, packets are classified to FECs and an edge node maintains one burst assembly queue per FEC.

Control of the burst assembly process may depend on a variety of parameters like, e. g., monitoring of the duration of states, thresholds of assembly queues and feedback on network state. This is indicated in Figure 3.5 by arrows pointing to the burst assembly control unit. In Section 3.2.2, a detailed discussion and classification of assembly mechanisms is presented.

At certain instances in time, a number of packets are taken out of an assembly queue and forwarded to the network as one burst. Herefore, a burst header packet, BHP, is created which contains information like, e. g., burst length, current wavelength, service class and information

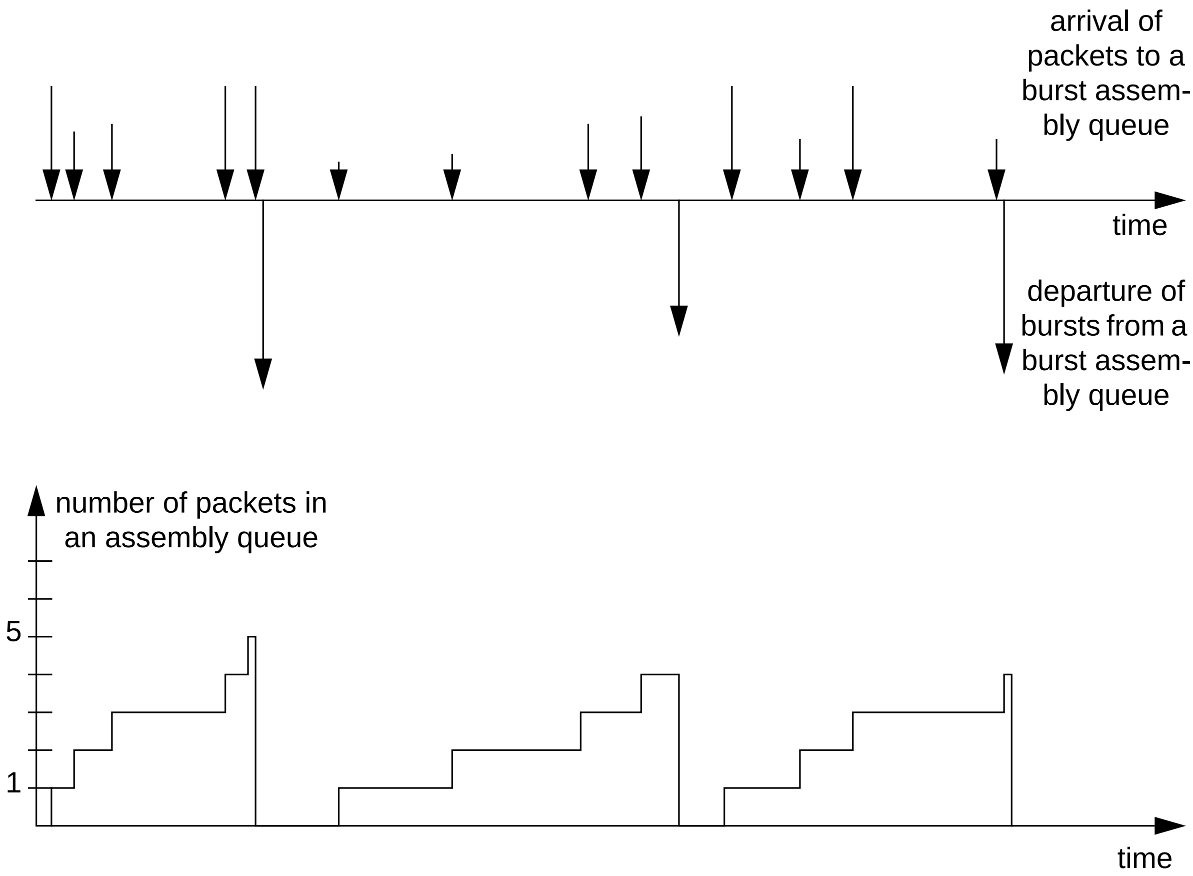

arrival of packets to a burst assembly queue
time
departure of bursts from a burst assembly queue
number of packets in an assembly queue
5
1
time

Figure 3.6: Packet arrival and burst departure process

on the destination. As the functionality of OBS is not yet standardized, additional fields may be required in a BHP. Figure 3.6 depicts an example of the packet arrival process to a burst assembly queue and the burst departure time of that queue. Also depicted in this figure is a state diagram of the assembly queue. Hereby, it is assumed that, at the time a burst is assembled, all packets[1] in the queue are assembled to a burst. It can be seen that – from view point of teletraffic theory – this behavior corresponds to a batch departure process [88].

3.2.2 Classification of Assembly Mechanisms

On of the most important classification criteria is whether the mechanism is controlled by an internal timer or not. Figure 3.7 presents a classification of possible burst assembly mechanisms which are (internally) timer-based whereas Figure 3.8 presents a comparable classification for non-timer-based assembly strategies. Additionally to those classifications, one can differentiate whether padding is applied or not. This is especially important for assembly strategies which may release bursts to the network that are shorter than a minimum burst size. In [57], a burst assembly mechanism which applies padding is proposed and evaluated.

[1] An approach where – in some cases – only a subset of packets contained in the assembly queue is assembled to a burst is presented in Section 6.3

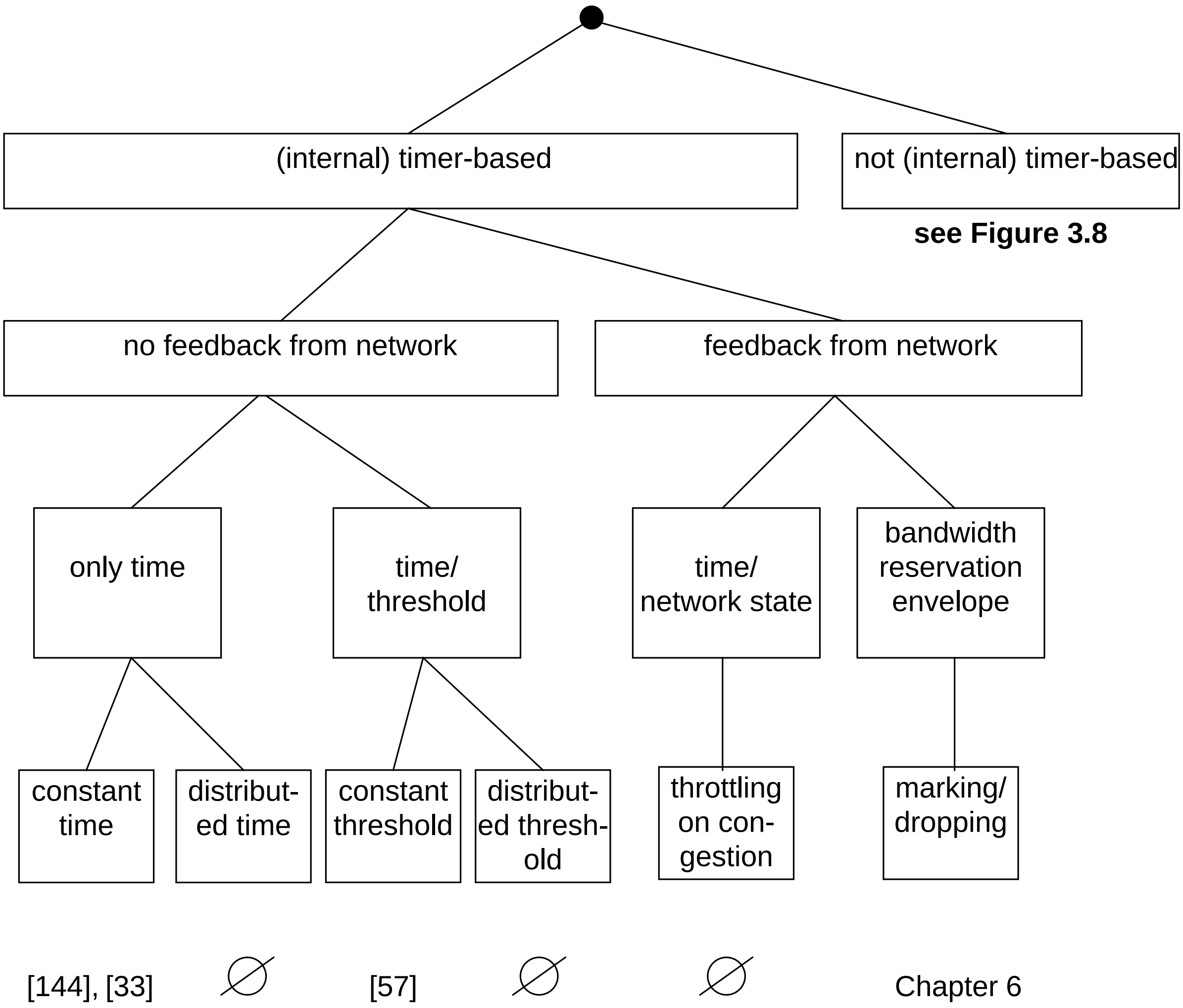


Figure 3.7: Classification of burst assembly mechanisms – timer-based

All burst assembly mechanisms reported in literature are basically timer-based in order to be able to guarantee a maximum waiting time in the assembly buffer. However, some slightly different flavours are available. In the classification in Figure 3.7, the simplest burst assembly control is only based on a timer. At arrival of a packet to an empty assembly queue, the time is set to a value which follows a certain distribution. At timeout, all packets contained in the respective assembly queue are assembled to one burst which is forwarded towards the network. This strategy realizes the well-known token bucket shaper. In [144] and [33], such a purely time-based solution with constant time is suggested. Additionally to that, [33] introduces an offset setting scheme which only allows bursts to leave an additional sending queue in the ingress node with a negative-exponentially distributed interarrival time. By doing so, traffic is shaped in order to control the outgoing rate as well as to smooth the traffic characteristics. A disadvantage of this scheme is the fact that bursts may have to wait at the network ingress although enough resources are available in the network. Furthermore, possible achievable multiplexing gain is small as an assembly queue cannot exceed its rate. However, this additional

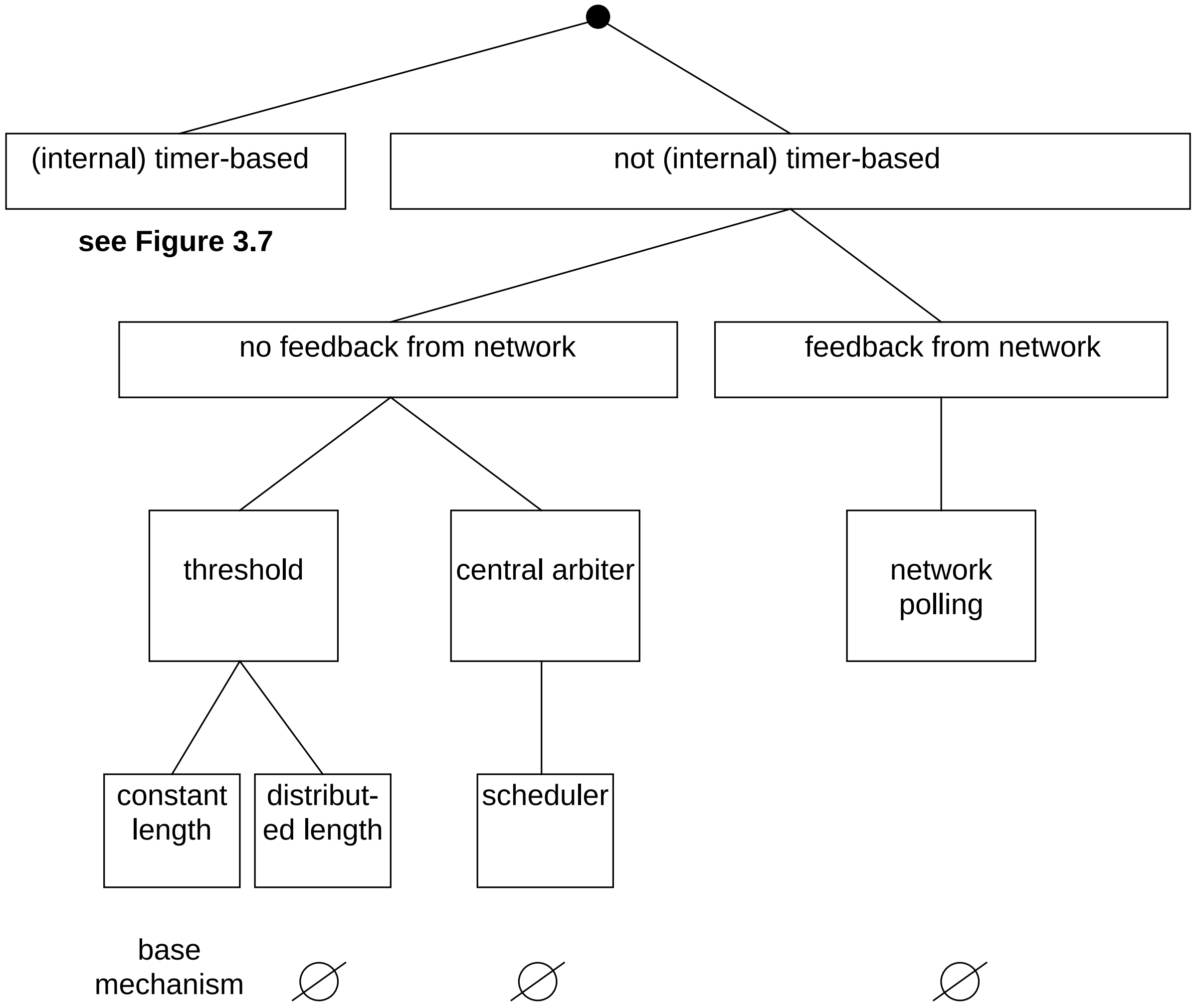


Figure 3.8: Classification of burst assembly mechanisms – not timer-based

traffic engineering functionality is not directly related to the burst assembly process as it is carried out later and independently.

A more sophisticated mechanism also considers the level of an assembly queue. If the level reaches a certain threshold prior to a timeout, a burst containing all packets of the respective queue is assembled and the timer is reset. Comparable to the mechanism which is only timer based, the threshold can be variable and, e. g., follow a distribution. However, no suggestion for non-constant threshold exists in literature. In [57], a time/threshold-based mechanism with constant time and constant threshold is proposed.

The next two categories in Figure 3.7 also incorporate some sort of feedback from the network in their assembly decisions. This is new and – besides own contributions – not reported in literature. The third category additionally considers the network state which is assumed here to be directly signaled back to an assembler on a coarse time granularity, or when the network state changes significantly. Dependent on this feedback, the burst assembly mechanism can react by adapting the rate and/or the resulting burst length.

The last category depicted in Figure 3.7 also considers feedback from the network which is not direct but indirect via a (coarse grained) bandwidth reservation for every assembly queue. As will be introduced in Section 6.3, feedback from the network is contained in a reservation envelope to which the burst assembler adapts its outgoing rate. In order to achieve multiplex gain between connections, some mechanism allows to send non-reserved bursts.

Burst assembly mechanisms which are not timer-based are classified in Figure 3.8. Comparable to Figure 3.7, they can also be further classified in mechanisms which consider feedback from the network and those which do not. The only mechanism which can be found in literature only considers a constant threshold of the assembly queue. Nevertheless, as these mechanisms cannot guarantee any upper bound for the waiting time in the assembly queue, it has no practical relevance and is only used as base component for burst assembly mechanisms.

Another assembly mechanism could be thought as a central arbiter which triggers all assembly queues in an edge node, e. g., according to a scheduling algorithm. Finally, another conceivable mechanism which is not timer-based but reacts on feedback from the network is polling. Hereby, the network triggers the assembly node when to send a burst.

An ongoing discussion in the research community is the question whether burst assembly reduces the self-similarity of the traffic. [57] states that self-similarity can be reduced whereas in [154], [94] and in [66], [65] it is argued that self-similarity is not reduced. [65] proves this by wavelet transformation [1]. This can be explained as burst assembly does not work like a traffic shaper which controls the amount of Bytes per time unit which leave the shaper. Instead, the arriving information to an assembly buffer is only delayed until a timer expires or enough information has arrived. The output rate from an assembly queue is hereby not considered. As a consequence, bursts with smooth traffic characteristics cannot be taken as basis for performance evaluations, or, if the self-similarity should be reduced, traffic shaping has to be included into the assembly process.

3.3 Reservation Mechanisms Supporting one Service Class

Recently, several so-called separate control delayed transmission (SCDT)-based reservation mechanisms have been proposed. They can be distinguished based on their way of indicating the end of a burst and the time when allocation of a WDM channel starts [43]. In the following Subsections, the classification depicted in Figure 3.9 and published in [43] is described in detail. Hereby, the difference between those reservation mechanisms is explained by Figure 3.10 where for every reservation mechanism a scenario of one wavelength with some already reserved bursts and three bursts called burst n, burst n+1 and burst n+2 which compete for a reservation is depicted.

Performance evaluations of those reservation mechanisms are presented in Chapter 5.

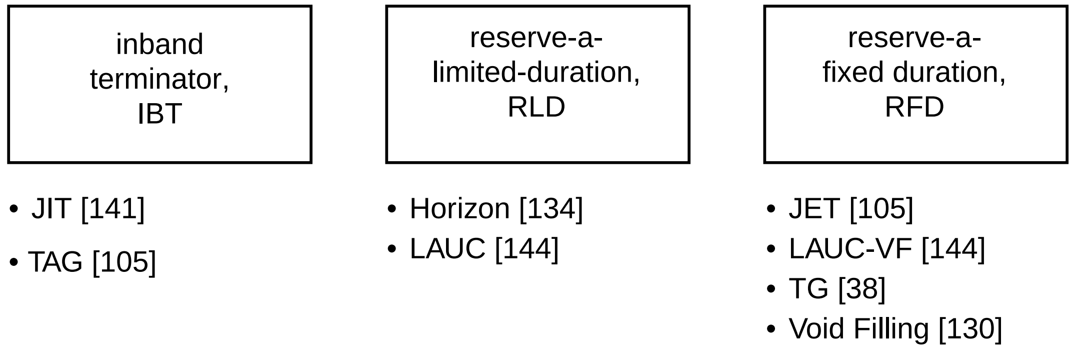


Figure 3.9: Classification of reservation mechanisms

3.3.1 Inband Terminator

A rather simple approach is to indicate the end of a burst by an additional trailing control packet or using an in-band terminator, IBT. In both cases there is no information about burst length when the heading control packet containing the reservation request arrives. Mechanisms which follow that principle are just-in-time, JIT, reservation [141] and tell-and-go, TAG [105] which are very similar if not the same. Upon arrival of the reservation request a wavelength channel is immediately allocated if available. Otherwise, the request is rejected and the corresponding data burst is discarded. The wavelength channel remains allocated until burst transmission has finished. The only information which has to be kept record of in network nodes is whether a wavelength channel is currently available or not. This makes JIT and TAG light weight approaches with low complexity in both edge and core nodes. The drawback is, however, its reduced efficiency as losses also occur in cases without transmission conflict between different bursts on the same wavelength.

This scenario is depicted in Figure 3.10a. Whereas it is obvious that burst n cannot make a reservation on that wavelength, burst n+1 is discarded although there would be no contention. Burst n+1 has to be discarded at the time of arrival of the control packet as the end of the actual burst on that wavelength is not known.

3.3.2 Reserve-a-Limited Duration

An improvement to schemes classified as IBT can be achieved by using reserve-a-limited-duration, RLD, based mechanisms. They require the sender to add the burst length in the control packet. A wavelength channel is only allocated for a limited duration so that subsequent burst transmission requests with a start time greater than the finishing time of an allocated burst may be accepted. This means the basic offset interval of a burst may overlap the transmission phase of a previously accepted burst. Thus, the burst n+1 in Figure 3.10a which is discarded by an IBT mechanism can be accepted by an RLD mechanism (Figure 3.10b). However, as also

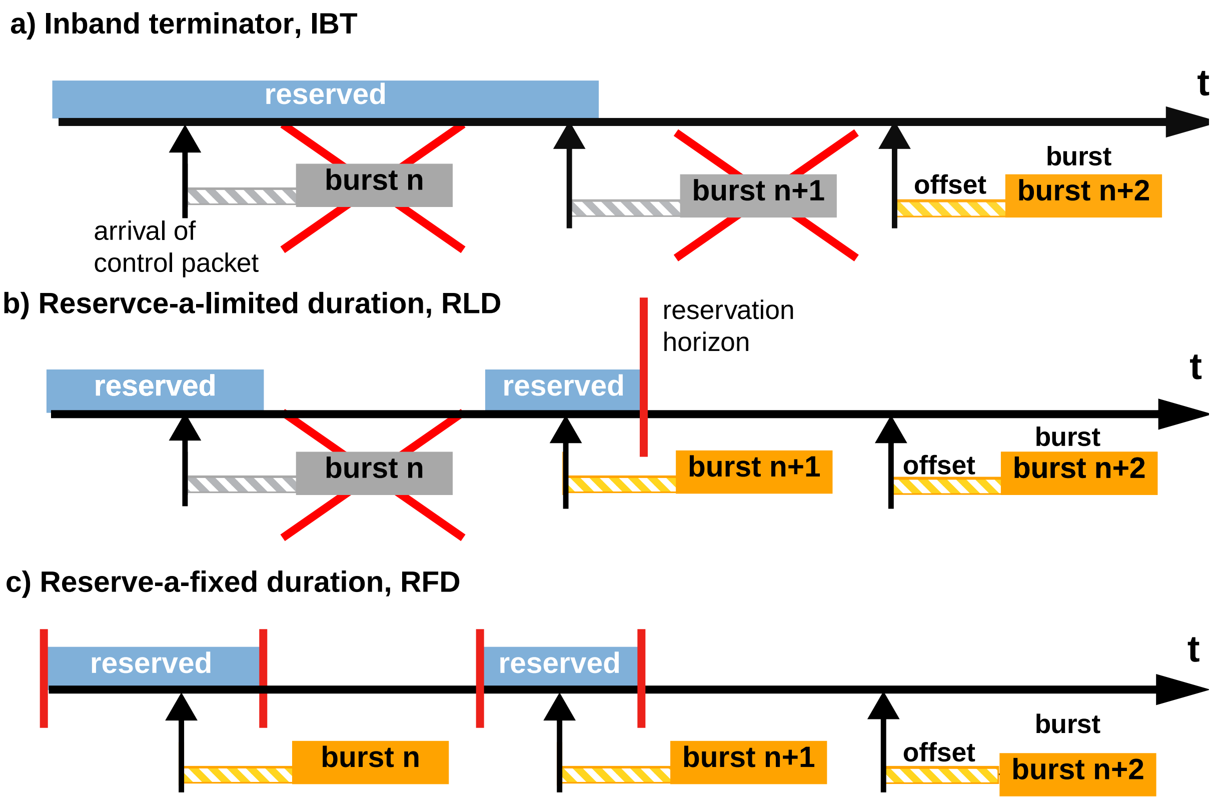


Figure 3.10: Scenarios for one-class reservation mechanisms

depicted in this figure, burst n cannot be accepted in a gap between two already reserved bursts as the reservation mechanism is not aware that this gap exists.

The Horizon mechanism proposed by Turner in [134] as well as the very similar Latest Available Unscheduled Channel mechanism, LAUC proposed by Xiong et. al. in [144] are representatives of RLD-based mechanisms. In Horizon and LAUC wavelength channel state information is enhanced by the so-called reservation horizon, i.e., the time until which the wavelength is allocated. When a new request arrives an available wavelength with a reservation horizon less than the start time of the new burst is looked for. The difference between Horizon and LAUC is that the minimization of the emerging gap is only an option in Horizon, whereas it is mandatory for LAUC. Like in IBT mechanisms, reservation starts immediately upon arrival of the control packet and lasts until the indicated end of burst transmission, which is the new reservation horizon of the corresponding wavelength. This makes RLD-based reservation mechanisms light-weight and efficient.

3.3.3 Reserve-a-Fixed Duration

Even higher efficiency may be achieved if start times of burst transmissions are also considered for reservation, i.e., reservation does not begin immediately when a request arrives but is delayed by the offset. This approach is called RFD, reserve-a-fixed-duration, as the channel is allocated for a fixed duration corresponding to the burst transmission time. Proposal of RFD-based reservation mechanisms are just-enough-time, JET, developed by Qiao and Yoo [106], [104], Latest Available Unscheduled Channel With Void Filling, LAUC-VF by Xiong et. al. [144], Tell & Go, TG, by Detti and Listani [38] and Void Filling by Tancevski et. al. [130]. JET can be considered as the basic algorithm which is slightly modified in the other algorithms.

Tell & Go, TG, is a simplified version of JET (an option of JET) where no offset exists between control and data. In order to compensate processing times, an FDL compensates header processing at every node, see also Figure 3.2b.

LAUC-VF enhances JET by first trying to reserve bandwidth for a burst in a gap/void left over from bursts and only if this fails trying to reserve a wavelength which is not yet segmented. In Void Filling, an algorithm even sorts and eventually delays bursts in order to minimize gaps on a wavelength.

State information comprises both, starting and finishing times of all accepted bursts, which makes such a system rather complex. On the other hand and in contrast to Horizon and LAUC, RFD-based mechanisms are able to detect situations where no transmission conflict occurs although the start time of a new burst is prior to the finishing time of the already accepted burst, i.e., burst n can be transmitted in between two already reserved bursts, see Figure 3.10c. Hence, bursts can be accepted with a higher probability than in Horizon and LAUC especially in case of large offset time variation, see Section 5.1.

3.4 OBS-QoS Mechanisms

An evolving questing in the context of IP-over-WDM is whether the optical layer can provide service differentiation as service to the IP layer and thus plays the role of a convergence layer. Therefore, OBS has to be enhanced, as yet, all reservation mechanisms discussed in Section 3.3 only support one service class. However, the ability of service differentiation of at least a small number of service classes is crucial for the application of OBS in optical transport networks. It allows for control of traffic which is the basis for sophisticated networking (traffic engineering, VPNs, ...) and also for charging and thus sophisticated marketing strategies. In order to enhance OBS to also support service differentiation, three major challenges are faced [41], namely

1. Limited time for burst header processing in core nodes.
2. No buffers in the core (beyond FDLs) to carry out scheduling.
3. No feedback about network status to the edge nodes in case of one-pass reservation.

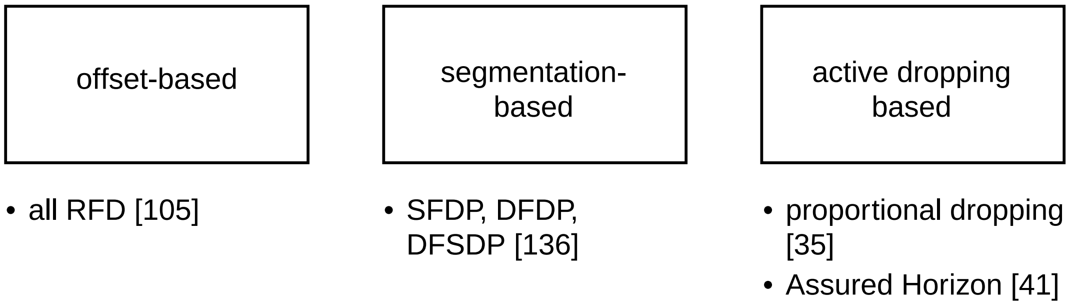


Figure 3.11: Classification of OBS-QoS mechanisms

While the first challenge is compensated by electronically processing the BHP and delaying data by a constant offset, the two latter post an outstanding problem. The requirement is to find an algorithm to schedule bursts to outgoing WLs. Hereby, without relying on buffers, isolation between FECs or service classes has to be achieved. Furthermore, the third challenge requests for an answer how to control possible overload that can significantly degrade the QoS.

In order to find an appropriate solution to overcome some of the aformentioned challenges, different OBS-QoS mechanisms have been proposed. Figure 3.11 presents a classification of those reported in literature which will be discussed in the following Subsections.

3.4.1 Offset-Based OBS-QoS Mechanisms

Offset-based OBS-QoS mechanisms add an additional offset between control information and data, called QoS offset to the basic offset which compensates processing times of the BHP. Depending on the priority of service class, the duration of such a QoS offset is varied. Hereby, in contrast to intuition, higher priority classes have a greater offset. This offset-based OBS-QoS mechanism is proposed by Qiao and Yoo in [105] for JET, however, any RFD-based one-class reservation mechanism can be applied.

In Figure 3.12, a scenario with three wavelengths containing some reserved bursts each is depicted. Herein, a low priority burst with no additional QoS offset and a high priority burst with a QoS offset try to make a reservation on one of these three wavelengths. As can be seen, the burst with the greater offset is able to reserve resources in advance to the low priority burst and thus can make a reservation whereas the low priority burst cannot. In general, this results in a lower burst loss probability than that of lower priority classes which tend to fill gaps (voids) of higher priority bursts.

Borrowing books from a library can be considered as example to further explain this prioritizing mechanism. The number of wavelengths of a link corresponds to the number of copies of a book owned by the library. At the time a book is borrowed, the time when the book is returned

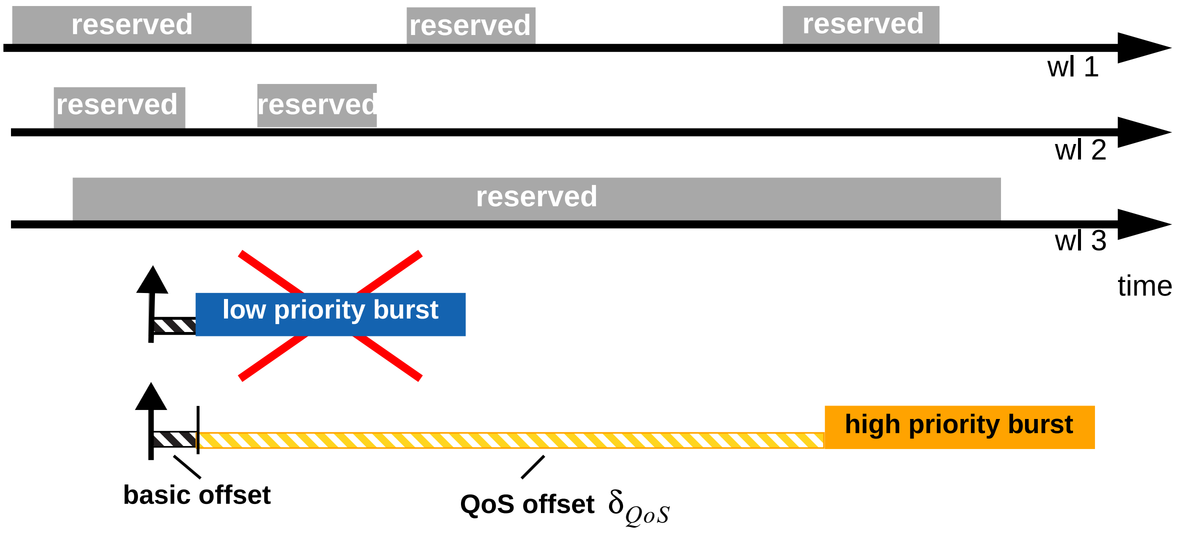


Figure 3.12: Offset-based OBS-QoS mechanism

is fixed. Prioritization is realized by introducing an offset time between reservation request and reservation, e. g., a high priority group (e. g., professors) always reserve a book one week in advance whereas a low priority group (e. g., students) make a reservation request for the same day. Consequently, students tend to only be able to reserve books which are left over from professors. If the mean borrowing time of a book is shorter than the additional QoS offset (than a week), most of the books borrowed by the students are returned during the offset and thus are available again for professors. In an extreme, professors do not realize students at all and only compete for books among themselves.

A detailed modelling and performance evaluation of the offset-based OBS-QoS mechanism is presented in Section 5.2, however, from this example, several system characteristics are immediately obvious:

- The ratio between QoS offset and mean holding time of lower priority classes determines the degree of service differentiation.
- Bursts of higher priority classes have a longer waiting time prior to being served. The greater the mean burst size of low priority bursts, the longer the waiting time for higher priority classes in order to obtain the same degree of differentiation.
- Higher priority bursts segment wavelengths. As a result, lower priority bursts tend to reserve only the gaps left over by higher priority bursts. Thus, shorter low priority bursts have a lower burst loss probability than longer low priority bursts as they have a higher probability to fit into those gaps. This is contradicting to control overhead which is low when low priority bursts are long.

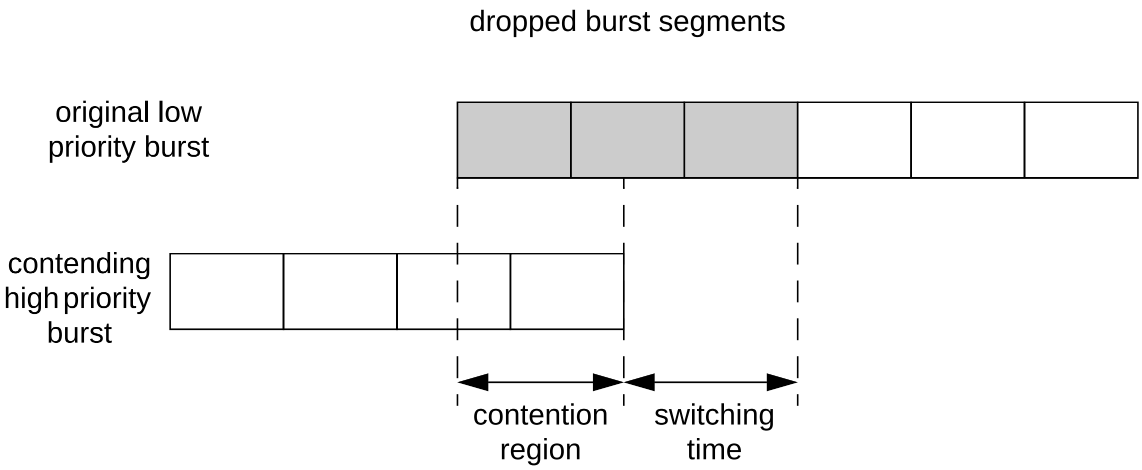


Figure 3.13: Segmentation-based OBS-QoS mechanism [137]

- This scheme is non-preemptive. As long low priority bursts can block wavelengths, no complete isolation is achieved. On the other hand, starvation of low priority classes is possible if the offered traffic load of high priority bursts is not controlled.

3.4.2 Segmentation-Based OBS-QoS Mechanisms Based

Mechanisms classified as segmentation-based give up the principle that a burst is an atomic unit and subdivide it in several independent segments. Independence is bought at the price of an extra header for every segment which at least needs to contain routing information, the burst length, segment number and a type field. The segment size compromises between the amount of lost information in case of contention and the overhead (of header information) per burst.

In case of contention in the network, some segments[1] of a burst are either discarded or deflected whereas the remaining part of the burst can still be delivered to the egress. Compared to a solution with the granularity of whole bursts, less bytes are lost. In Figure 3.13, an example is depicted where the tail of an original low priority burst is discarded in order to be able to carry the entire contending high priority burst.

Strategies which realize discarding segments of a burst can be differentiated in head or tail discarding. Head discarding requires the adaption of the offset between header and start of burst. On the other side, tail discarding has to compensate the problem that the header of a burst with discarded tail has already been forwarded to the next node. For reservation of this burst, the next node assumes the wrong (original) length which may result in another contention which is no contention in reality as the burst has been shortened earlier.

[1] The lost segments do not only use up the contention region, but also the switching time, see Figure 3.13. Thus, a slower switching technology results in a greater amount of lost segments.

OBS-QoS mechanisms which realize a segmentation-based scheme are proposed in [137], [136] and evaluated in [123]. Here, a modified tail discarding mechanism is proposed which only discards the tail of a burst if the total size of the contending burst is greater than the segments to be discarded. In addition to discard segments, deflection of segments is proposed. Deduced from these options, a variety of different strategies can be found that differentiate whether a burst is deflected first or segmented first and, in case it is segmented, whether the segments are deflected or discarded.

Without modelling or analysis of such mechanisms, the following characteristics are immediately obvious:

- Overhead in Bytes is introduced by additional burst headers for every segment.
- Overhead in signalling is introduced in the core in order to inform successing nodes that a burst was shortened.
- Greater networks result in smaller mean bursts, as low priority bursts tend to keep loosing segments on their way towards the network egress.
- Smaller transport units require faster switching technology or are less efficient as switching times get more significant, see Section 3.1.3. In contradiction to this, the avoidance of small transport units is one of the major drivers for OBS.
- Depending on the strategy, complete isolation of the highest priority class can be achieved. However, as this scheme is preemptive, it can result in starvation of lower priority classes.
- A major disadvantage of segmentation-based OBS-QoS is the increase of processing complexity in the core as a node has to determine what to do in case of a contention and also manipulate bursts on the data path. This stands in contrast to the assumption of transparency and simplicity of OBS.

Therefore, segmentation-based OBS-QoS mechanisms will not be considered throughout the rest of this thesis.

3.4.3 Active Dropping-Based OBS-QoS Mechanisms

Mechanisms which are based on active dropping implement a burst dropper in front of every core node. Dependent on a dropping policy, some BHPs and their associated data are dropped prior to reaching the reservation unit. Thus, the dropping policy represents an admission control policy for outgoing wavelengths, see Figure 3.14. By doing so, the offered load of a service class can be controlled locally by every core node in order to maintain bandwidth on wavelengths for other service classes. Accordingly, active dropping-based OBS-QoS mechanisms prophylactically drop bursts and hence intervene before congestion occurs. In order to

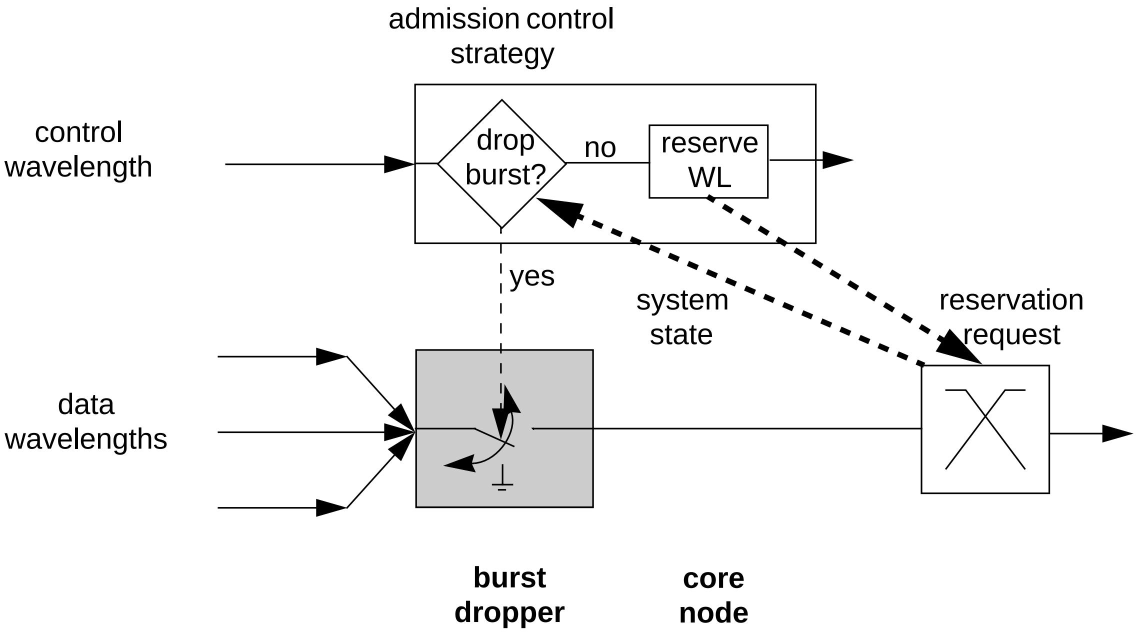


Figure 3.14: Active dropping-based OBS-QoS mechanism

perform the admission control fair, the burst length has to be known prior to dropping. This is especially true as bursts originated from different service classes may have different traffic characteristics like, e. g., different mean burst length, it is not sufficient to only consider the number of dropped bursts. Accordingly, RLD as well as RFD mechanisms are suitable for active dropping based OBS-QoS mechanisms.

Figure 3.14 depicts a burst dropping unit performing admission control in front of a core node. Based on information the dropping control receives, e. g., from measurements, statistics on BHPs, or from the reservation mechanism of the core node, it determines whether a BHP and its corresponding burst is being dropped.

An example of a dropping scheme is called ‘intentional burst dropping’ and is reported in [35]. Hereby, the dropping policy realizes a proportional differentiation model[1]. The burst loss probabilities of different service classes are kept proportional according to

$$\frac{q_i}{q_j} = \frac{s_i}{s_j} \tag{3.1}$$

where q_k is the QoS metric and s_k the differentiation factor for class k.

Without modelling or analysis of such a mechanism, the following characteristics are immediately obvious:

- In general, burst loss (drop) probabilities do not depend on burst characteristics

[1] Such a model is known from the DiffServ architecture (Section 2.2.2) where a scheduler realizes proportional QoS. For an comprehensive overview, see [18].

- Depending on the admission control strategy, the realization can be very simple
- Burst loss probability is (slightly) higher compared to another class of OBS-QoS mechanism as a case can occur where a lower priority burst is actively dropped in order to maintain bandwidth for a higher priority burst, but no higher priority burst arrives.
- Relative burst dropping implies that no absolute guarantees can be made as increased offered traffic of any class may result in increased burst dropping for all classes.
- Active burst dropping is non-preemptive.

As shown by simulations in [35], bursts of different classes are always dropped according to the desired ratio. Furthermore, an advantageous characteristic of this mechanism is that losses do not depend on traffic characteristics and hence, can be completely controlled. These characteristics, together with its simplicity, make such an OBS-QoS mechanism very promising.

However, a major disadvantage of this scheme is that no feedback is provided from the network core to the edges and thus traffic volume of different classes cannot be controlled. Furthermore, isolation between classes cannot be guaranteed. If the traffic volume of a low priority class is significantly increased, and as a consequence the overall burst loss probability rises, burst loss probabilities of all classes are increased. Therefore, such a scheme requires additional traffic control mechanisms.

Another example of dropping-based OBS-QoS is the Assured Horizon framework which is published in [41] and [40]. This framework forms the core of this thesis, will be described in detail in Chapter 6 and modelled as well as evaluated in Chapter 7.

3.4.4 Additionally Conceivable Mechanisms

Besides the OBS-QoS mechanisms classified in Figure 3.11, additional mechanisms are conceivable. These mechanisms are based on assumptions which are not valid in general for all OBS systems. However, for reasons of completeness, they will be shortly mentioned in the following.

Differentiation through buffering

If an OBS system makes use of sophisticated buffering, well-known concepts of electronic scheduling and buffer management can be applied. Even if only simple FDL buffers are used, differentiation of service classes with respect to burst loss probability can be obtained by controlling access to these buffers.

Deflection routing

As already indicated in Section 3.4.2, deflection routing can be carried out selectively for different service classes.

Protection and restoration

Protection and restoration schemes and thus differentiation with respect to reliability and service availability can also be applied selectively for specific (various) service classes.

Chapter 4

Teletraffic Fundamentals on Loss Systems

Because of its generality, the theory of loss systems is today still applied as basis to dimension resources according to blocking probabilities. In such a loss system, requests arrive to a system with n servers. In case any server is idle, it is occupied by an arriving request for a generally distributed holding time, corresponding to complete sharing of all servers. If all servers are occupied at the time a request arrives, the request is discarded[1].

Dependent on the modelling of a real system, a request may be a connection setup request in a (mobile) telephone network and the service time corresponds to the duration of the call, or the arrival of a burst header packet is modelled as request and the transmission time of a burst as service time. Both applications of the theory of loss systems have in common that requests are not buffered and several servers are available to service requests. In case of OBS nodes, this is due to the fact that buffers are not mandatory in the core and wavelengths can be modelled as servers.

In the following, major teletraffic fundamentals on the M/G/n loss system as representative of classless systems as well as multi-class systems with admission control strategies are reviewed in Section 4.1 and Section 4.2, respectively.

4.1 M/G/n Loss Systems

The M/G/n loss system is a model in teletraffic theory where requests arrive according to a negative-exponentially distributed interarrival time at a system with n servers. Requests are

[1] According to [118], this model description is identical to the dimensioning of a stochastic knapsack.

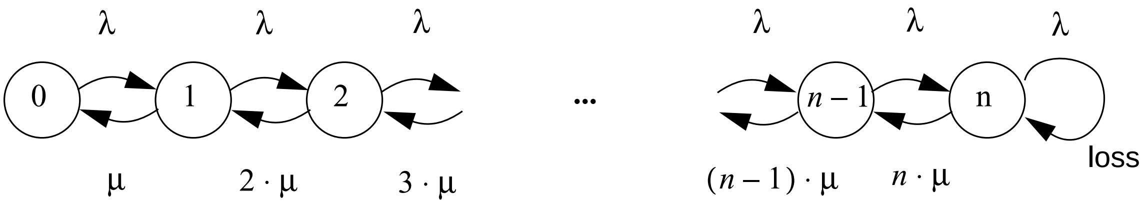


Figure 4.1: State transition diagram of a one-dimensional loss system

not distinguished in this model, i. e., it only covers one service class. In such a loss system, no waiting time is caused as there are no buffers. Instead, the loss probability B is of interest.

The loss probability can be obtained by calculating the state probabilities of a one-dimensional Markov chain. Although, for this calculation, a Markovian holding time distribution has to be assumed, it can be shown that the result is also valid for generally distributed holding times, see, e. g., [91] and [88]. Figure 4.1 depicts a state transition diagram of a loss system with n servers where requests arrive with an arrival rate λ. An occupied server terminates with rate $\mu = 1/h$ where h denotes the mean service time.

From Figure 4.1, the steady state system equilibrium probabilities p_i can be obtained solving

$$p_{i-1} \cdot \lambda = p_i \cdot i \cdot \mu \qquad i = 1 \ldots n \tag{4.1}$$

and applying the normalization

$$\sum_{i=0}^{n} p_i = 1 \tag{4.2}$$

Hereby, according to the PASTA-theorem (Poisson Arrivals See Time Averages), the loss probability B equals the probability that the system is in state n. B is described by the famous Erlang-loss formula that has been derived in 1917 [49], [50] (also called Erlang-B-formula).

$$B = p_n = E_{1,n}(A) = \frac{A^n/n!}{\sum_{i=0}^{n} A^i/i!} \tag{4.3}$$

Hereby, $A = \lambda/\mu$ denotes the traffic which is offered to the system, and A/n the normalized offered traffic. As this formula is of central interest, the loss probability is evaluated depending on its parameters A and n.

Figure 4.2 depicts B against the number of servers n with A/n as parameter. It can be seen that a greater number of servers yields an exponential decrease in B, e. g., if n is increased from 40 to 160 in a scenario with $A/n = 0.6$, B is decreased by about six orders of magni-

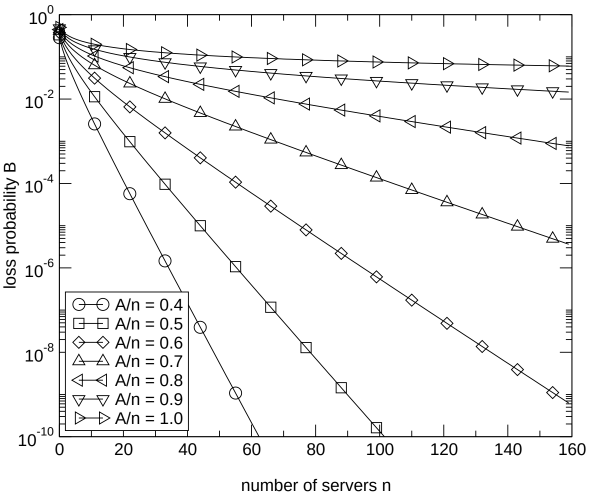


Figure 4.2: Loss probability B against number of servers n

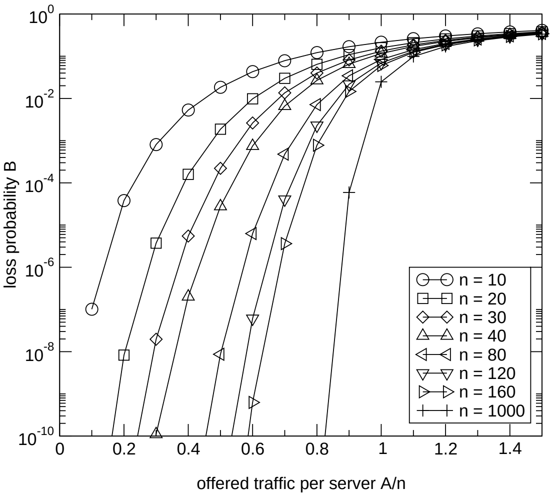


Figure 4.3: Loss probability B against normalized offered traffic A/n

tude. This reduction of B is called economy of scale as the greater the number of servers, the greater the probability that a request can find a server which is not occupied at the moment. The slope of the decrease depends on A/n. If A/n approaches 1, the slope is very small resulting in an only hardly reduced loss probability with increasing number of servers. Thus, the increase of n from 40 to 160 results in a loss probability which is only halved in case $A/n = 1$.

In Figure 4.3, the loss probability B is depicted against A/n with the number of servers n as parameter. In this graph, also overload situations with $A/n > 1$ are depicted. It can be seen that an increased A/n also yields an increased loss probability. For greater overload, B approaches roughly the same value despite the number of servers, i. e., a greater n does not yield a lower B in greater overload situations. For smaller values of A/n it can be seen that a reduced A/n results in a decrease in B. The slope of this decrease depends hereby from the number of servers and can be again explained by the economy of scale.

Summarizing this short discussion, a low loss probability can only be achieved by a large number of servers and A/n which is controlled far below 1. This proposition is central for the design of Assured Horizon in Chapter 6.

4.2 Multi-Class Loss Systems

The consideration of different service classes leads to a multi-dimensional Markov chain. Hereby, comparable to the previous section, the interarrival time distribution of requests as well as the holding time distribution of a server is assumed to be negative-exponential in this section.

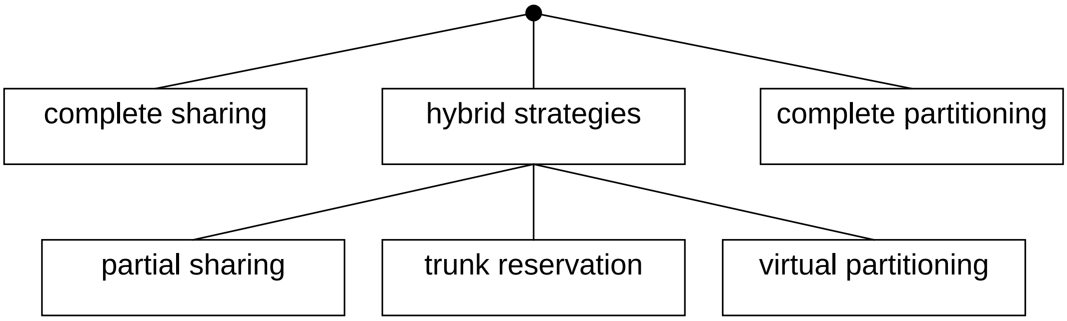


Figure 4.4: Classification of admission control strategies

In Figure 4.5, an example is depicted distinguishing two classes in a system with 4 servers where requests always require one server. In this example, all requests are allowed to allocate a server as long as one is available, i. e., no admission control mechanism is applied. This strategy is known as complete sharing of resources. One major disadvantage of this strategy in the here considered scenario where all requests have the same bandwidth requirement is that no protection between classes is possible as one class can consume all system resources (system states $p_{4,0}$ and $p_{0,4}$).

In most cases, a differentiated treatment of service classes is required in order to obtain different loss probabilities. In practice, an admission strategy controls the number of requests per service class [114], [67], [118]. Figure 4.4 presents a classification of admission control strategies. The opposite of complete sharing is complete partitioning where a fixed number of servers is allocated exclusively for a certain service class. As no class is allowed to use temporarily unused servers of other classes, no multiplexing gain between classes is possible in such a system. For calculation, it is simply a one-class loss system per service class and thus can be solved with (4.3) obtained for the M/G/n loss system considering the reduced number of servers.

However, more interesting are policies where some resources are shared while a certain amount is exclusively reserved to a class. Those policies which are denoted in Figure 4.4 as hybrid strategies are partial sharing, trunk reservation and virtual partitioning.

In *partial sharing* admission control strategies, a part of the resources is exclusively reserved per service class whereas the remaining resources can be reserved by all classes on a first come first served, FCFS, strategy. A large subset are strategies which are coordinate convex (departures are never blocked, see also Figure 4.6). If additionally all transitions come in pairs, the state probabilities can be obtained by a product form solution [118]. In [86] the validity of the product form solution was extended to holding time distributions with rational Laplace transform. In the example with an overall number of 4 servers depicted in Figure 4.6, one server is exclusively reserved per service class whereas the remaining 2 servers can be occupied by any

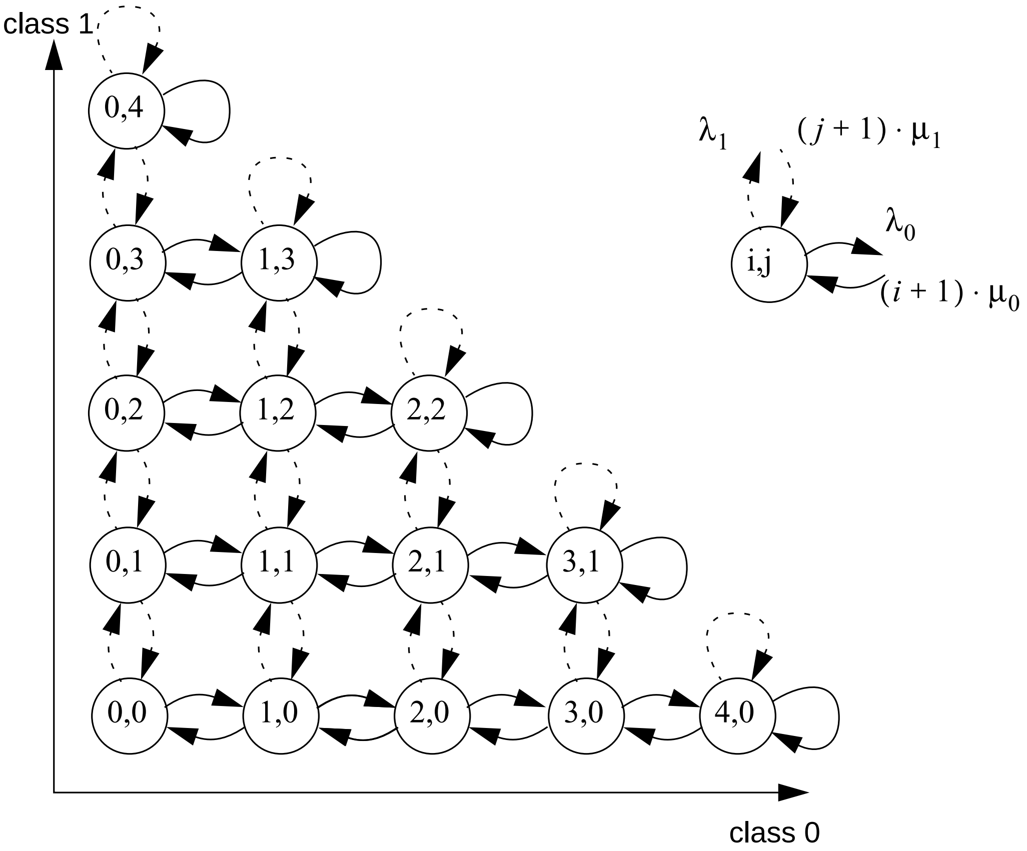


Figure 4.5: State transition diagram of a two-dimensional system with four servers with complete sharing admission control policy

of the two service classes. Therefore, in Figure 4.6, transitions to the system states $p_{4,0}$ and $p_{0,4}$ are not allowed and these two states cannot be reached. For reasons of clarity, transitions indicating a loss (which start and end at the same transition) are omitted.

In *trunk reservation* admission control strategies [114], [115], a request of class i is only admitted if not more than a class-dependent threshold q_i number of servers are currently occupied by requests of any class. Hence, service differentiation by a one-way protection of higher priority classes against lower priority classes is realized by rejecting requests of lower priority when the available number of resources in the system are less than a specified threshold. In literature, trunk reservation admission control is mainly applied in order to obtain the same loss probability for classes of requests with different bandwidth requirements. A detailed overview on research activities in the field of trunk reservation can be found in [132].

In [115], it is stated that the trunk reservation policy is generally better with respect to an objective function value (e. g., total reward, see Section 4.2.3) and stability (e. g., choice of the threshold q_i and number of calls in the system) than the best coordinate convex strategy. Furthermore, in [69] it is shown that under the assumption of same bandwidth requirements of all classes as well as same mean holding times, a policy maximizing the reward rate is a trunk res-

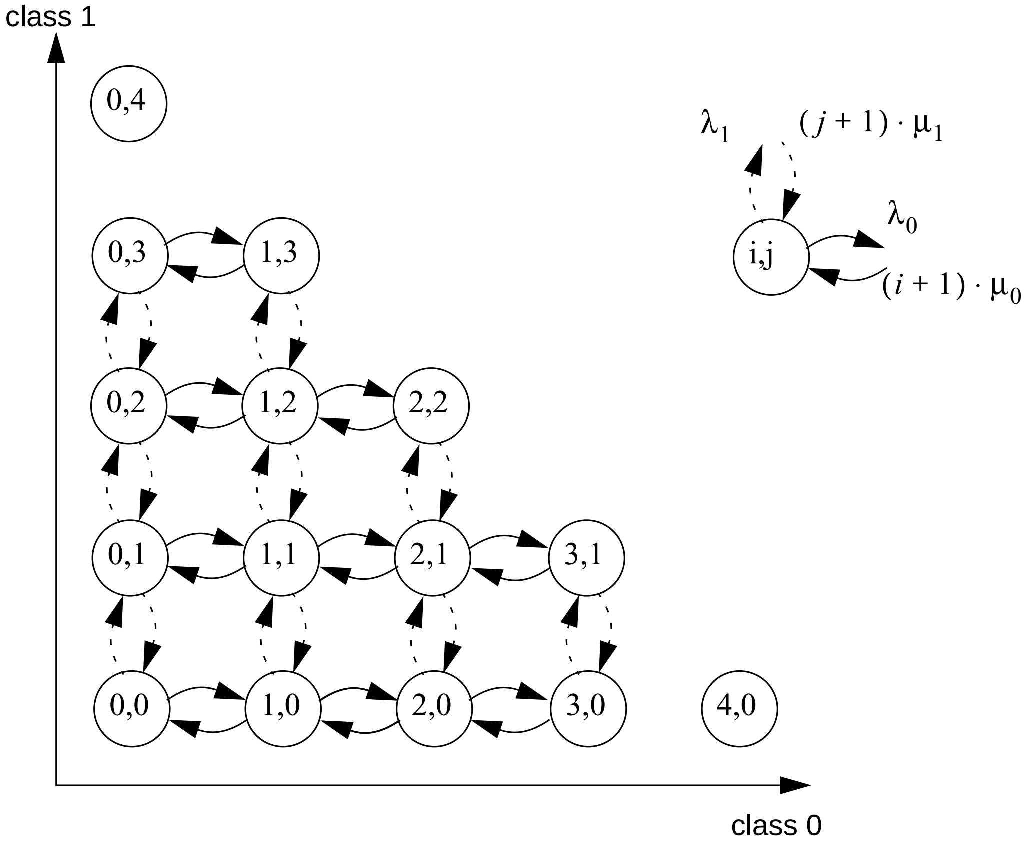


Figure 4.6: State transition diagram of a two -dimensional system with four servers with partial sharing admission control policy

ervation policy with multiple priority levels. Because of the just listed favorable attributes of trunk reservation, its theoretical background will be discussed in the following.

In Figure 4.7, the already known example with 4 servers is depicted for a trunk reservation admission control strategy with no access restrictions for class 0 ($q_0 = 4$) whereas class 1 requests are only admitted if not more than 2 servers are occupied ($q_1 = 2$). From Figure 4.7, it can be seen that not all transitions come in pairs and hence the product form solution cannot be applied. Formulæ for the loss probability are derived in Chapter 4.2.2.

V*irtual partitioning* strategies see, e. g., [20], extend trunk reservation strategies by a dynamic threshold $q_i(n_i)$ which depends on the number of currently accepted calls of class i. By doing so, relative guarantees for blocking probabilities can be realized, see also [18]. However, as detailed state information is required for this mechanism, it is not in the focus of this thesis and will not be considered in the following.

4.2.1 Formulæ for Loss Systems with Complete Sharing

The system state is described by the number of requests n_i of each class i. Thus, the multi-dimensional state space is of the order of the number of classes. In Figure 4.5, a state transition

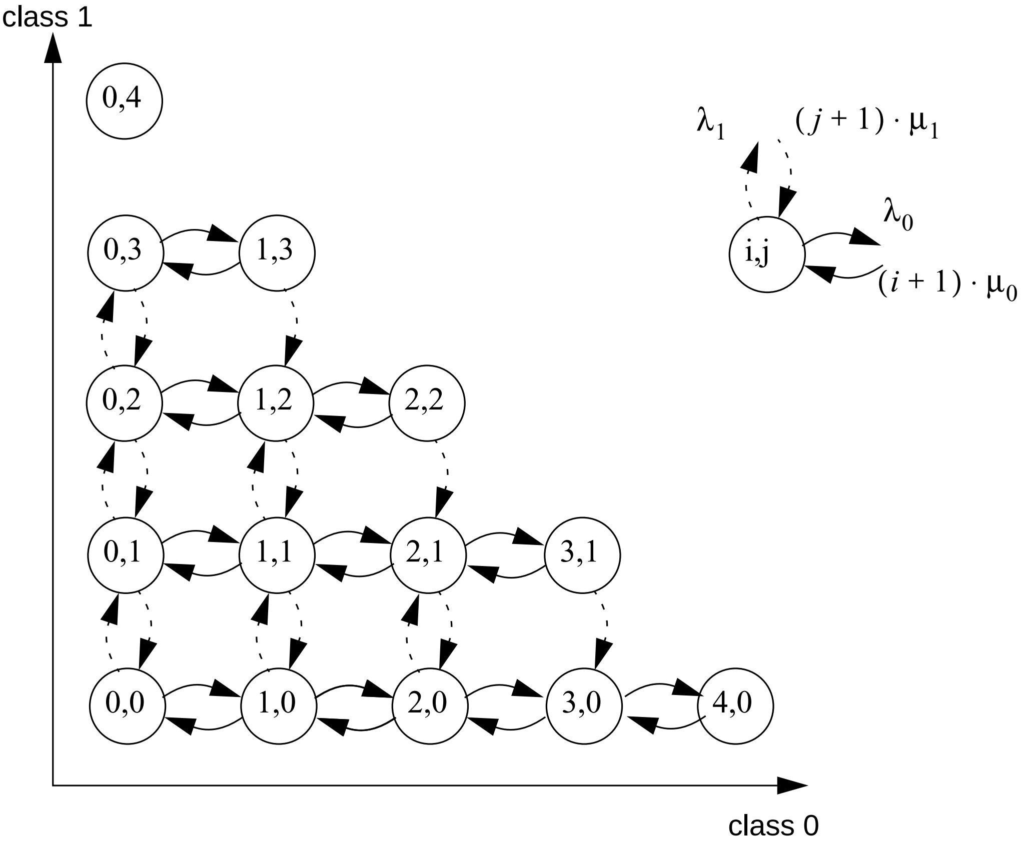


Figure 4.7: State transition diagram of a two-dimensional system with four servers with trunk reservation admission control policy

diagram is depicted for the case of two classes which require one server per request in a system with 4 servers. The probabilities of state can be either obtained by a product form solution or a recursive solution [132]. As the recursive approach is applied in Chapter 4.2.2, both approaches will be introduced here shortly.

Product form solution

Probabilities of state can be obtained according to

$$p(n_0, n_1, \ldots, n_{N-1}) = \frac{\prod_{i=0}^{N-1} \frac{(\lambda_i/\mu_i)^{n_i}}{n_i!}}{\sum_{n_1=0}^{\lfloor n/C_0 \rfloor} \cdots \sum_{n_1=0}^{\lfloor n/C_{N-1} \rfloor} \prod_{i=0}^{N-1} \frac{(\lambda_i/\mu_i)^{n_i}}{n_i!}} \tag{4.4}$$

Here, N denotes the number of classes, n the overall system capacity in number of servers and C_i the required number of servers per request of class i with $C_i \equiv 1$ in all scenarios. However, for generality, the formulæ are listed with C_i in the following. Derived from (4.4), for the blocking probabilities B_i of class i it follows

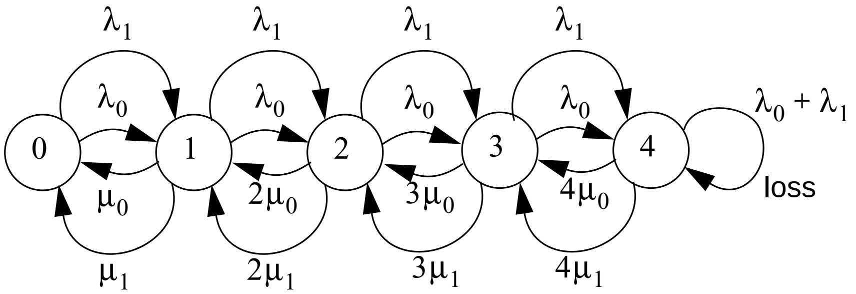


Figure 4.8: One-dimensional state transition diagram of a two-dimensional system with complete sharing admission control policy

$$B_i = \sum_{(n_0, \dots, n_{N-1}) \in S_i} p(n_0, n_1, \dots, n_{N-1}) \tag{4.5}$$

with

$$S_i = \left\{ (n_{0, \dots}, n_{N-1}) \middle| (n_i + 1) \cdot C_i + \sum_{\substack{j = 0 \\ j \neq i}}^{N-1} n_j C_j > n \right\} \tag{4.6}$$

Recursive solution

For a large number of classes and servers, the following recursive approach yields the results faster. Hereby, the multi-dimensional state space is mapped to a one-dimensional state space according to a proper bandwidth discretization [132]. Under the assumption already made in Figure 4.5 with a basic bandwidth requirement of one server, the reduced state space depicted in Figure 4.8 follows.

Unnormalized steady state probabilities can be calculated by

$$\tilde{p}(m) = \begin{cases} 1 & \text{for m = 0} \\ 0 & \text{for m < 0} \\ \frac{1}{m} \cdot \sum_{i=0}^{N-1} \tilde{p}(m - C_i) \cdot C_i \cdot \frac{\lambda_i}{\mu_i} & \text{for } 0 < \text{m} \leq n \end{cases} \tag{4.7}$$

Herby, m denotes the overall number of occupied servers.

Normalization yields

$$p(m) = \frac{\tilde{p}(m)}{\sum_{m=0}^{n} \tilde{p}(m)} \tag{4.8}$$

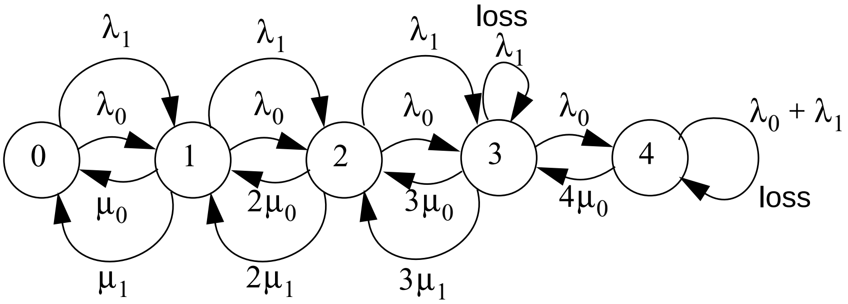


Figure 4.9: One-dimensional state transition diagram of a two-dimensional system with trunk reservation admission control policy

And finally, the blocking probabilities B_i for classes i can be obtained from

$$B_i = \sum_{m = n - C_i + 1}^{n} p(m) \tag{4.9}$$

In [116], it is has been shown that the results for B_i obtained by this recursive solutions are exact, also for generally distributed holding times.

4.2.2 Formulæ for Loss Systems with Trunk Reservation

Based on the formulæ for the strategy ‘complete sharing’ obtained in Chapter 4.2.1, [132] presents also a recursive solution for loss systems with trunk reservation admission control.

In loss systems with trunk reservation admission control, a request of class i is accepted if – upon arrival – C_i servers are available and not more than q_i servers are occupied. For a class which is not subject to trunk reservation (e. g., the highest priority class), $q_i = n$. For the approximation of probabilities of state, it follows:

$$\tilde{p}^*(m) = \begin{cases} 1 & \text{for m} = 0 \\ 0 & \text{for m} < 0 \\ \dfrac{1}{m} \cdot \displaystyle\sum_{i=0}^{N-1} \tilde{p}^*(m - C_i) \cdot C_i(m) \cdot \dfrac{\lambda_i}{\mu_i} & \text{for } 0 < \text{m} \le n \end{cases} \tag{4.10}$$

with

$$C_i(m) = \begin{cases} C_i & m \le q_i \\ 0 & m > q_i \end{cases} \tag{4.11}$$

In this context, the parameter $\theta_i = q_i/n$ is introduced which denotes the highest occupancy where a request of class i is still admitted. Again, normalization yields

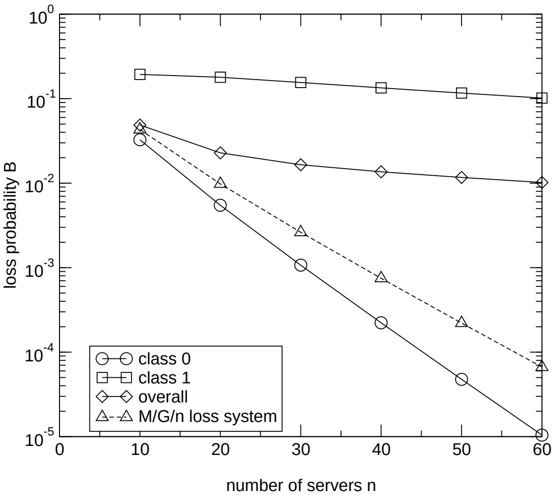


Figure 4.10: Loss probability against number of servers ($\theta_1 = 0.7$)

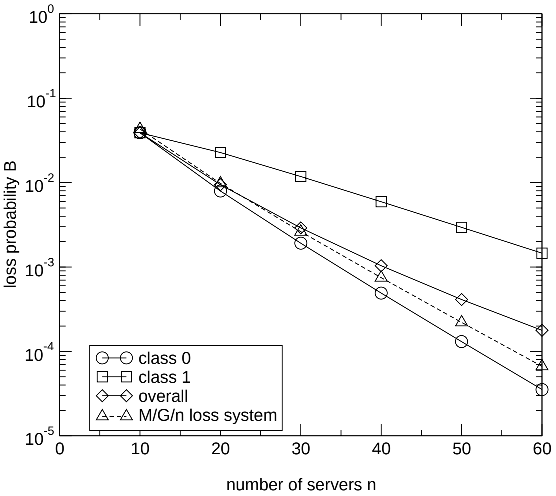


Figure 4.11: Loss probability against number of servers ($\theta_1 = 0.9$)

$$p^*(m) = \frac{\tilde{p}^*(m)}{\sum_{m=0}^{n} \tilde{p}^*(m)} \tag{4.12}$$

And finally, an approximation for the blocking probabilities B_i of class i follows

$$B_i^* = \sum_{m = \min\{n - C_i, q_i\}}^{n} p^*(m) \tag{4.13}$$

4.2.3 System Reward

In order to determine the thresholds q_i appropriately, class dependent rewards R_i are granted per admitted request in the system, see, e. g., [118], [69]. Thus, the overall reward R can be calculated according to

$$R = \sum_{n_0, \ldots, n_{N-1}} p_{n_0, \ldots, n_{N-1}} \cdot (n_0 \cdot R_0 + \ldots + n_{N-1} \cdot R_{N-1}) \tag{4.14}$$

Hereby, $p_{n_0, \ldots, n_{N-1}}$ denotes the probability of system state. If the rewards R_i are given, an optimized system is obtained by maximization of (4.14). For the following discussion, R is normalized by the maximum achievable reward $R_{\max}$ yielding

$$R_{\text{norm}} = \frac{R}{R_{\max}} \tag{4.15}$$

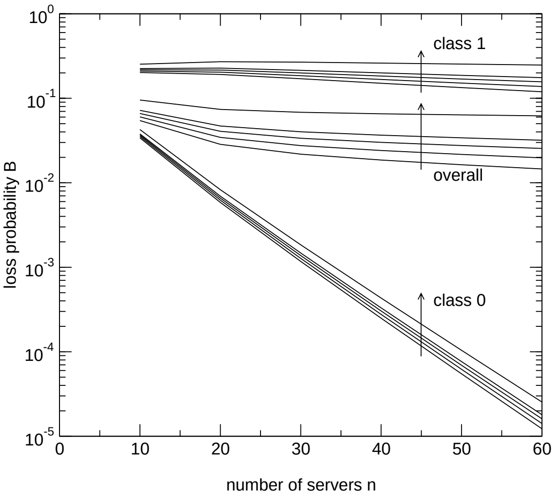


Figure 4.12: Loss probability against number of servers with increased A_1 (100% - 200%)

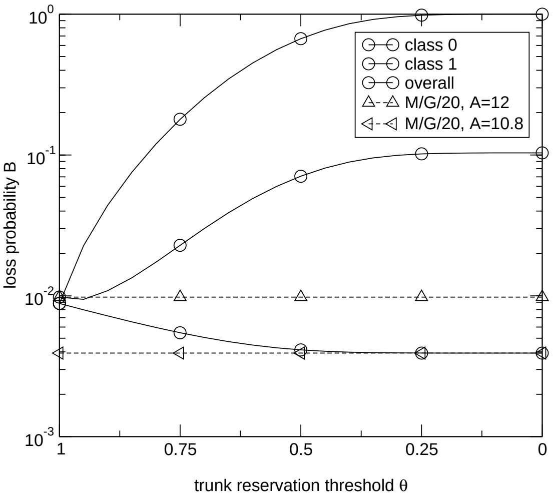


Figure 4.13: Loss probability against normalized trunk reservation threshold θ

4.2.4 Performance Evaluation of Loss Systems with Trunk Reservation

All following evaluations are carried out in a scenario of a loss system with n servers and $A/n = 0.6$ if not denoted differently. The offered traffic is generated from two classes with the same traffic characteristics, $A_0 = 0.9 \cdot A$ and $A_1 = 0.1 \cdot A$. Thus, in contrast to intuition, the amount of high priority traffic is greater than the amount of low priority traffic. Such a scenario is chosen because of the application in Chapter 6. Trunk reservation admission control is applied where requests of class 0 can use all servers whereas requests of class 1 are not admitted if more than θ_1 servers are currently occupied.

In Figure 4.10 and Figure 4.11, the loss probability of both classes as well as the overall loss probability are depicted against the number of servers. Additionally, for orientation, the loss probability of the classless M/G/n loss system is also depicted. In Figure 4.10, a greater degree of service differentiation is chosen ($\theta_1 = 0.7$) in contrast to $\theta_1 = 0.9$ in Figure 4.11.

From both figures, it can be seen that B_0 is lower than B_1 as a result of the service differentiation by trunk reservation admission control. Furthermore, B_0 is also lower than the loss probability in an M/G/n loss system. However, this improvement in performance of class 0 has to be paid by an increase in B_1 as well as the overall loss probability B_{all}. For a smaller value of θ_1, more requests of class 1 are blocked by the admission control mechanism and thus the overall loss probability does not decrease as fast as is would without admission control. This can be seen in Figure 4.10 where B_{all} only slightly decreases with increasing number of servers.

In Figure 4.12, the grade of service differentiation between class 0 and class 1 is evaluated. Like in the previous figures, the loss probability is depicted against the number of servers. In this graph, the amount of class 0 traffic is kept constant whereas the amount of offered traffic of

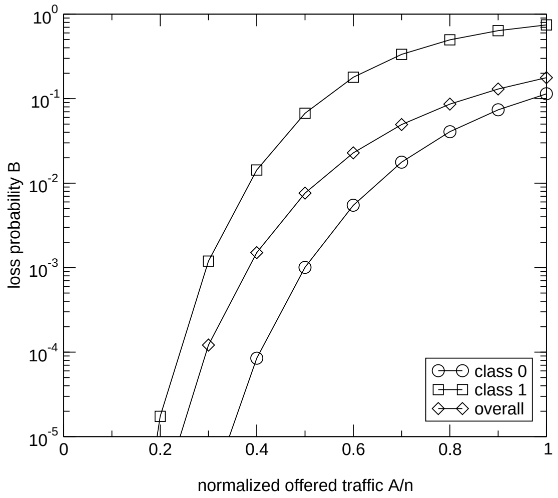


Figure 4.14: Loss probability against normalized offered traffic A/n $(\theta_1 = 0.7)$

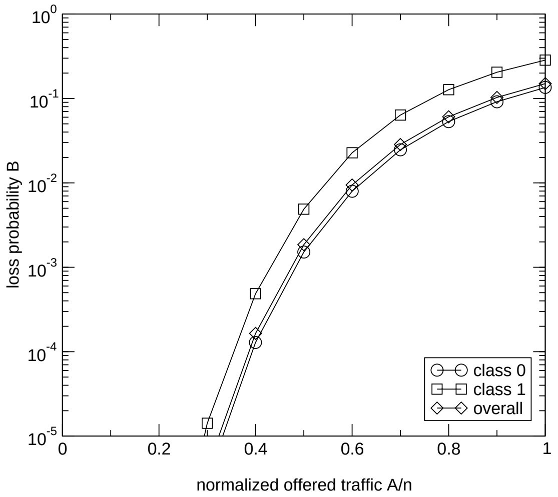


Figure 4.15: Loss probability against normalized offered traffic A/n $(\theta_1 = 0.9)$

class 1 (and thus the overall offered traffic) is increased. In Figure 4.12, A_1 is multiplied with factors 1, 1.25, 1.5, 1.75 and 2, respectively. Hereby, an arrow indicates the direction of the increase. It can be seen that an increase in A_1 also slightly influences B_0. This behavior can be also explained because of the small A_1 which has only a small impact on the overall system. Thus, as already explained earlier, no total isolation between classes is provided.

The compromise between lower loss probability of class 1 and higher overall loss probability as well as loss probability of class 1 is also visible in Figure 4.13. Here, the loss probabilities are depicted against θ_1 in a scenario with 20 servers. In case $\theta_1 = 1$ (requests of class 1 are allowed to use all servers, independent of the current occupancy and thus a complete sharing admission control policy is applied) both loss probabilities are the same and can be calculated by the M/G/n loss system. The other extreme, where $\theta_1 = 0$ (no request of class 1 is admitted) the loss probability of class 0 can be calculated from an M/G/n loss system with only class 0 traffic whereas B_1 equals 1. B_{all} also strongly increases as all class 1 request are blocked.

In Figure 4.14 and Figure 4.15, the loss probability is depicted against A/n for $\theta_1 = 0.7$ and $\theta_1 = 0.9$, respectively, in a scenario with 20 servers. In both figures, it can be seen that B_0 is smaller than B_1 over the entire range of A/n. This confirms that a service differentiation is achieved between class 0 and class 1 over the entire range of A/n. However, as already indicated from Figure 4.10 and Figure 4.11, the grade of differentiation is smaller with greater θ_1.

In Figure 4.16, the normalized reward R_{norm} is depicted against the number of restricted servers for class 1 in a scenario with 4 servers which is also depicted in Figure 4.7. Hereby, $R_0 + R_1 = 1$ is always satisfied. The bound that is reached if only class 0 requests are admitted is obtained in this scenario from

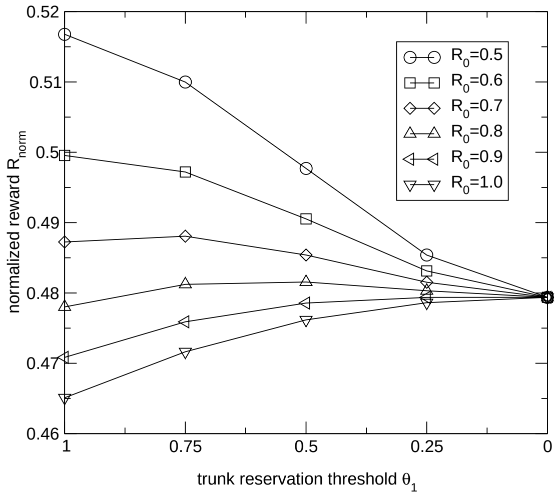


Figure 4.16: Normalized reward against trunk reservation threshold θ_1

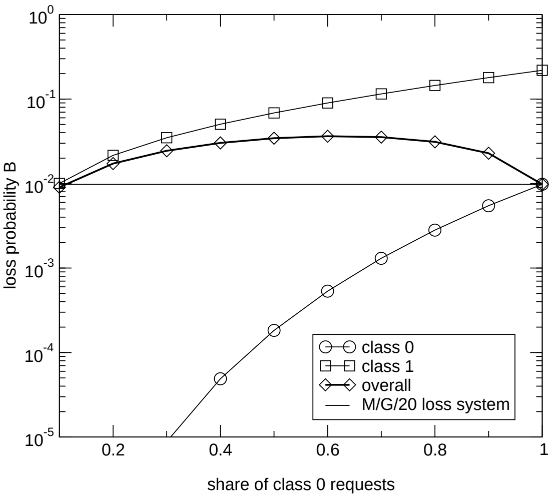


Figure 4.17: Loss probability against share of class 0 requests

$$R_{\text{norm},4} = \frac{\sum_{i=1}^{4} p_{i,0} \cdot i \cdot R_1}{n \cdot R_1} = \frac{Y_1 \cdot R_1}{n \cdot R_1} = \frac{Y_1}{n} = \frac{A_1}{n} \cdot (1 - B) \tag{4.16}$$

For small R_0, the overall system performance is optimized by not restricting any class 1 requests whereas for great values of R_0, total restriction of class 1 request yields best results. In between (see, e. g., $R_0 = 0.7$ and $R_0 = 0.8$), R_{norm} has a flat maximum indicating the number of servers (trunks) which should be restricted for class 1.

Finally, in Figure 4.17, the loss probability is depicted against the share of class 0 requests for the scenario of 20 wavelengths with $\theta_1 = 0.7$. Both, B_0 and B_1 increase with increasing share of class 0 requests. The increase of B_0 is caused by the increasing offered load of class 0 while the increase of B_1 can be explained by an increasing number of class 0 requests which occupy wavelengths which are also available for class 1.

B_{all} equals the loss probability of the M/G/20 system which does not differentiate between classes if either all requests are originated from class 0 or from class 1, respectively. In between, the graph of B_{all} has a flat maximum, i. e., B_{all} is increased if trunk reservation admission control is applied. This is obvious, as trunk reservation rejects requests although not all system resources are occupied. As a consequence, in order to minimize B_{all}, it is advantageous to have most requests of just one class.

4.2.5 Summary

Summarizing, trunk reservation provides service differentiation between a number of classes over the entire range of load. The grade of service differentiation is determined by a fixed parameter per service class which compromises between a low overall loss probability and small loss probability of high priority classes. Thus, the price that has to be paid for service differentiation in loss systems with trunk reservation admission control is an increased overall loss probability.

As trunk reservation thresholds as well as the number of servers which are currently allocated is sufficient to determine whether a request is admitted or not, such an admission control mechanism is very simple, robust and independent of system states. However, it should be emphasized here, that although higher priority traffic is protected from lower priority traffic, no total isolation is achieved.

Chapter 5

Modelling and Performance Evaluation of OBS-QoS Mechanisms

In this chapter, only those reservation mechanisms reported in literature are evaluated whose granularity in the core is a whole burst, i. e., segmentation-based mechanisms introduced in Section 3.4.2 are not considered. In Section 5.1, reservation mechanisms which only support one service class are evaluated analytically and compared with each other. As already indicated in Chapter 4, the theory of loss systems can be taken as basis for all evaluations. This is due to the fact that buffers are not mandatory in OBS nodes and wavelengths can be modelled as servers. Thus, the basic conclusions drawn for loss systems in Chapter 4 are also valid for the following performance evaluations.

Section 5.2 concentrates on the offset-based OBS-QoS mechanism as this is the first and most important mechanism reported in literature. Here, an approximative analysis of the offset-based OBS-QoS mechanism JET is presented which was published first in [43] and an extension in [42]. In Section 5.3, performance evaluations are carried out by the approximative analysis presented in Section 5.2 as well as by simulations. The focus of this section is on shortcomings of the offset-based OBS QoS mechanisms which motivates the introduction of a new OBS-QoS mechanism which overcomes these shortcomings.

All following evaluations have in common that collisions of burst header packets, BHPs, are not considered and thus, an ideal signalling is assumed throughout the rest of this thesis. Instead, only collisions on the data path are taken into account. Furthermore, all evaluations are carried out in a one-node scenario. A discussion of additional effects arising in a network scenario are presented in Appendix A.

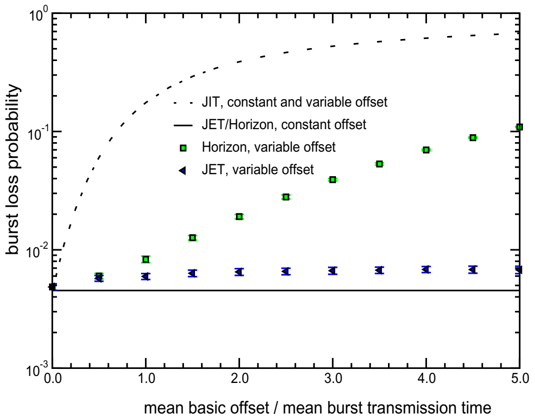


Figure 5.1: Impact of the offset on burst loss probability

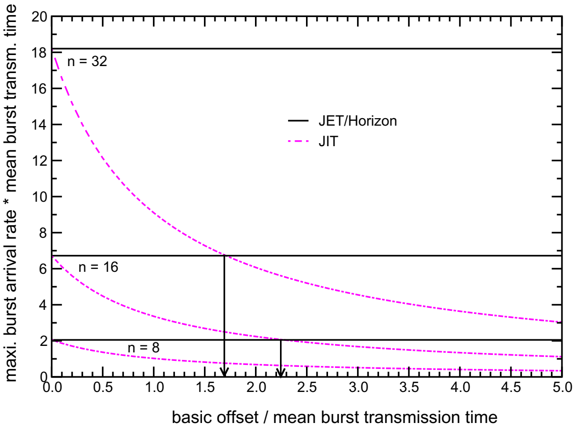


Figure 5.2: Maximum burst arrival rate for a given loss probability of 10^{-3}

5.1 One-Class OBS Reservation Mechanisms

The performance of reservation mechanisms introduced in Section 3.3 can be expressed in terms of burst loss probability. If evaluation is restricted to a single node case with fixed offsets δ for all bursts the loss probability may be obtained analytically, see [43]. In the case of JET, this also means that only a single service class is considered.

Under the assumption that BHPs (and in consequence data bursts) arrive according to negative-exponentially distributed interarrival times with rate λ, Erlang's B formula (4.3) can be taken for calculating the loss probability of an M/G/n loss system. Consequently, all properties of the M/G/n loss system which are discussed in Section 4.1 also apply for such a system.

In Erlang's loss formula n represents the number of servers in a loss system which in this context corresponds to the number of wavelengths on a link. Hence, it is assumed that full wavelength conversion is possible, i. e., a burst can change to any wavelength in case the wavelength it arrives on is currently occupied on the outgoing fiber. The offered traffic A relevant for loss computation depends on the reservation mechanism. For Horizon and JET, the offered traffic is simply the product of arrival rate λ and mean transmission time h of a data burst. So the loss probability of a burst is given for Horizon and JET by

$$P_{\text{Loss, Horizon}} = P_{\text{Loss, JET}} = B(\lambda \cdot h, n) . \tag{5.1}$$

Note, that Horizon and JET have the same performance under above given assumptions as the scenario where a burst is reserved in between two already reserved bursts does not occur in the single node case with constant δ.

If JIT is applied as reservation mechanism the system behaves like a loss system with increased offered traffic, resulting in the loss probability:

$$P_{\text{Loss, JIT}} = B(\lambda \cdot (h + \delta), n) \tag{5.2}$$

The reason for this behavior is that each request blocks a channel for an interval whose length is the sum of basic offset δ and burst transmission time h. The increased load leads to a higher loss probability of JIT compared to Horizon and JET, especially for large δ as demonstrated in Figure 5.1. Therein, as well as in several following graphs, the burst loss probabilities are depicted against the mean offset normalized by h, i. e. δ/h, in order to ease interpretation.

A derived measure, especially interesting for dimensioning, is the maximum burst arrival rate λ_{max} which can be allowed in order to not exceed a certain loss probability on a link with a given number of wavelengths. From (5.2) it can be concluded that in case of JIT λ_{max} is reduced by a factor of

$$\frac{\lambda_{max,\,JIT}}{\lambda_{max,\,JET}} = \frac{1}{1 + \delta/h} \tag{5.3}$$

as compared to Horizon and JET. Figure 5.2 indicates that JIT drastically remains behind JET and Horizon even for relatively small δ. One can see from this figure that a JET/Horizon system with 16 wavelength channels is even better than a 32 wavelength channel system using JIT if $\delta > 1.7h$.

In a network scenario (see also Appendix A), the offset values occurring in a node will not be constant. Therefore, the influence of randomly varying δ is also investigated in Figure 5.1 by simulations [161] as the analysis does not cover varying offsets. For JIT this has no effect, i. e., the loss probability can still be determined using (5.2). In the case of JET and Horizon, however, simulations show that this variation leads to higher losses (variable offset results in Figure 5.1 are obtained for negative-exponentially distributed offsets and burst lengths). While this effect is minor for JET, the loss probability significantly increases for a larger mean offset when Horizon is applied. The conclusion is that the higher complexity of JET as compared to Horizon results in better performance for varying offsets.

5.2 Approximative Analysis of Offset-Based QoS Mechanisms

In this section, an analysis of the burst loss probabilities of a JET-OBS node is presented, which distinguishes multiple classes of bursts and was originally published in [42]. The mean burst length is the same for all classes whereas the offsets can be arbitrary. The loss probability is calculated for a WDM output link assuming full wavelength conversion capability. In Section 5.2.1 the analysis is presented for two classes and in Section 5.2.2 it is extended to multiple classes.

Unlike the single class case where all bursts have the same fixed basic offset δ_{basic} to compensate switching and processing times – as mentioned in Section 3.4.1 – offset-based differentiation is applied to introduce additional offsets for all but the least priority class, called *QoS offset* δ_{QoS}, which provide service class differentiation. For the following analysis, it is assumed that class i has priority over class j if $i < j$ for positive i, j, i. e., the highest priority class has index 0.

If the basic offset and all QoS offsets are constant, the degree of isolation between two arbitrary classes solely depends on their effective offset difference, i. e., the constant basic offset has no impact on isolation. This stems from the fact that a constant basic offset δ_{basic} for all classes can be interpreted as a constant shift in time of the reservation process and thus neither arrival nor reservation events are reordered in time. This result has also been proven by simulation for various arrival and service time distributions and offsets. Hence, $\delta_{\text{basic}} = 0$ is assumed without loss of generality and the effective offset difference $\Delta_{i,j}$ between class i and j can be introduced as

$$\Delta_{i,j} = \delta_i - \delta_j > 0 \qquad \text{for } i < j \tag{5.4}$$

5.2.1 Single Node with Two Classes

5.2.1.1 Basic Formulæ

As introduced in Section 5.1, the loss probability of a one-class system can be obtained by Erlang's loss formula (4.3). In [151], it has been stated that the overall burst loss probablity $P_{\text{Loss, all}}$ is kept constant for equal mean burst lengths regardless of the number of classes which is called in [151] a conservation law. Thus, $P_{\text{Loss, all}}$ on the considered output link in a two-class OBS node with total offered traffic $A_0 + A_1$ can be obtained independently of service differentiation as

$$P_{\text{Loss, all}} = B(A_0 + A_1, n) . \tag{5.5}$$

In order to calculate the burst loss probability of the high priority class $P_{\text{Loss, 0}}$, not only the offered traffic A_0 of the high priority class has to be considered but also a fraction of the carried traffic of the low priority class. This low priority traffic $Y_1(\Delta_{0,1})$ represents bursts which started transmission prior to the arrival of the high priority BHP and are still being served when the high priority burst starts, i. e., $\Delta_{0,1}$ after the high priority QoS offset began. This additional traffic stems from the fact that high priority traffic is not totally isolated from low priority traffic. Thus, $P_{\text{Loss, 0}}$ is approximated by

$$P_{\text{Loss, 0}} = B(A_0 + Y_1(\Delta_{0,1}), n) . \tag{5.6}$$

The burst loss probability of the low priority class $P_{\text{Loss, 1}}$ can be obtained solving

$$(\lambda_0 + \lambda_1) \cdot P_{\text{Loss,all}} = \lambda_0 \cdot P_{\text{Loss,0}} + \lambda_1 \cdot P_{\text{Loss,1}} \tag{5.7}$$

with arrival rates λ_0 and λ_1 for this output link, respectively. This averaging weights burst loss probabilities with respect to their occurrence.

For the carried traffic $Y_1(\Delta_{0,1})$ it follows

$$Y_1(\Delta_{0,1}) = A_1 \cdot (1 - P_{\text{Loss,1}}) \cdot (1 - F_1^f(\Delta_{0,1})) \tag{5.8}$$

where $A_1 \cdot (1 - P_{\text{Loss,1}})$ is the carried traffic of the low priority class at the time when the high priority control packet arrives. $1 - F_1^f(\Delta_{0,1})$ is the complementary distribution function of the forward recurrence time of the burst transmission time at time $\Delta_{0,1}$. It describes the probability that a low priority burst which has already started transmission prior to some random observation time τ has not finished transmission within the period $[\tau, \tau + \Delta_{0,1}]$, see, e. g., [91]. In the considered case, this observation time corresponds to the arrival time of a high priority BHP. Finally, (5.8) is an approximation because in reality, longer bursts are discarded with a higher probability, see also Section 5.3 for simulation results.

5.2.1.2 Iterative Solution

According to (5.6), (5.7) and (5.8), there is a mutual dependency between $P_{\text{Loss},0}$ and $P_{\text{Loss},1}$. Therefore, [42] suggests an iterative solution for above formulæ. The iteration is initialized with estimates for loss probabilities of high and low priority classes, $P^{(0)}_{\text{Loss},0}$ and $P^{(0)}_{\text{Loss},1}$. These zero order estimates are given in (5.9) and can be derived from (5.5) - (5.7) by decoupling the high priority class from the low priority class which is equivalent to neglecting $Y_1(\Delta_{0,1})$.

$$\begin{aligned} P^{(0)}_{\text{Loss},0} &= B(A_0, n) \\ P^{(0)}_{\text{Loss},1} &= 1/\lambda_1 \cdot (\lambda_{\text{all}} \cdot P_{\text{Loss, all}} - \lambda_0 \cdot P^{(0)}_{\text{Loss},0}) \end{aligned} \tag{5.9}$$

Similar formulæ are also published by Qiao and Yoo [152] and yield lower boundaries for the analysis if the QoS offset is very large (Figure 5.6 and Figure 5.7, see below).

The distribution function of forward recurrence time (see, e. g. [91]) of burst transmission time is given by

$$F_1^f(t) = 1/h_1 \cdot \int_{u=0}^{t} (1 - F_1(u))du \tag{5.10}$$

where h_1 and $F_1(u)$ represent the mean and the distribution function of the burst transmission time, respectively. Finally, the amount of carried low priority traffic is determined by (5.8) using (5.9) and (5.10)

$$Y_1^{(0)}(\Delta_{0,1}) = A_1 \cdot (1 - P_{\text{Loss},1}^{(0)}) \cdot (1 - F_1^f(\Delta_{0,1})) \tag{5.11}$$

and can be inserted in (5.6) yielding a first order result for the loss probability of the high priority class $P_{\text{Loss},0}^{(1)}$. By application of (5.7) and the just derived result for $P_{\text{Loss},0}^{(1)}$ a first order result for the low priority class $P_{\text{Loss},1}^{(1)}$ is obtained. Iteration until some precision criterion is satisfied leads to $P_{\text{Loss},0}$ and $P_{\text{Loss},1}$.

5.2.2 Single Node with Arbitrary Number of Classes

5.2.2.1 Basic Formulæ

The burst loss probabilities for k service classes with different QoS offsets is obtained by heuristically generalizing basic formulæ (5.5) - (5.8) to an arbitrary number k of classes [42]. This is performed by considering all interferences from a class m of lower priority on a class i of higher priority ($0 \le i < m \le k-1$). $P_{\text{Loss, all}}$ again follows Erlang's loss formula as given in (4.3). $P_{\text{Loss},0}$ is calculated by taking into account its own offered traffic A_0 and the interfering carried traffic components $Y_m(\Delta_{0,m})$ originating from all lower priority classes

$$P_{\text{Loss},0} = B\left(A_0 + \sum_{m=1}^{k-1} Y_m(\Delta_{0,m}), n\right). \tag{5.12}$$

In the multi-class case, an equation corresponding to (5.7) can be formulated for every set of classes $S_j = \{0, \ldots, j\}$ with $0 < j \le k-1$

$$\left(\sum_{i=0}^{j} \lambda_i\right) \cdot P_{\text{Loss,S}_j} = \sum_{i=0}^{j} \lambda_i \cdot P_{\text{Loss,i}} \tag{5.13}$$

where $P_{\text{Loss,S}_j}$ is the total loss probability of all classes in S_j. Each class i in S_j experiences additional interfering traffic $Y_m(\Delta_{i,m})$ from each class m not belonging to S_j

$$Y_m(\Delta_{i,m}) = A_m \cdot (1 - P_{\text{Loss},m}) \cdot (1 - F_m^f(\Delta_{i,m})). \tag{5.14}$$

These interference components are weighted by the arrival rate of class i within S_j – representing relative occurrence of class i bursts in S_j – and summed up over all i and m for given j

$$P_{\text{Loss,S}_j} = B\left(\sum_{i=0}^{j} A_i + \sum_{m=j+1}^{k-1} \sum_{i=0}^{j} \frac{\lambda_i}{\sum_{l=0}^{j} \lambda_l} \cdot Y_m(\Delta_{i,m}), n\right). \tag{5.15}$$

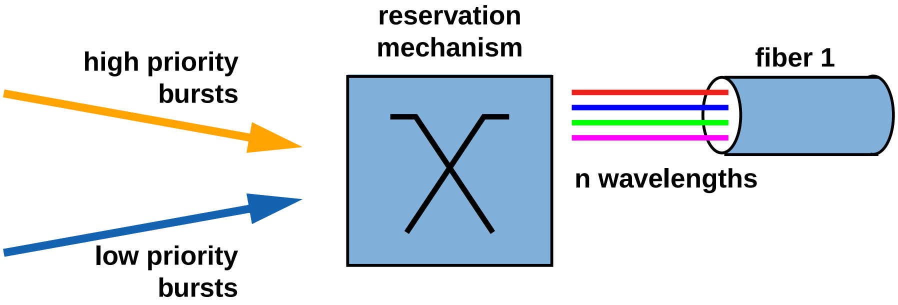


Figure 5.3: Evaluation scenario of a two-class OBS node

Consequently, (5.12) and the set of $k-1$ equations in (5.13) completely describe approximations of burst loss probabilities for all k classes.

5.2.2.2 Iterative Solution

Again, [42] suggests the iterative solution of (5.12) - (5.15). Starting with (5.12) for the highest priority class, (5.13) is repeatedly solved for $P_{\text{Loss},\, j}$ with increasing class indices j. Initial values for $P^{(0)}_{\text{Loss},\, 0}$ from (5.12) are calculated and for all other $P^{(0)}_{\text{Loss},\, j}$ from set of equations (5.13) assuming no interference, i. e.,

$$Y_m(\Delta_{i,\, m}) = 0 \qquad \text{for all valid combinations of m and i} \; . \tag{5.16}$$

These zero order estimates have been described in [152]. They yield lower boundaries in case of perfect isolation with $\Delta_{i,\, i+1} \to \infty$, i. e., no interference of classes. By evaluating (5.14) for zero order estimates and inserting results in (5.12) and (5.13) first order results for all $P_{\text{Loss},\, j}$ can be calculated. Iteration until some precision criterion is satisfied leads to all burst loss probabilities.

5.3 Performance Evaluation of Offset-Based QoS Mechanisms

In this section, a two-class OBS node is evaluated which provides service differentiation by an additional QoS offset for a higher priority class indexed with 0 which is also depicted in Figure 5.3. The number of wavelengths is assumed to be 8. The example of two different classes is sufficient to work out the main characteristics of an offset-based QoS mechanism without introducing unnecessary complexity. In, e. g. [151] and in Appendix A, evaluations of a system with more classes are shown. As analytically obtained results for a higher number of wavelengths indicate, the principle behavior of an offset-based OBS node does not change for an increased number of wavelengths. For the following, a scenario is considered where the normalized offered traffic is 0.6 with a share of 30% of high priority traffic. Comparable to the

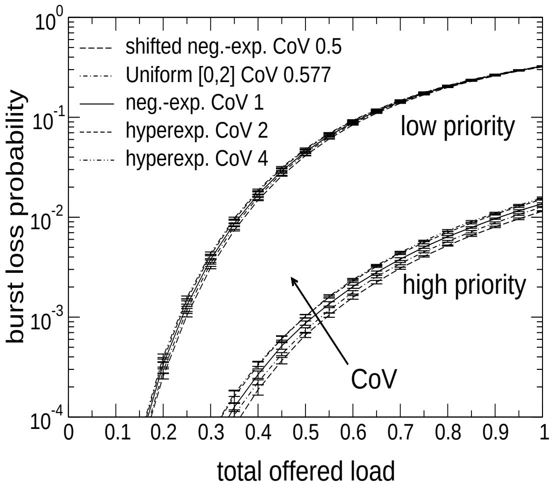


Figure 5.4: Impact of interarrival time distributions on $P_{\text{Loss, all}}$

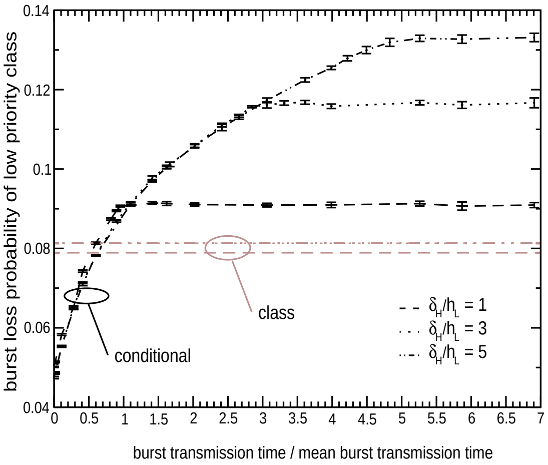


Figure 5.5: Impact of actual low priority burst length on $P_{\text{Loss, 1}}$

above introduced approximative analysis, the following results are published first in [43], [42] and [55].

In Section 5.3.1, the impact of the interarrival time distribution is evaluated in order to motivate why the M/G/n loss system is taken as basis for the approximation. Section 5.3.2 compares the burst loss probability obtained by analysis and simulation and discusses the dependencies of the low priority burst length distribution on the high priority burst loss probability. In Section 5.3.3, the impact of the actual burst length on low priority burst losses is shown. Section 5.3.4 discusses the impact of different burst lengths on the burst loss probability. Finally, Section 5.3.6 summarizes the results of this chapter and draws conclusions with respect to the applicability of offset-based OBS-QoS.

In all evaluations in this section, bursts are not assembled from, e. g., IP packets. Instead, bursts are generated directly according to a certain interarrival time and a burst length distribution. This traffic model is chosen in order to work out the impact of traffic characteristics on the burst loss probability.

5.3.1 Impact of Interarrival Time Distribution

As the assumption in the approximative analysis that the burst interarrival time has Markovian property seems to be very restrictive, simulations varying the burst interarrival time distribution of both classes are carried out. In Figure 5.4, burst loss probabilities of a high and a low priority class for different uncorrelated interarrival time distributions[1] and negative-exponentially distributed burst lengths ($h_0 = h_1$) are depicted against the overall offered traffic. The various interarrival time distributions are further characterized by the coefficient of variance,

[1] The hyperexponential distribution satisfies the symmetry condition $p \cdot h_1 = (1 - p) \cdot h_2$ where p is the branch probability and h_1 and h_2 are the mean values of the respective phases.

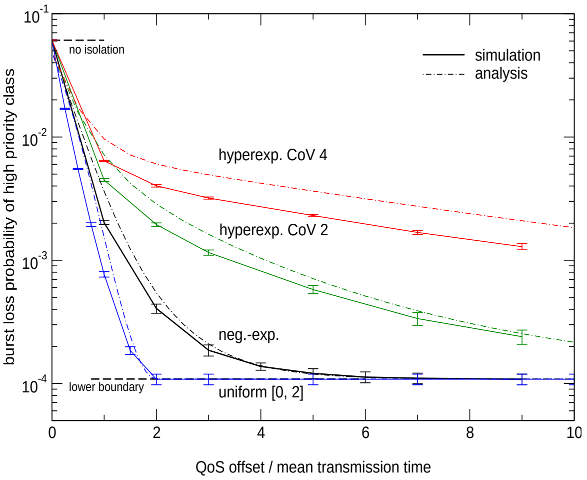


Figure 5.6: Impact of low priority burst length distribution on $P_{\text{Loss, 0}}$; $n_{\text{WL}} = 8$

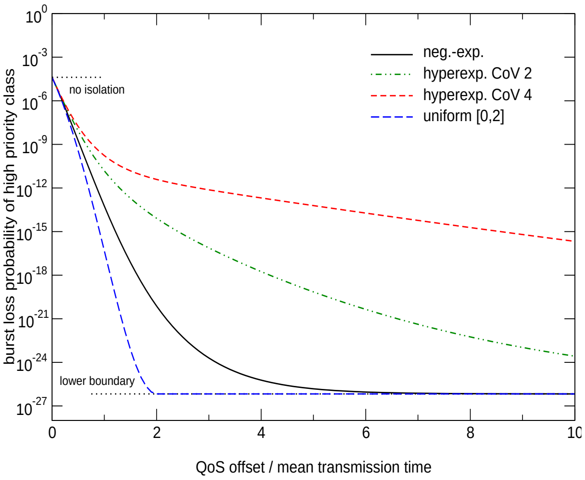


Figure 5.7: Impact of low priority burst length distribution on $P_{\text{Loss, 0}}$; $n_{\text{WL}} = 64$

CoV. It can be seen that changes in the arrival process have only small impact on the burst loss probabilities of both classes. Thus, the model of a Poisson arrival process assumed for the approximative analysis yields reasonable results even for different interarrival time distributions.

5.3.2 Impact of Low Priority Burst Length Distribution

In this section, mean burst transmission times of high and low priority bursts h_0 and h_1, respectively, are assumed to be identical, i. e., $h_0 = h_1$. Figure 5.6 shows $P_{\text{Loss, 0}}$ against the QoS offset δ_0 normalized by h_1 for different low priority burst length distributions. An upper bound for the case of no isolation as well as a lower bound for perfect isolation (see Section 5.2.1.2) are included. The upper bound corresponds to a system without isolation. According to Erlang's loss formula (4.3) results can be obtained considering $A = A_0 + A_1$. The lower bound reflects the case of total isolation and thus is obtained by Erlang's loss formula and an offered traffic A_0. Hence, this bound reflects a minimal burst loss probability of the high priority class for all possible OBS-QoS mechanisms.

From Figure 5.6, it can be seen that the presented approximative analysis matches the simulated curves quite well for all distributions. Furthermore, the strong impact of the forward recurrence time of low priority bursts as indicated by (5.6) and (5.8) on the high priority burst loss probability, $P_{\text{Loss, 0}}$ is visible. If the coefficient of variation, CoV, of the low priority burst length distribution is increased, the probability also increases that very long low priority bursts occupy wavelengths for a long time and thus reduce the service differentiation. In case of, e. g., a hyperexponential distribution with CoV $= 4$, the lower bound of $P_{\text{Loss, 0}}$ is approached very slowly. The herefore required offset of the high priority class is too large to be realized and hence, service differentiation according to the theoretical lower bound cannot be assumed.

This result is especially critical as recent publications, e. g. in [65], proof that burst assembly does not reduce the self-similar traffic characteristic. As a consequence, if no additional traffic engineering mechanism is applied, it cannot be expected that the burst length distribution has a small CoV.

Figure 5.7 shows that the principle shape of curves depicted in Figure 5.6 remains unchanged for a higher number of wavelengths. Only the order of magnitude of losses changes drastically, e. g., for 64 wavelengths the lower bound reduces to about 10^{-26}. This characteristic system behavior also emphasizes why it is acceptable to carry out simulations in a scenario with a reduced number of wavelengths. Another intention of this graph is to show that very low burst loss probabilities and thus reasonable performance characteristics for transport networks can be achieved. Hereby, as already indicated during the discussion of loss systems in Section 4.2, the significantly reduced burst losses are obtained through multiplexing gain for a large number of servers/wavelength and a normalized offered traffic which is far below 1.

5.3.3 Impact of Low Priority Burst Length

In Figure 5.5, an important impact of offset-based service differentiation on loss probabilities is depicted. In this graph, $P_{\text{Loss, 1}}$ is depicted against the burst transmission time normalized by the mean burst transmission time. It demonstrates for a system with different offsets that the loss probability of a low priority burst depends on the actual length of a burst. This behavior is inherent to a system in which low priority bursts tend to occupy wavelengths in between already reserved high priority bursts (see also Section 3.4.1). Thus, the probability to find a gap of appropriate length is higher for bursts which are shorter than δ_0. It can be seen that the conditional low priority burst loss probability as depicted in Figure 5.4 increases until the respective burst transmission time is as long as δ_0 and stays constant from there on. The longer the offset time, i. e., the greater the isolation, the larger is the difference between the burst loss probabilities of a short burst and a very long burst. For the scenario of a QoS offset of five times the mean burst transmission time, $P_{\text{Loss, 1}}$ more than doubles for long bursts. A solution to this problem could be to bound burst lengths within a short interval. However, this has the disadvantage that several short bursts produce much more overhead concerning connection management and signalling, which is especially undesirable for the low priority class.

5.3.4 Impact of Ratio of Mean Burst Lengths

In order to reduce processing overhead and increase efficiency for large volume bulk traffic, longer low priority bursts might be advantageous. However, in order to maintain a certain degree of isolation, larger low priority bursts result in a larger QoS offset and consequently a longer pre-transmission delay for the high priority class. With respect to this trade-off, the per-

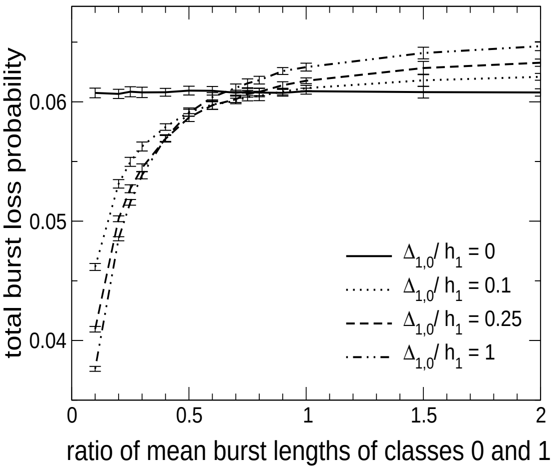


Figure 5.8: Impact of different mean burst lengths on $P_{\text{Loss, all}}$

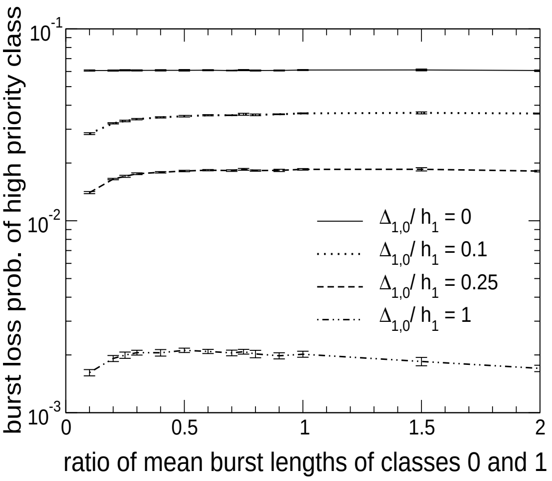


Figure 5.9: Impact of different mean burst lengths on $P_{\text{Loss, 0}}$

formance of an OBS node is evaluated depending on the ratio of the mean burst lengths $h_{0,1} = h_0/h_1$. In order to keep the offered traffic $A_i = \lambda_i \cdot h_i$ unchanged within each class the arrival rates are adapted. Figure 5.8 shows $P_{\text{Loss, all}}$ against $h_{0,1}$. In this graph, curves are drawn for several offsets. As expected, $P_{\text{Loss, all}}$ is unchanged for varying $h_{0,1}$ if no offset distinguishes the classes. But even for very small offsets $P_{\text{Loss, all}}$ changes significantly with $h_{0,1}$. For shorter high priority bursts $P_{\text{Loss, all}}$ decreases while it increases for longer high priority bursts. Thus, a decreased $P_{\text{Loss, all}}$ can be achieved by operating the system with bursts satisfying $h_{0,1} < 0.7$. This scenario contradicts $P_{\text{Loss, all}} = \text{const}$ and therefore is not covered by the analysis presented in Section 5.2.

In order to get a deeper inside into this effect, the burst loss probabilities of both classes are observed separately by simulations. From Figure 5.10 it can be seen that $P_{\text{Loss, 1}}$ significantly increases for decreasing $h_{0,1}$. As already discussed in Section 5.3.3, this effect is caused by the reservation mechanism itself, as low priority bursts in most cases fill gaps left over by high priority bursts. Due to the higher number of arriving high priority bursts per time interval, the link is fragmented and the length of gaps left for low priority bursts is reduced. It can be seen that the burst loss probability increase is larger for lower $h_{0,1}$. If the burst transmission time is longer than the offset duration, a boundary value is reached. This boundary value increases for decreasing $h_{0,1}$. Again, very short bursts are not affected as they fit into small gaps left over by higher priority bursts.

Resuming the above discussion, Figure 5.9 indicates that $P_{\text{Loss, 0}}$ slightly decreases for shorter high priority bursts. Together with the description of $P_{\text{Loss, 0}}$ in (5.6) and (5.8) and the increase of $P_{\text{Loss, 1}}$, the decrease of $P_{\text{Loss, 0}}$ can be explained: High priority traffic experiences reduced low priority interference due to higher low priority losses. Considering the significant changes of the arrival rates over $h_{0,1}$ in (5.7) as well as the behavior of $P_{\text{Loss, 0}}$ and $P_{\text{Loss, 1}}$, the dependence of $P_{\text{Loss, all}}$ on $h_{0,1}$ depicted in Figure 5.8 can now be explained.

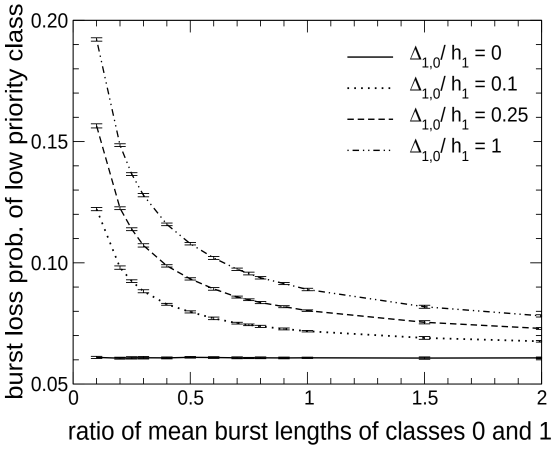


Figure 5.10: Impact of different mean burst lengths on $P_{\text{Loss, 1}}$

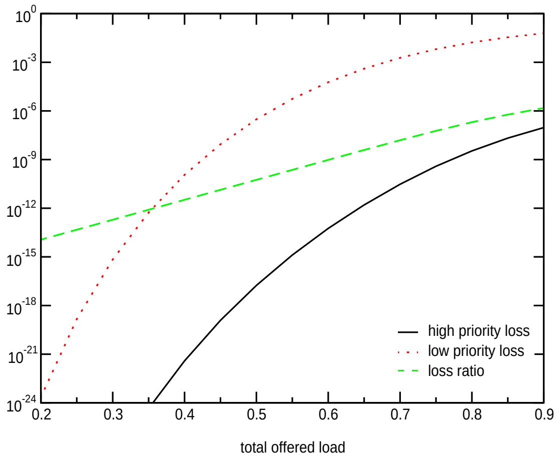


Figure 5.11: Impact of the total offered load on $P_{\text{Loss, 0}}$ and $P_{\text{Loss, 1}}$

Summarizing, on the one hand, it is desirable to have a small $h_{0,\,1}$ because it fits the idea of short high priority, potentially real-time bursts and long bulk traffic low priority bursts, and it results in a reduced $P_{\text{Loss, all}}$. On the other hand, if $h_{0,\,1}$ is small, $P_{\text{Loss, 1}}$ increases significantly for longer low priority bursts. This is undesirable, especially as from the signalling and processing point of view, it is much more efficient to transmit long low priority bursts.

5.3.5 Impact of Occupancy

For completion, $P_{\text{Loss, 0}}$, $P_{\text{Loss, 1}}$ as well as their ratio $P_{\text{Loss, 0}}/P_{\text{Loss, 1}}$ are depicted in Figure 5.11 against the total offered load which is equally increased for both classes. Here, like in Figure 5.7, a scenario of 64 wavelengths is chosen. The offset δ_1 is set equal to the mean burst transmission time, which yields $P_{\text{Loss, 0}} \approx 10^{-10}$ at a total offered traffic of 0.7. It can be seen that a good grade of isolation is kept over the whole range of offered load. However, the ratio $P_{\text{Loss, 0}}/P_{\text{Loss, 1}}$ increases with increasing occupancy. Furthermore, like already indicated in graphs in Chapter 4, it can be seen that the occupancy has to be controlled in order to keep $P_{\text{Loss, 1}}$ within reasonable boundaries.

5.3.6 Summary and Conclusions

Summarizing the qualitative evaluation in Section 3.4.1 as well as the quantitative evaluations in Section 5.3, it can be stated for an offset-based OBS-QoS mechanism:

- The type of interarrival time distribution does not significantly influence burst loss probabilities.

- The burst loss probability of a higher priority class strongly depends on the burst length distribution of lower priority classes. Hereby, burst length distributions with a greater CoV yield significantly increased losses.
- The burst loss probability of lower priority classes depends on the actual length of a burst. Shorter bursts experience a lower probability to be lost.
- If bursts of different classes do not have the same mean burst length, $P_{\text{Loss, all}} = \text{const}$ is not satisfied any more. Additionally, in a desirable scenario of short high priority bursts and long low priority bursts, the burst loss probability of low priority bursts is significantly increased.
- A good grade of isolation is kept over the whole range of occupancy. However, an additional mechanism to control the offered load is required in order to bound burst losses and thus guarantee a certain grade of service, a requirement which is already generally pointed out for loss systems in Section 4.1.

These results, especially the dependencies of burst loss probabilities on traffic characteristics like, e. g., burst length distribution and mean burst length, are undesirable. A QoS mechanism should yield service differentiation independent on traffic characteristics, especially from lower priority traffic and thus, make the performance of service classes predicable and calculable. Furthermore, feedback from the network to the edges is required in order to be able to provide service guarantees. Such a feedback should be integrated into the reservation mechanism or the OBS-QoS mechanism. Therefore, a new OBS-QoS mechanism called Assured Horizon, see Chapter 6, is designed to overcome these shortcomings.

Chapter 6

The new OBS-QoS Framework Assured Horizon

Based on shortcomings of OBS-QoS mechanisms presented in Chapter 3 and Chapter 5, a new OBS-QoS framework called Assured Horizon is introduced in this chapter. Assured Horizon is a combined framework for a burst assembly mechanism, a burst reservation mechanism as well as the communication between them in optical burst switched networks [40], [41].

In Section 6.1, design goals of Assured Horizon and an overview are presented and the major new contributions of Assured Horizon to the research community are outlined. Corresponding to the three parts of the framework, the remainder of this chapter also consists of three parts: the new bandwidth reservation mechanism is discussed in Section 6.2, the new burst assembly mechanism is discussed in Section 6.3 and the new wavelengths reservation mechanism is introduced in Section 6.4.

6.1 Design Goals and Overview of Assured Horizon

Section 6.1.1 lists the main design goals of Assured Horizon and Section 6.1.2 presents a brief overview of this framework whose parts are discussed in more detail in Section 6.2 - Section 6.4. Section 6.1.3 outlines the major new contributions of Assured Horizon.

6.1.1 Design Goals

In general, the design goal of Assured Horizon is to overcome the three main challenges which are faced when realizing QoS differentiation (see Section 3.4), namely (i) limited time for burst header processing in core nodes, (ii) no buffers in the core (beyond FDLs) to perform scheduling, and (iii) no feedback about the network status provided from the core nodes to the edge nodes by the one-pass reservation. Derived in the context of general loss systems in Chapter 4 as well as from the performance evaluation of an offset-based OBS-QoS mechanism in Chapter 5, more specific requirements for a new OBS-QoS mechanism, which is in the focus of this thesis, are:

- isolation of service classes,
- independence of burst losses from burst characteristics,
- invariance with respect to networking effects,
- provisioning of feedback from the network to the ingress without two-way signalling, and
- control of offered traffic by ’burst admission control’.

More general design goals are

- integration/utilization of GMPLS control, see Section 2.4.2, and
- accomplishment of as much header processing as possible at the ingress.

Thus, a kind of (distributed) scheduler has to be designed which considers and also takes advantage of special requirements of the optical layer. Hereby, one of the greatest differences to the electronic domain is the fact that optical random access memory is not available and thus most scheduling principles which are well-known from the electronic domain [155] cannot be applied here.

Assured Horizon can be classified to the class of active dropping-based OBS-QoS mechanisms, see Section 3.4.3. This class of OBS-QoS mechanisms inherently satisfies most of the requirements and design goals which have just been introduced. In this class of OBS-QoS mechanisms, the burst dropper controls admission to the wavelengths reservation process according to a dropping function which should be dimensioned in that way, that the burst loss probability equals to the burst drop probability, i. e., every burst which passes the dropper can reserve a wavelength. This allows to directly engineer the burst loss probability by the burst dropping function.

Finally and perhaps most important, the overall design goal is a mechanism which is as simple/ incomplex as possible. Hence, it can satisfy the stringent time requirements which arise with increased data rates and thus can be applied in real OBS networks.

6.1.2 Overview of Assured Horizon

Assured Horizon consists of three major building blocks

1. Coarse-grained (or static) reservation of a bandwidth envelope for every FEC

   The basic idea of Assured Horizon is the reservation of a timely coarse-grained or even static amount of bandwidth for every forwarding equivalent class, FEC, between an ingress node and an egress node by a protocol of the GMPLS architecture or statically by management, respectively. This bandwidth envelope allows to control the accepted traffic by giving feedback from the core to the edges whether the reservation envelope can be increased in case it is dynamically adapted to the mean rate of a FEC. Additionally, as consequence of this feedback, a timely coarse-grained or static ‘burst admission control‘ can be provided by this framework which is the basis for any QoS guarantees.

   The experienced burst loss probability is determined by the ratio of reserved bandwidth and mean bandwidth of a FEC and hence can be determined independently for each FEC.

2. Policing and marking at the ingress by the burst assembly mechanism in a distributed way

   In order to consider the challenge of little available time to perform burst header processing in the core, a major part of the policing functionality can be carried out by the burst assembly mechanism at the ingress in a distributed way. This functionality includes observation whether the used bandwidth exceeds the reserved bandwidth envelope. Therefore, each burst assembly mechanism has an assembly queue per FEC and marks bursts in a new field in the BHP as compliant, C, or non-compliant, NC, with respect to its reserved bandwidth envelope. Hereby, an algorithm compromises between the proportion of NC bursts and the waiting time in the assembly buffer.

3. Central enforcement of policed bursts

   The last building block of Assured Horizon is the active dropping-based enforcement of the policing at each core node and thus the performance of a distributed burst admission control. Dependent on the load situation of a core node (whether a node is congested or not), a core node either admits all bursts to the wavelengths reservation process in order to allow for multiplexing gain or drops NC bursts in order to increase the probability of C bursts to successfully reserve a wavelength and hence guarantee a certain burst loss probability of C bursts.

In order to further explain the meaning of the three building blocks, they are compared with an electronic scheduler with queueing which is an analogon in the electronic domain. The reserved bandwidth envelope in Assured Horizon corresponds to a weight at a weighted scheduler, e. g., weighted fair queueing, WFQ [155]. This comparison confirms that the bandwidth is only adapted in a timely coarse-grained or even static manner as weights at a weighted scheduler are also not adapted at the same time level as an IP packet.

The traffic share in this framework which is marked as compliant corresponds to packets which are guaranteed to experience less than a certain queueing delay whereas non-compliant traffic uses the excess bandwidth and thus allows for multiplexing gain. So, generally, the major difference to the electronic analogon is that queueing in the electronic world corresponds to dropping in this approach. However, the queueing probability of an electronic scheduler cannot be mapped directly to the dropping probability in this approach as some contention can be resolved in the frequency domain by wavelength conversion. Furthermore, FDL buffers might be included in an architecture to further reduce the burst loss probability and thus to reduce retransmissions which are possibly initiated by higher layer protocols [53], [56], [122].

6.1.3 Major new Contribution of Assured Horizon

The framework Assured Horizon provides the following major new contributions:

- Introduction of feedback from the core to the edges

  This is the basis to control the offered traffic while still allowing for multiplexing gain. As a consequence, some sort of ‘burst admission control’ can be carried out. Hereby, the burst admission control mechanism is based on the simple but very efficient trunk reservation admission control mechanism introduced in Section 4.2.

- Intelligent ingress nodes

  The intelligence is shifted from the core to the edges and thus from optical core routers to electronic assembly nodes. By doing so, a distributed scheduler can be realized without mandatory use of queues.

- Provisioning of service guarantees

  Burst admission control allows to guarantee a certain burst loss probability. As these guarantees are independent of traffic characteristics of other classes, isolation between classes can be obtained.

- Stateless core network

  In order to have a very simple optical core network, Assured Horizon realizes a stateless core, an idea which was published first in a different context in [128].

6.2 Bandwidth Reservation Mechanism

For every FEC i between an ingress node and an egress node including all intermediate core nodes, a bandwidth reservation envelope r_i has to be set up and maintained. This reserved bandwidth envelope allows for traffic isolation between FECs and is the basis for feedback from core nodes to the ingress node about the network congestion status. Thus, r_i is the major building block for QoS guarantees in the Assured Horizon framework.

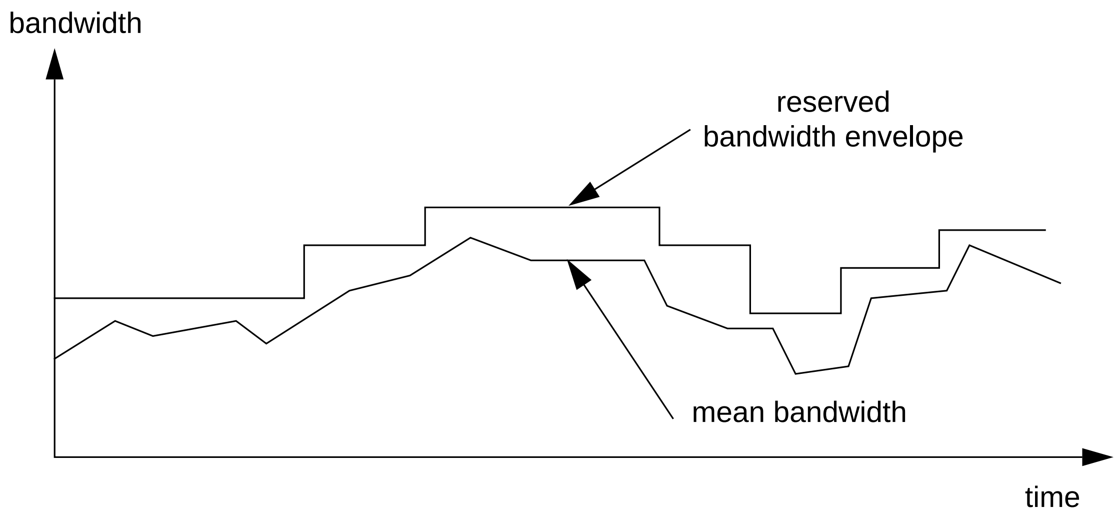


Figure 6.1: Dynamic bandwidth management scheme

Bandwidth reservation for a FEC results in a QoS architecture with a granularity between class-based (e. g., DiffServ, Section 2.2.2) and flow-based (e. g., IntServ, Section 2.2.2). In order to comprise between the advantages of both approaches and to get rid of their disadvantages, reservation in Assured Horizon is ‘class-based per path’, i. e., aggregates of flows with the same ingress and egress nodes. This is especially efficient as bandwidth is only reserved on a per-class basis. Additionally, protection between traffic of the same service class but originated from different ingress nodes /destined to different egress nodes is provided.

The allocation factor $f_i = r_i/m_i$ is the ratio of the reserved bandwidth envelope r_i of a FEC and its mean bandwidth m_i. Starting from $f_i = 1$, the greater f_i, the greater the probability that also a temporarily greater arrival rate is fully covered within r_i and thus can be delivered without any losses to the egress node. Consequently, f_i directly determines the QoS that is experienced in the network and thus has to be carefully determined. This concept is also reported in literature as reservation for effective bandwidth, see, e. g., [87] for a comprehensive overview. A discussion how to determine f_i is presented in Section 7.1 and Section 7.3.

As f_i is crucial for the QoS, r_i may be adapted to m_i in order to keep f_i reasonable constant. However, this adaption should be of a granularity which is much coarser than the one of a mean burst duration. A protocol, e. g., of the GMPLS family is required to perform this dynamic bandwidth management. Such a scheme will be assumed throughout the rest of this thesis. The functionality of such a dynamic bandwidth management scheme is depicted in Figure 6.1 and – in a slightly different context – a more detailed description and analysis can be found in, e. g. [96].

Hereby, an ingress node and a core node play a different role. Whereas the core node only has to remember the accumulated reserved bandwidth of all FECs passing it, an ingress/assembly node is responsible to observe the mean bandwidth of every FEC and – in case of a significant

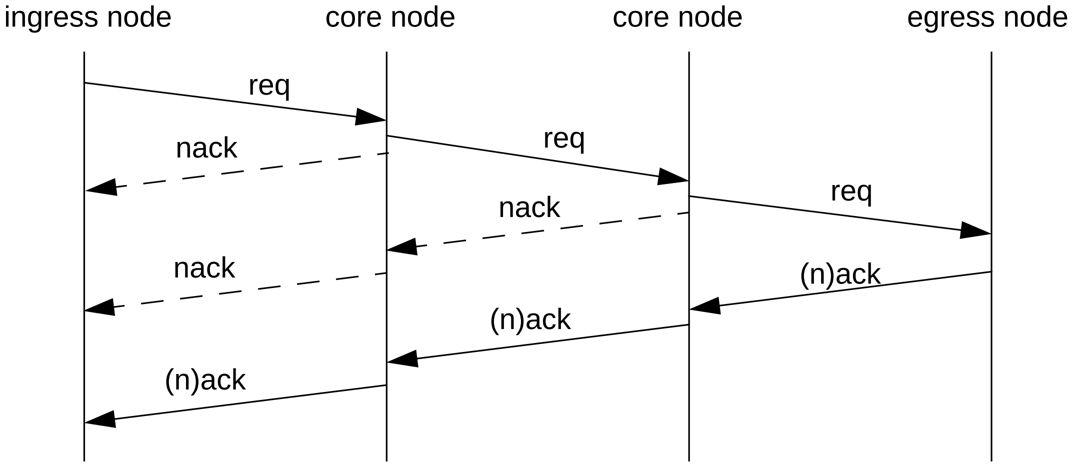


Figure 6.2: Signalling to increase/decrease reserved bandwidth envelope

change of f_i – signal towards the egress node the adapted new bandwidth requirement. In Figure 6.2, the signalling is depicted schematically. This signalling can be carried out by, e. g., a signalling protocol of GMPLS, see Section 2.4.2.1.

The functionality which has to be covered by such a signalling protocol is the request of more/ less bandwidth for a FEC. As a core node only remembers the accumulated reserved bandwidth[1], the delta of bandwidth which is additionally required/can be released is sufficient to be contained in the signalling message. In case of greater required bandwidth, every intermediate node checks whether the request can be granted. If the additional bandwidth can be granted, the node pre-reserves the bandwidth and forwards the request towards the egress node. If the egress node is reached, it signals back a positive acknowledgement, ack, in order to indicate that the pre-reserved bandwidth has to be reserved. In case it cannot be granted, the request is rejected and dependent on the applied signalling protocol, this information is signalled back towards the ingress node as negative acknowledgement, nack, and the pre-reserved bandwidth is released. A more detailed specification of such a signalling protocol is out of the scope of this thesis.

This signalling is the basis for a so-called coarse-grained ‘burst admission control‘, BAC. As every core node between ingress and egress node has to agree to an increase in bandwidth, a reservation request and thus an increase in offered traffic to the wavelengths reservation process can be rejected. Thus, a burst switch can control the amount of (admitted) traffic in order to avoid overload situations, a requirement which is the outcome of general evaluations of loss systems in Chapter 4 as well as the performance evaluation of the OBS-QoS mechanism JET in Chapter 5. This is also the basis for QoS guarantees in the Assured Horizon framework.

[1] Remembering only the accumulated reserved bandwidth allows to keep the core stateless with respect to FECs and also simplifies/accelerates the reservation. However, a solution where core nodes hold state information for different FECs would also be possible.

However, for simplicity, it is also possible to apply static reservation envelopes which partition the link according to any (non-technical) policy. Such a static reservation envelope can be setup and changed by a network management system.

6.3 Burst Assembly Mechanism

Besides aggregating flows of arriving IP packets and assembling them to bursts, the burst assembly mechanism in the Assured Horizon framework has a variety of different tasks in the context of QoS support. Among these tasks which are integrated in the assembly mechanism are the observation and policing of the bandwidth reservation envelope per FEC. Hereby, classification and aggregation of arriving IP packets to a FEC, the decision how many FECs exist between an ingress node and an egress node as well as management of a FEC (e. g., setup and constrained-based routing) are assumed to be carried out by GMPLS, see Section 2.4.2.

The burst assembly mechanism has an assembly queue and a timer for every FEC, see also Figure 6.4. It observes whether the offered traffic exceeds r_i by marking bursts as compliant, C, or non-compliant, NC, dependent on a burst conforming to r_i or not. As already discussed in the previous section, the greater r_i, the greater the share of C bursts and thus the better experienced QoS. The length of a burst which is assembled follows

$$l_i = r_i \cdot \tau_{\mathrm{iat}} \tag{6.1}$$

with the interarrival time τ_{iat} since the last timeout (which may be greater than the timeout interval τ_i). Hence, greater r_i as well as greater τ_{iat} yield longer bursts. As in a real implementation, it is not meaningful to split an IP packet in two bursts, l_i may be exceeded by the maximum length of an IP packet.

In order to consider the greater amount of Bytes that has been sent or to use some bandwidth of previous intervals which have not been used, l_i may be adapted according to the well-known exponential weighting following

$$l_i^{(j)} = \left(1 - e^{-\frac{\tau_{\mathrm{iat}}}{\tau_{\mathrm{const}}}}\right) \cdot (r_i \cdot \tau_{\mathrm{iat}}) + e^{-\frac{\tau_{\mathrm{iat}}}{\tau_{\mathrm{const}}}} \cdot l_i^{(j-1)} \tag{6.2}$$

with $l_i^{(j)}$ being the estimate of l_i in the j^{th} interval and a normalization constant τ_{const}. Dependent on τ_{const}, this exponential weighting allows also to use same amount of bandwidth from previous timeouts that has not been used so far. However, exponential weighting of l_i is just an option and the burst assembly mechanism also works fine if l_i is always determined according to (6.1).

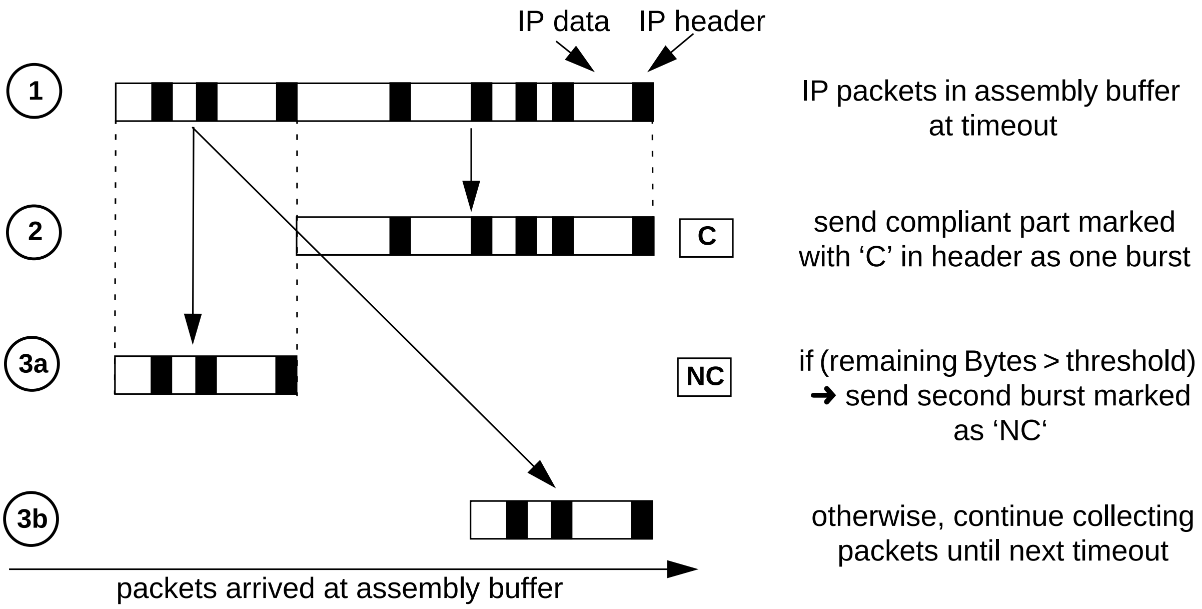

IP data
IP header
1
IP packets in assembly buffer at timeout
2
C
send compliant part marked with ‘C’ in header as one burst
3a
NC
if (remaining Bytes > threshold) ➜ send second burst marked as ‘NC‘
3b
otherwise, continue collecting packets until next timeout
packets arrived at assembly buffer

Figure 6.3: Burst assembly mechanism

In order to indicate whether a burst is C or NC, a new field in the BHP called burst drop priority, BDP, is introduced. The name of this field follows an idea in ATM where a bit in the ATM header called ‘cell loss priority’, CLP, indicates a cell exceeding its reservation, see, e. g. [89]. In the core, all evaluations to support service differentiation are solely based on that new BDP field and thus, no calculation has to be carried out upon arrival of a BHP. In the following, a purely time-based algorithm is presented and also illustrated in Figure 6.3 which applies the basic marking scheme:

1. Upon arrival of an IP packet, the packet is classified by GMPLS to a FEC and forwarded to the respective queue and the respective timer is set to τ_i (if it is not already set).

2. When a timer of a FEC expires, the assembly unit assembles a burst of maximum length which is still compliant to r_i according to (6.1) or (6.2). In case the resulting burst length is shorter than a defined minimum burst length, the burst may be padded. This burst is released into the network.

3a. If the accumulated length of the IP packets remaining in the assembly queue exceeds a threshold σ_i, they are all sent in a second burst marked as NC. In order to further control the offered traffic of a FEC, an option of this algorithm may bound the number of Bytes per FEC sent as NC.

3b. Otherwise, the non-compliant IP packets remain in the assembly buffer, the timer is set to τ_i and arriving IP packets are added until the next expiration of the timer.

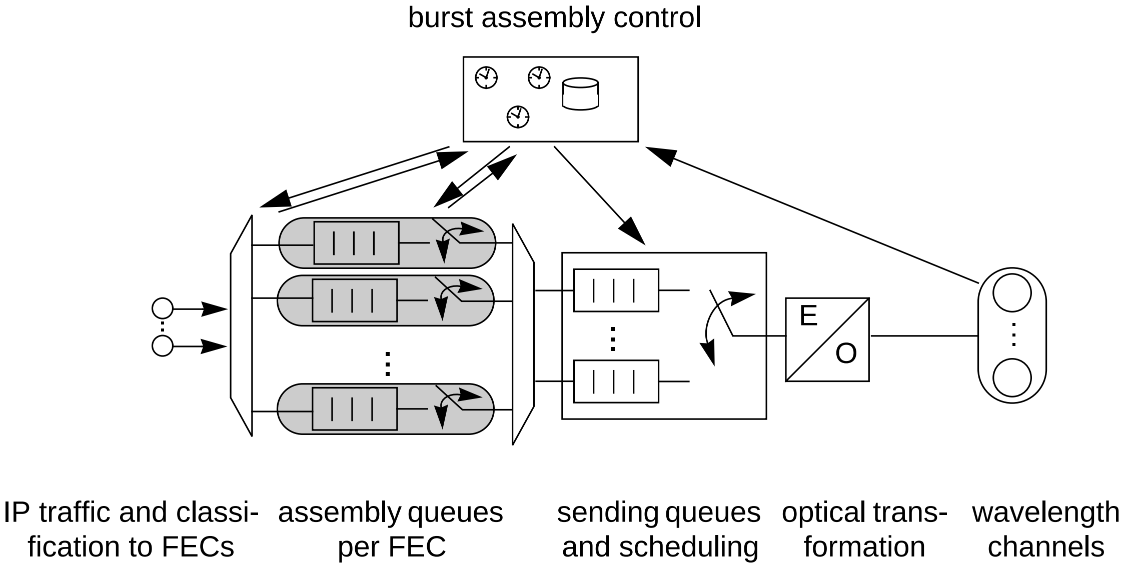


Figure 6.4: Model of burst assembly and local reservation mechanism

The threshold σ_i compromises between the proportion of NC bursts and the waiting time in the assembly buffer. If σ_i is very small, a second burst marked as NC is generated every time not all packets are sent out in a C burst. Thus, as a limit, the amount of burst may be doubled. However, the waiting time in an assembly buffer is bounded to τ_i for this case. On the contrary, a great σ_i results in very little additional bursts but possibly long (unbounded) waiting time, see Section 7.1.3.

6.4 Burst Reservation Mechanism and Burst Dropping

The design of a bandwidth reservation mechanism has to distinguish between a reservation mechanism at an access node where bursts are assembled and are waiting in an electronic buffer and a reservation mechanism in the core where no buffers are available.

6.4.1 Reservation at the Edge

A local reservation mechanism controls the access to wavelengths at an edge node where bursts are assembled. The difference to a core node is hereby, that these bursts are still in the electronic domain and hence can be buffered in a random access memory before being transformed to the optical domain. Consequently, an electronic scheduling algorithm with queueing can be applied to schedule bursts to outgoing wavelength channels. A model of this is depicted in Figure 6.4. Here, burst assembly queues sent bursts to ‘sending queues‘ which are still electronic and can be scheduled according to any scheduling strategy.

The most simple solution is to apply just one sending queue and hence schedule all burst according to first come first serve, FCFS. However, especially for a large number of FECs and different service classes, this might result in undesired waiting time for C bursts. Therefore, it is more appropriate to apply two sending queues, one for C bursts and one for NC bursts. In order to minimize the waiting time for C bursts, a scheduling algorithm always prioritizing C bursts fulfills the requirements and is very simple to realize. Additionally, it already follows the design principle in the core where only C and NC bursts are distinguished no matter to which service class they belong to. In case of reasonable dimensioning of the reservation envelopes, their overall sum is smaller than the link capacity and thus, starvation of NC bursts cannot occur. If, however, additional isolation between bursts of the same service class but different FEC is desired, more sending queues can be applied. These sending queues of the same priority can be scheduled according to weighted fair queueing, WFQ, or an approximation of it, e. g., self clocked fair queueing, SCFQ, see [155] for a comprehensive overview on scheduling algorithms.

Assuming reasonable dimensioning of the reserved bandwidth, it can be expected that no C burst is lost at the edge of the network due to an overflow of a sending buffer. On the contrary, depending on the traffic characteristics, NC bursts may be lost in case of great overload.

6.4.2 Reservation in the Core Supported by Active Dropping

In contrast to the just described local reservation mechanism at ingress nodes, the global reservation mechanism at core nodes cannot be based on queueing-based scheduling as no optical random access memory is available. However, in order to be able to control burst losses, a reservation mechanism which does not distinguish between bursts, see Chapter 3.3 is supported by an active dropper which realizes an admission control to the wavelengths reservation process, see Chapter 3.4.3. For simplicity, Horizon [134] is chosen as wavelengths reservation mechanism. However, any other mechanism which carries out complete sharing of wavelengths can be applied in this framework.

The major task of the burst dropper is to support the reservation mechanism by dropping NC bursts before contention occurs and thus before reserved C bursts have to be discarded because they cannot find an available wavelength[1]. Hence, the burst dropper carries out burst admission control functionality which is distributed throughout the whole core network. The difficulty hereby is the grade of dropping. If too few bursts are dropped, the isolation between FECs is only weak and as a result NC bursts can interfere with C bursts. However, if too many bursts are actively dropped, the overall burst loss probability increases as there may be the case that

[1] The intention of this proposed active dropping is very different from the active dropping called random early detection, RED [51] which is proposed in the context of TCP. Whereas dropping in RED aims to signal back to the TCP sender to throttle the sending rate, dropping in the context of Assured Horizon aims to leave available resources (wavelengths) unused in case a reserved burst will arrive in future.

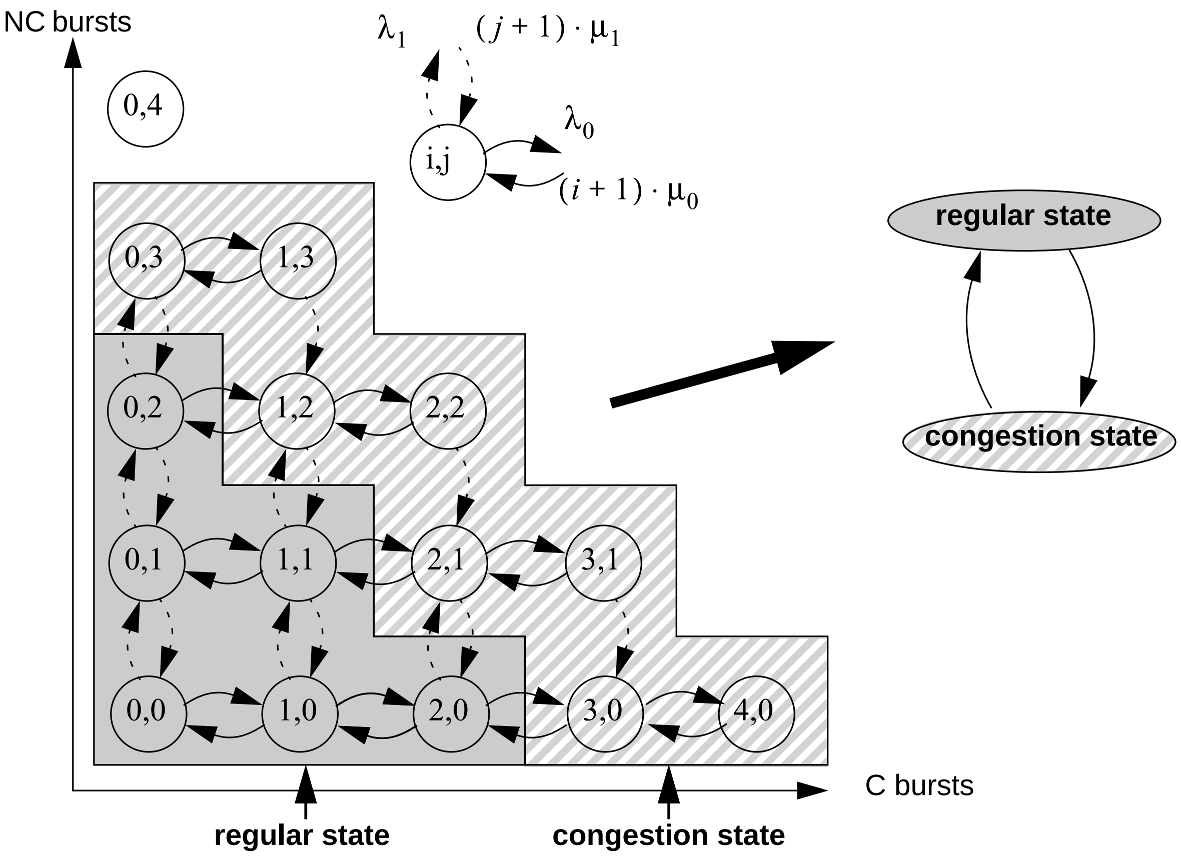


Figure 6.5: Trunk reservation admission control in a scenario with four servers and $\theta_{NC} = 0.5$; complete system states and reduced model

an NC burst is not admitted and no C burst follows that can reserve the left over wavelength. Thus, a compromise for a good dropping strategy has to be found.

In order to reduce the complexity and thus also the processing time in the core, a very simple dropping mechanism is suggested for the Assured Horizon framework. Hereby, the decision whether a burst has to be dropped or not is directly based on information which is available in the BHP without the necessity to carry out any calculations.

Following the general evaluations of multi-class loss systems presented in Section 4.2, the dropping mechanism of Assured Horizon is based on the trunk reservation admission control mechanism. By trunk reservation admission control, a burst of class i is only admitted if the normalized occupancy of the system does not exceed a class-dependent threshold θ_i whereas the highest priority class is always admitted, i. e., $\theta_i = 1$.

As core nodes of Assured Horizon only know two classes of bursts, C and NC, trunk reservation admission control with two classes, comparable to the introduction of trunk reservation in Section 4.2.2 is applied. Here, C bursts have the highest priority and are always admitted to the wavelengths reservation process whereas NC burst have lower priority and are only admitted if

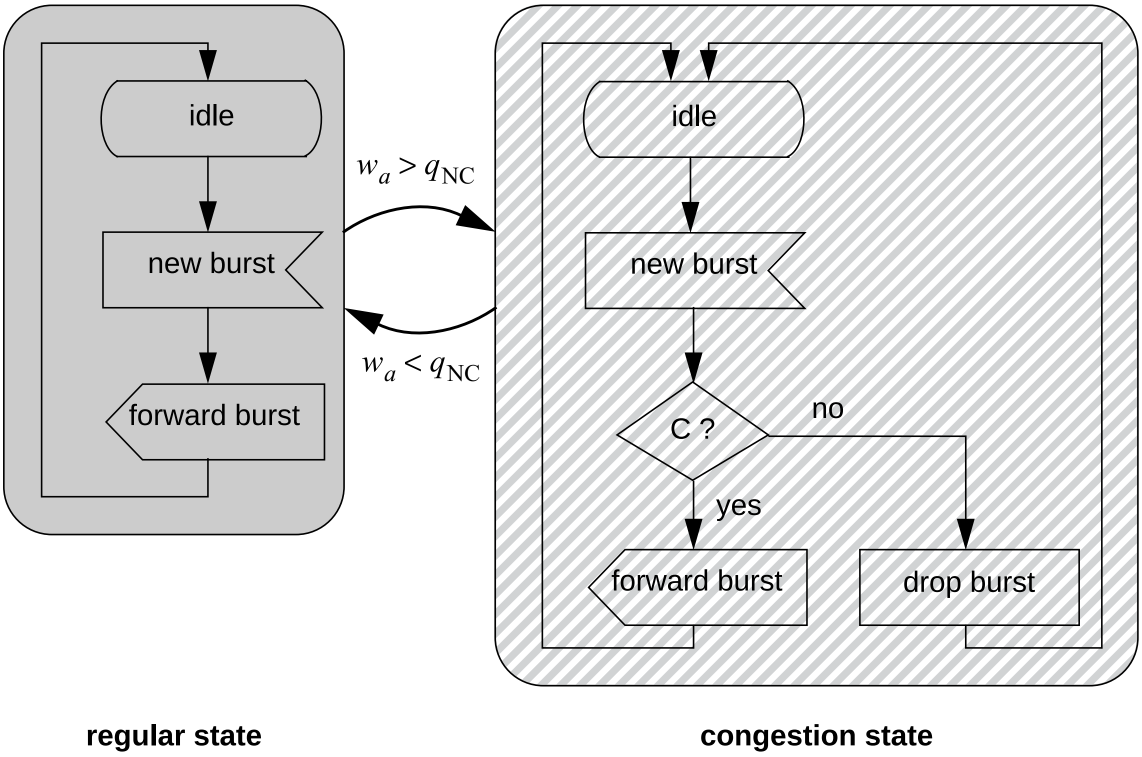


Figure 6.6: SDL-specification of the functionality of admission control

– at the time a BHP arrives – the normalized occupancy of the system does not exceed $\theta_{\mathrm{NC}} = q_{\mathrm{NC}}/n$. Hence, as also depicted in Figure 6.5, a core node only knows two states, a regular state where all bursts are admitted and a congestion state where only C bursts are admitted. In the example depicted in Figure 6.5, the already known system from Chapter 4 with four wavelengths is depicted with $\theta_{\mathrm{NC}} = 0.5$. Here, the regular state comprises the system states (0,0), (1,0), (2,0), (0,1), (1,1) and (0,2) and the congestion state all other states that can be reached. The system state (0,4) does neither belong to the regular state nor to the congestion state as it is never reached. As can be seen from Figure 6.5, trunk reservation admission control leaves from all states the last $n - q_{\mathrm{NC}}$ wavelengths for the exclusive use of C bursts.

The functionality of admission control is also depicted in Figure 6.6. Here, for the regular state and the congestion state, a respective specification in specification and description language, SDL (see, e. g., [90]), of the functionality is depicted. In regular state, upon arrival of a new burst, the burst is forwarded to the wavelengths reservation process and the burst dropper waits for the next arrival. In congestion state, a new arrived burst is only forwarded if it is marked as C whereas NC bursts are dropped.

State changes of the burst dropper are triggered from the wavelengths reservation mechanism dependent on the number of currently allocated wavelengths w_a which is taken as a measure of the current occupancy. If $w_a > q_{\mathrm{NC}}$ the dropper is in congestion state. Hence, neither the

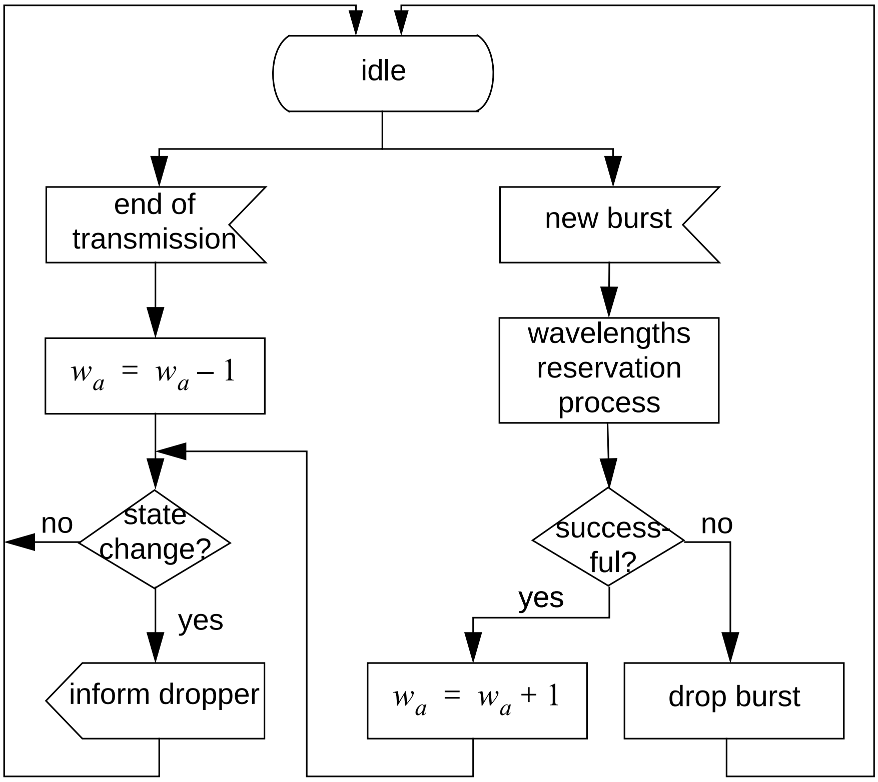


Figure 6.7: SDL-specification of functionality of reservation mechanism

dropper nor the reservation mechanism is required to perform any calculations to determine if a burst is dropped. An SDL-specification of the functionality of the reservation process mechanism is depicted in Figure 6.7. In this graph, it is distinguished whether a burst transmission ends or a new burst arrives.

Upon arrival of a new burst, the wavelengths reservation mechanisms (according to Horizon) is carried out. If it is successful, an internal variable counting the number of currently allocated wavelengths is increased and – if the system state changes – the dropper is informed and the system waits for the next arrival. If the wavelengths reservation process is not successful, the burst is dropped. If the end of a burst transmission occurs, the counter is decreased and, depending whether the system state is changed, the dropper is informed before the system returns to the idle state.

As already generally discussed in the context of Figure 4.13, dimensioning of q_{NC} depends on the objective for the carried traffic and is a trade-off between overall burst losses and isolation between FECs. The aim is that at the objective for the carried traffic, only a negligible number of C bursts cannot find an outgoing wavelength and thus have to be discarded from the reservation mechanism. By doing so, this OBS-QoS mechanism realizes isolation between FECs, as most bursts which are marked as C can find an outgoing wavelength. Accordingly, this traffic is guaranteed a negligible burst loss probability. In this scheme, there is no isolation between NC bursts which all experience the same service. Multiplexing gain is achieved by (dynamic) partial sharing of wavelengths between C and NC bursts. The dynamic results from dedicating the last $n - q_{\text{NC}}$ not allocated wavelengths to C bursts.

A possible extension of the trunk reservation mechanism to more than two classes is also possible. In such a scheme, marking at the network ingress has to differentiate NC bursts of different service classes whereas it does not make sense to distinguish between C bursts of different service classes. However, as this option of Assured Horizon makes the system more complicated, it is not in the focus of this thesis and will not be considered in the following.

Chapter 7

Modelling, Analysis and Performance Evaluation of Assured Horizon

In this chapter, the framework Assured Horizon which was introduced in Chapter 6 is evaluated with respect to its performance, see also [39]. In Section 7.1, an approximative analysis of the NC traffic share and the resulting burst loss probability is presented. In Section 7.2, the evaluation scenario as well as traffic model and system parameters are introduced. Finally, in Section 7.3, a detailed analytical and simulative performance evaluation is presented and discussed.

7.1 Approximative Analysis

The approximative analysis of Assured Horizon is also subdivided in two parts – policing including marking of bursts and enforcement – like the framework itself, see also Section 6.3 and Section 6.4 as well as Figure 7.2 and Figure 7.3. Policing of burst traffic at the edge of the network possibly leads to generation of NC bursts. This process depends on the reserved bandwidth envelope r_i, the threshold σ_i and the timeout interval τ_i of an assembly buffer i as well as IP traffic characteristics. The probability to generate an NC burst is analyzed in Section 7.1.1. Based on the share of NC bursts and the theory of trunk reservation admission control introduced in Section 4.2 the drop probability as well as the loss probability are obtained analytically in Section 7.1.2. Finally, in Section 7.1.3 an upper bound of the waiting time in an assembly buffer is obtained.

For simplicity of the formulæ, an index indicating an assembly queue is only considered in Section 7.1.2 where different assembly queues interact.

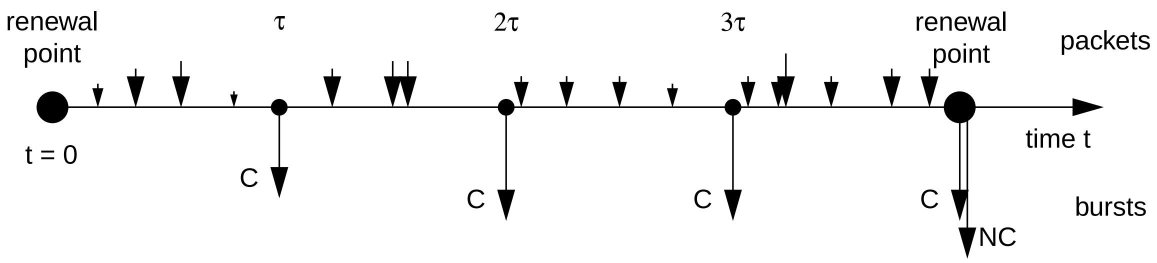


Figure 7.1: Packet arrival and burst departure point processes

7.1.1 Analysis of the NC Share

Marking of bursts in Assured Horizon can be considered as a renewal process where the departure of an NC burst is a renewal point as at that instance in time, an assembly buffer is completely emptied, see Figure 7.1. Therefore, the following analysis starts at such a renewal point. Random variable $Y(\tau)$ denotes the amount of Bytes arriving at an assembly buffer within time interval τ.

$$P\{Y(\tau) > y\} \tag{7.1}$$

is the probability that $Y(\tau)$ exceeds y. An NC burst is generated if the inflow to an assembly buffer is greater than its outflow plus a threshold σ. Under assumption of heavy traffic, a burst assembly buffer is a time-discrete system where C bursts are generated at multiples of τ, see Figure 7.1 and Figure 7.2. This is an approximation, as – in case an assembly buffer is empty – a new time interval begins if a new packet arrives to the assembly buffer and thus two consecutive timeouts may happen at a greater interval than τ. Thus, for the generation of an NC burst,

$$\text{inflow}(j \cdot \tau) > \text{outflow}(j \cdot \tau) + \sigma \text{ with } j = 1, 2, \ldots \tag{7.2}$$

has to be satisfied. As a second heavy traffic approximation, the outflow is estimated by the maximum outflow which is $r \cdot \tau$. Then, (7.2) can be inserted in (7.1) yielding the excess probability of σ in any interval smaller or equal to $j \cdot \tau$

$$P_{\text{excess}}(j \cdot \tau) = P\{Y(j \cdot \tau) > j \cdot r \cdot \tau + \sigma\} \text{ with } j = 1, 2, \ldots \tag{7.3}$$

However, for smaller inflow within a time interval and a small amount of Bytes in the assembly buffer, the outflow within the considered time interval may also be smaller. Thus, within several accumulated time intervals, an NC burst may be generated although $j \cdot r \cdot \tau + \sigma$ is not exceeded. As a consequence of this second approximation, $P_{\text{excess}}(j \cdot \tau)$ may underestimate the real excess probability.

In order to obtain the probability that an NC burst is generated in the j^{th} interval ($j > 1$), it has to be also considered that no NC burst is generated in previous intervals, thus

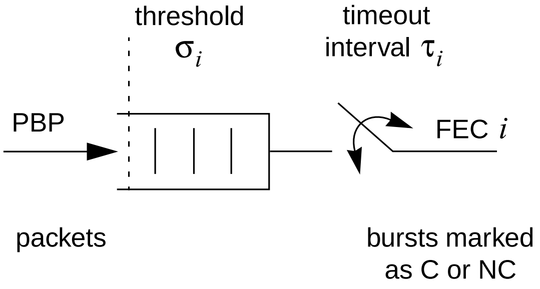


Figure 7.2: Burst assembly including policing and marking at an assembly buffer per FEC at the network edge

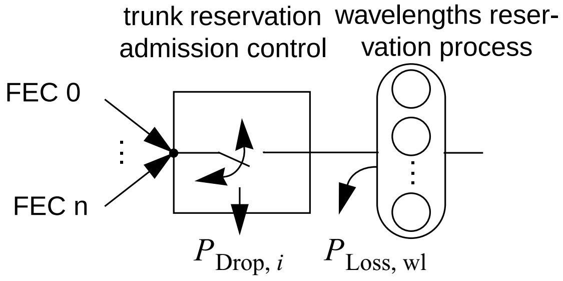


Figure 7.3: Enforcement of the policed bursts by dropping at the network core with trunk reservation admission control

$$P_{\text{NC}}(j \cdot \tau) = P_{\text{excess}}(j \cdot \tau) \cdot \prod_{i=1}^{j-1} (1 - P_{\text{excess}}(i \cdot \tau)), \; j > 1 \tag{7.4}$$

needs to be satisfied. Finally, the share of NC bursts, S_{NC}, is obtained from (7.4) by summation over all intervals j and multiplication with the factor $1/(1+j)$ which considers, depending on the interval j, the percentage of NC bursts.

$$S_{\text{NC}} = \sum_{j=1}^{\infty} \frac{1}{1+j} \cdot P_{\text{NC}}(j \cdot \tau). \tag{7.5}$$

Hereby, it is assumed that a non-zero reservation envelope exists and thus a C burst is sent out after each timeout interval as also depicted in Figure 7.1. For the share of C bursts

$$S_{\text{C}} = 1 - S_{\text{NC}} \tag{7.6}$$

follows directly. Summarizing, (7.5) and (7.6) describe the outcome of the marking process of one assembly queue. As the marking processes of different assembly queues are independent of each other, S_{NC} can be obtained independently for each assembly queue. S_{NC} is a very important performance metric, as admission to the wavelengths reservation process is only based on the share of NC bursts of class i, $S_{\text{NC}, i}$, and the occupancy of outgoing wavelengths.

7.1.2 Analysis of Burst Loss Probability

In order to obtain the burst loss probability $P_{\text{Loss}, i}$ of a class i, the probability that an NC burst of any class is actively dropped, $P_{\text{Drop, NC}}$, and thus not admitted to the wavelengths reservation process is calculated, see Figure 7.3 for the applied scenario. Herefore, $S_{\text{NC, all}}$ has to be considered according to

$$S_{\mathrm{NC,\,all}} = \frac{\sum_i A_i \cdot S_{\mathrm{NC},\,i}}{\sum_i A_i} \tag{7.7}$$

with offered traffic A_i of FEC i yielding separation of traffic in C and NC bursts (without distinction of classes) as it is seen by a core node. Based on this traffic, a two-class system (C and NC bursts) with TR has to be solved according to (4.13) yielding $P_{\mathrm{Drop,\,C}}$ and $P_{\mathrm{Drop,\,NC}}$. This now explains why the focus in Chapter 4 is on loss systems with only two classes. However, as C bursts are always admitted, only $P_{\mathrm{Drop,\,NC}}$ is required. Afterwards, the drop probability of a class i, $P_{\mathrm{Drop},\,i}$ can be obtained according to

$$P_{\mathrm{Drop},\,i} = S_{\mathrm{NC},\,i} \cdot P_{\mathrm{Drop,\,NC}} \tag{7.8}$$

as $P_{\mathrm{Drop,\,NC}}$ is independent of service classes. Thus, differentiation with respect to $P_{\mathrm{Drop},\,i}$ is only caused by different marking of bursts, i. e., $S_{\mathrm{NC},\,i}$. If, in contrast to the recommended operation in Chapter 6, the system is not properly dimensioned or no internal FDLs are applied, bursts that have already passed the admission control unit may fail to reserve a wavelength. Those losses, called $P_{\mathrm{Loss,\,wl}}$, have to be considered to obtain $P_{\mathrm{Loss},\,i}$ according to

$$P_{\mathrm{Loss},\,i} = P_{\mathrm{Drop},\,i} + (1 - P_{\mathrm{Drop},\,i}) \cdot P_{\mathrm{Loss,\,wl}}. \tag{7.9}$$

In case of recommended operation, $P_{\mathrm{Loss,\,wl}}$ can be neglected and $P_{\mathrm{Loss},\,i}$ follows

$$P_{\mathrm{Loss},\,i} \approx P_{\mathrm{Drop},\,i}. \tag{7.10}$$

7.1.3 Upper Bound for the Waiting Time in an Assembly Buffer

Besides the marking process at an assembly buffer which influences the burst loss probability P_{Loss}, the maximum waiting time $w_{\max}$ in an assembly buffer is of interest and will be evaluated in the following.

The worst case of the waiting time in an assembly buffer happens to a packet that – together with the packets already in the buffer – just exceeds threshold σ but not $\sigma + r \cdot \tau$. Furthermore, until the time when the packet is assembled to a burst and sent out, $\sigma + r \cdot \tau$ is not exceeded, i. e., the packet is sent out in a C burst.

According to (6.1), C bursts of length $r \cdot \tau$ Bytes[1] are sent every τ seconds. Consequently, $w_{\max}$ can be calculated individually for each assembly buffer by

[1] As packets have a discrete length and are not split in different bursts, the burst length of a C burst can exceed $r \cdot \tau$ by the maximum length of a packet minus one Byte.

$$w_{\max} = \left\lceil \frac{\sigma}{r \cdot \tau} + 1 \right\rceil \cdot \tau . \qquad (7.11)$$

Where $\lceil x \rceil$ denotes the smallest integer which is greater or equal to x. Form (7.11) it is obvious that a greater threshold σ or a greater timeout interval τ yield longer maximum waiting time. On the other hand, $w_{\max}$ can be reduced by greater reservation envelope r individually for each assembly buffer.

7.2 Evaluation Scenario

7.2.1 Traffic Model

The traffic model which is applied in the following evaluations is the M/G/∞ burst process[1] which is also called Poisson burst process, PBP, see, e. g., [114], [18]. M denotes here the negative-exponentially distributed interarrival time, G denotes a general burst length distribution and ∞ indicates that an infinite number of sources are superimposed. Mathematically, it is a limit of the superposition of on-off sources where the number of sources as well as the mean off duration of a source approaches infinity whereas the mean on time as well as the total mean rate are kept unchanged. As a result, each source is active only once [114].

A PBP is often suggested as simple model for aggregated traffic as it is close to the real traffic behavior (see, e. g., [100], [18], [17]) and already considers correlations between packets. In case of a Pareto distributed holding time with parameter $\alpha < 2$ it creates a self-similar process. An additional interesting property of such a process is that a superposition of independent PBP with the same burst length distributions is again a PBP with the sum of the arrival rates of the original PBPs [100]. However, it is still a model and thus a rough approximation as protocol specific characteristics, e. g., from transport control protocol, TCP, are not considered.

Bursts which are generated by such a PBP are called files in the following in order not to confuse these bursts with data bursts in OBS. Furthermore, this notation also indicates that data generated by a PBP can be interpreted as files. A file is segmented in packets of maximum length and a packet containing the remaining Bytes. These packets are sent out according to a maximum link speed. For the following analysis, the maximum link speed is assumed to be infinite, i. e., all packets belonging to a file arrive at the same time at the assembly buffer which is a worst case approximation. Evaluations per simulation will show the effects of a reduced maximum link speed which is equivalent to traffic shaping.

In Figure 7.4, an example of a file arrival process and the resulting packet arrival process is shown. In this figure, the correlation between packets is visible. Hereby, packets belonging to

[1] This burst process model should neither be confused with the queueing system which has the same notation, nor with the term burst in optical burst switching.

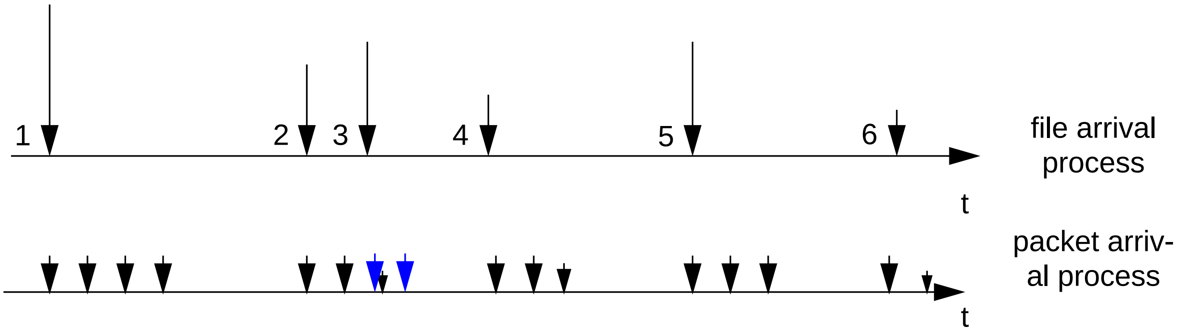


Figure 7.4: File arrival and packet departure processes

different files are treated if they have an access link each, i. e., packets of different files are not subject to be shaped. This is shown in Figure 7.4, where packets belonging to file 2 and file 3 overlap.

In order to determine the overall burst loss probability, $P_{\text{excess}}(j \cdot \tau)$ according to (7.3) is required and will be determined in the following sections.

7.2.1.1 File Length Independent Calculations

From, e. g. [91], it is known that the amount of arriving Bytes in a time interval τ of such a PBP is a compound distribution. The outer distribution is the distribution of the number of files which arrive within τ whereas the inner distribution covers the number of Bytes contained in a file.

According to, e. g. [91], the outer distribution is Poissonian with random variable N describing the number of arrivals in the time interval τ. Assuming an arrival rate λ, the distribution can be calculated according to

$$p_n(\tau) = P\{N = n\} = \frac{(\lambda \cdot \tau)^n}{n!} \cdot e^{-\lambda \cdot \tau} \tag{7.12}$$

The inner distribution – which will be specified later – has a distribution

$$p_x = P\{X = x\} \tag{7.13}$$

where X denotes the random variable describing the number of Bytes contained in a file.

The resulting distribution of the number of received Bytes in τ can either be obtained in frequency domain using the theory of generating function or in time domain by discrete convolution. In the following, a solution in time-domain will be presented.

Let Y be the number of received Bytes in τ. Then,

$$p_y = P\{Y = y\} = X_1 + X_2 + \ldots + X_N \tag{7.14}$$

is its distribution which can be obtained by n-fold discrete convolution. Under the assumption that the number N of arrivals is known, the probability that more than y Bytes arrive can be obtained by

$$p(y|n) = P\{Y > y|N = n\} = \sum_{y > (X_1 + X_2 + \ldots + X_n)} p_{x_1} \cdot p_{x_2} \cdot \ldots \cdot p_{x_n} . \qquad (7.15)$$

Applying the law of total probability and using (7.12) and (7.13),

$$p(y, \tau) = P\{Y > y\} = \sum_{n=0}^{\infty} P\{Y > y|N = n\} \cdot P\{N = n\} \qquad (7.16)$$

follows for the unconditioned probability.

7.2.1.2 Negative-Exponentially Distributed File Length

If the file length is negative-exponentially distributed, $P\{Y = y|N = n\}$ is an Erlang-n distribution (see, e. g. [91]). Thus, for (7.15) it follows

$$p(y|n) = P\{Y > y|N = n\} = \sum_{i=0}^{k-1} \frac{(\mu \cdot y)^i}{i!} \cdot e^{-\mu \cdot y} \qquad (7.17)$$

with rate μ. Finally, for the unconditioned probability comparable to (7.16) it follows

$$p(y, \tau) = P\{Y > y\} = e^{-(\lambda \cdot \tau + \mu \cdot y)} \cdot \sum_{n=1}^{\infty} \frac{(\lambda \cdot \tau)^n}{n!} \cdot \sum_{i=0}^{k-1} \frac{(\mu \cdot y)^i}{i!} . \qquad (7.18)$$

7.2.2 System Parameter

For the following performance evaluations, the scenario depicted in Figure 7.5 is taken as basis for simulations. Hereby, traffic for every FEC is generated by a PBP where files are generated and segmented in IP packets, see Section 7.2.1. The file length distribution is negative-exponential, hyperexponential[1] or Pareto[2] distributed with mean 10 kByte in accordance to measurements, see, e. g., [31]. All IP traffic of a FEC is assembled to bursts in an assembly queue according to the mechanism introduced in Section 6.3 and bursts are marked as C or NC in their BHP. After reservation in a local reservation mechanism and E-O conversion, optical bursts reach a core node. Prior to the wavelengths reservation process, a burst dropper controls admission to these wavelengths according to trunk reservation admission control strategy, see

[1] The hyperexponential distribution satisfies the symmetry condition $p \cdot h_1 = (1 - p) \cdot h_2$ where p is the branch probability and h_1 and h_2 are the mean values of the respective service times.

[2] In case file lengths follow a Pareto distribution, $\alpha = 1.6$ is chosen which results in self-similar packet traffic.

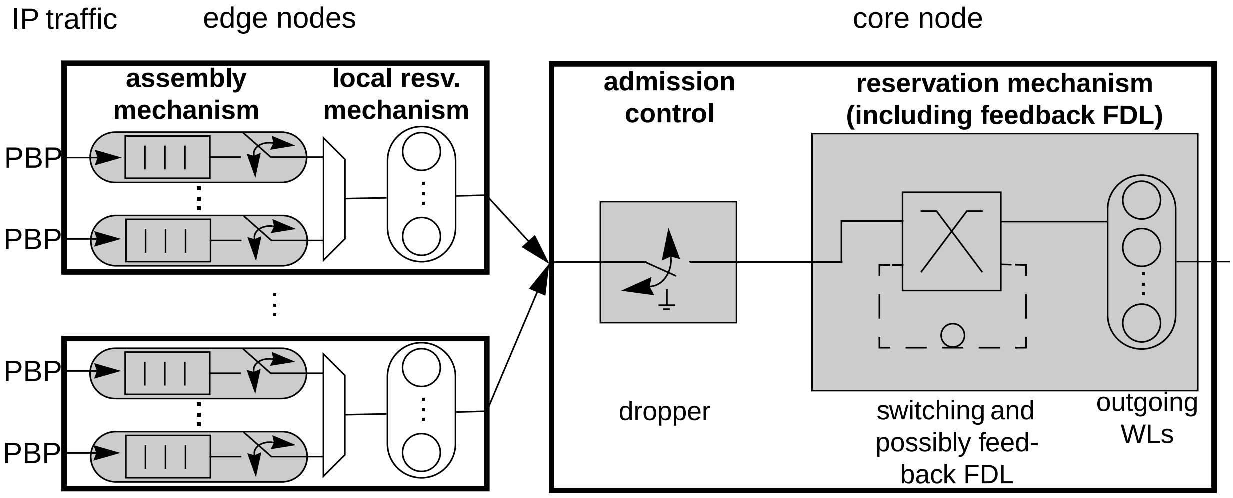


Figure 7.5: Block diagram of an evaluation scenario including edge nodes and a core node

Section 6.4.2. If all outgoing wavelengths are occupied, internal buffering with FDLs may be applied. However, in all simulations, no internal buffering is carried out. For studies with buffering, see, e. g., [54] and [56].

Parameters of such a system which influence the marking of NC bursts are reserved bandwidth envelope r_i, threshold σ_i in the assembly buffer, timeout interval τ_i, access bandwidth r_{access} as well as the file length distribution of a PBP. The impact of those parameters is evaluated in Section 7.3.1. Hereby, for the following evaluations, the overall reserved rate equals the overall mean rate m, i. e., $r = \sum r_i = m$ and is kept constant. Thus, in case of two classes and 30% share of high priority traffic, an increase in r_0 implies a decrease in r_1 according to

$$r_1 = (1 - 0{,}3 \cdot f_0) \cdot m_1 / 0{,}7 \tag{7.19}$$

with allocation factor $f_0 = r_0 / m_0$.

If not denoted differently, timeout interval $\tau_i = 1\text{ ms}$ and threshold $\sigma_i = m_i \cdot \tau_i$ are assumed for service class i. Section 7.3.2 discusses the burst characteristics which results from the burst assembly process.

Parameters which determine whether a burst is dropped in the core are TR threshold θ_{NC}, the number of wavelengths and possible internal buffering by FDLs. The influence of those parameters will be evaluated in Section 7.3.3.

7.3 Performance Evaluation

In this section, the dependencies of traffic and system parameters on the marking process at the assembly buffer are evaluated by the just introduced approximative analysis as well as simulations [161]. In Section 7.3.1, S_{NC} will be evaluated, i. e., the part of traffic which is marked as

NC, and thus, might be subject to policing by active burst dropping. The marking process is especially important as – according to (7.10) and (7.8) – the burst loss probability follows S_{NC} by reduction of a constant factor.

7.3.1 Marking at the Network Edge

7.3.1.1 Impact of the Reservation Envelope

In Figure 7.6, S_{NC} of both classes as well as $S_{\mathrm{NC,\,all}}$ are depicted against the allocation factor f_0. The reservation envelope of class 1 is hereby determined according to (7.19) in order to keep the overall reserved bandwidth constant. An increase in f_0 starting from $f_0 = 1$ leads to a decrease of f_1 and thus below the mean bandwidth of class 1.

It can be seen that an increase in f_i results in more bursts of class i to be within the reservation envelope, i. e., less NC bursts are required in order to transmit all Bytes which are in the assembly buffer. This figure is very important as it is the basis for service differentiation in the Assured Horizon framework. Dependent on f_0, the grade of service differentiation can be adjusted, e. g., for $f_0 = 1.6$ $S_{\mathrm{NC,\,1}}$ is two orders of magnitude greater than $S_{\mathrm{NC,\,0}}$ and – according to (7.10) – consequently the burst loss probability $P_{\mathrm{Loss,\,1}}$ will also be about two orders of magnitude greater than $P_{\mathrm{Loss,\,0}}$. If $f_0 < 1$, class 1 bursts have priority over class 0 bursts.

Also depicted in Figure 7.6 is the overall NC share $S_{\mathrm{NC,\,all}}$. It can be seen that if more weight is allocated to one class, $S_{\mathrm{NC,\,all}}$ is increased, i. e., more NC bursts are generated. As a consequence, the overall loss probability in the core is increased. This is the price that has to be paid for differentiated QoS in the Assured Horizon framework.

As already mentioned in Section 6.1.2, an analogon of the marking (and later enforcement) of bursts in the electronic world is a weighted scheduler with queueing. Hereby, the variation of the allocation factor f_0 corresponds to the variation of the weight at a weighted scheduler. This is also obvious from the progression of $S_{\mathrm{NC,}\,i}$ by, e. g., comparing the curves with the loss probability of a weighted scheduler depicted against the varying weight, see, e. g. [16].

Also from Figure 7.6, it can be seen that the approximative analysis matches the simulated results very well and consequently correctly reflects the dependencies of the marking process from system and traffic parameters. The small underestimation stems from both approximations indicated in Section 7.1.1. In the range $0.6 < f_i < 1.4$, the simulation is too optimistic due to rounding caused by granularity of packets and the adaption of the maximum burst size to the size of timeout intervals which – however – does not fully explain the difference between simulation and analysis. As the progression of S_{NC} is in principle the same for both classes, most of the following evaluations will focus on $S_{\mathrm{NC,\,0}}$.

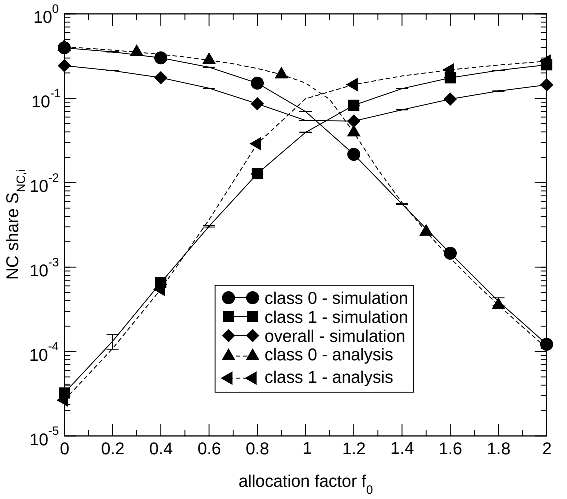


Figure 7.6: NC share of both classes $S_{\text{NC}, i}$ against allocation factor f_0

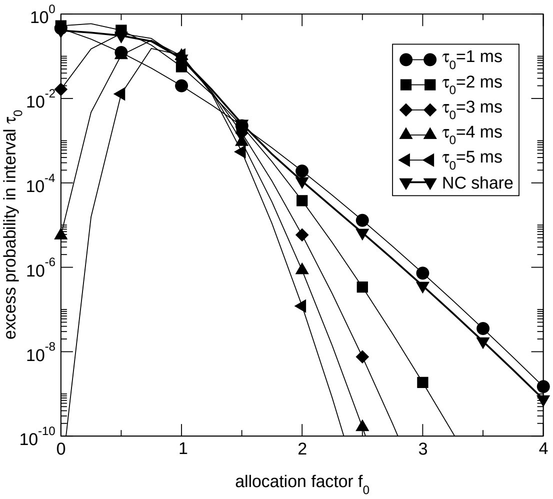


Figure 7.7: Overall NC share $S_{\text{NC, all}}$ against allocation factor f_0

In Figure 7.7, the shape of $S_{\text{NC}, 0}$ according to (7.5) and Figure 7.6 is explained for a broad range of f_0 by also depicting its components from (7.4), i. e., the probability that an NC burst is generated is depicted for the first five timeout intervals. Additionally, the NC share obtained by weighted summation of these five graphs is depicted. It should be mentioned here again, that the marking process strongly depends on the traffic characteristics and thus the graphs are not valid for general file length distributions, see Section 7.3.1.5. However, the principle dependency from f_0 can be seen. It can be also seen, that $P_{\text{NC}}(j \cdot \tau)$ is small if f_0 is either small or great. For small f_0 this can be explained with a great probability that σ_i is exceeded in a previous timeout interval whereas for great f_0, the probability that an NC burst is generated is generally small as C bursts are mostly sufficient to carry the offered traffic.

An upper bound for S_{NC} is 1 if only NC burst are sent. This is the case if no bandwidth is reserved. In case a small amount of bandwidth is reserved, an NC burst is sent out if the threshold σ_i in the assembly buffer is exceeded. If, additionally, σ_i is small, S_{NC} is bounded by $1/2$ as additionally to every C burst also a NC burst is sent. If σ_i is chosen to be the mean amount of Bytes which arrives within timeout interval τ_0 – which is the case in this scenario – the upper bound of S_{NC} is mainly determined by $P_{\text{excess}}(\tau)$ and $P_{\text{excess}}(2 \cdot \tau)$. The probability that a greater number of timeout intervals is required to exceed σ_i is small. Thus, a rough approximation for S_{NC} and small r_i assuming that $P_{\text{excess}}(\tau) = P_{\text{excess}}(2 \cdot \tau) = 1/2$ yields

$$S_{\text{NC}} \approx \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{5}{12} \approx 0.4167 \qquad (7.20)$$

which matches the simulated curves very well.

If f_0 is great, S_{NC} follows $P_{\text{excess}}(\tau)$ reduced by the factor 0.5. This can be explained with the fact that the considered traffic smooths out if the considered time interval increases (as it is not self-similar). Because the outflow of an assembly buffer strongly exceeds its inflow, the

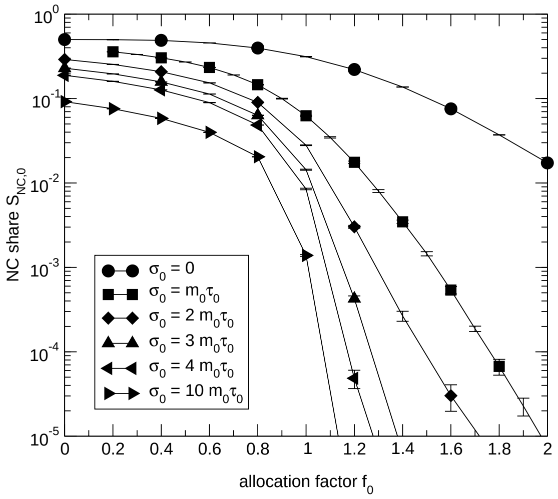


Figure 7.8: Excess probability and NC share $S_{\mathrm{NC},\,0}$ against allocation factor f_0

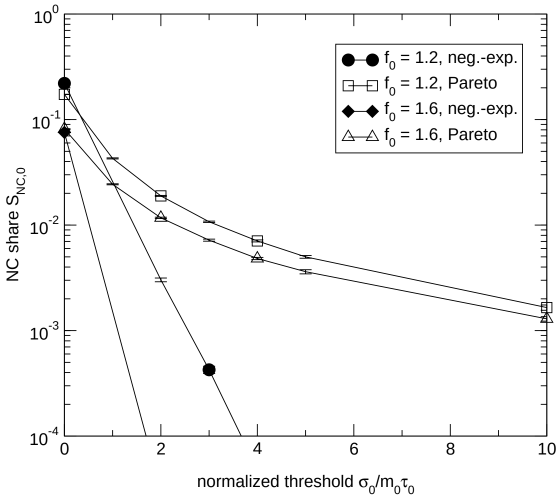


Figure 7.9: NC share $S_{\mathrm{NC},\,0}$ against normalized threshold $\sigma_0/(m_0 \cdot \tau_0)$

probability to exceed the threshold is primarily determined within one timeout interval. For the range between no bandwidth reservation and strong overallocation, a greater number of timeout intervals have to be considered in order to determine S_{NC}.

Summarizing this discussion, $S_{\mathrm{NC},\,0}$ is composed of three regions. The first region (small f_0) is determined by the threshold in an assembly buffer, the last region (great f_0) is determined by the excess of the outflow within one timeout interval and the region in between considers the excess probability in many timeout intervals.

7.3.1.2 Impact of the Threshold

The impact of threshold σ_i is evaluated in Figure 7.8 and Figure 7.9. In Figure 7.8, $S_{\mathrm{NC},\,0}$ is depicted against f_0, comparable to the scenario with negative-exponentially distributed file length depicted in Figure 7.6. In this graph, σ_i normalized by the mean amount of data within one timeout interval, i. e., $\sigma_0/(m_0 \cdot \tau_0)$ is an additional parameter. It can be seen that for an increase in $\sigma_0/(m_0 \cdot \tau_0)$, $S_{\mathrm{NC},\,0}$ is decreased for all values of f_0, which is already obvious from (7.2) and (7.3). Additionally, the slope of the decrease of $S_{\mathrm{NC},\,0}$ increases with increasing σ_i. It can be seen, that already an increase of $\sigma_0/(m_0 \cdot \tau_0)$ from 1 to 2 yields more than one order of magnitude less bursts marked as NC for $f_0 > 1$.

On the other hand, according to (7.19), a greater σ_i may also yield longer waiting time in the assembly buffer. Thus, the threshold is a compromise between more bursts which are sent marked as NC and longer waiting time in the assembly buffer.

In Figure 7.9 $S_{\mathrm{NC},\,0}$ is depicted against the normalized threshold $\sigma_0/(m_0 \cdot \tau_0)$. Here, not only results obtained for negative-exponentially distributed files are shown, but also results obtained from Pareto distributed file sizes. Like in the previous scenario, the access link rate r_{access} is infinite. It can be seen that an increase in $\sigma_0/(m_0 \cdot \tau_0)$ and a scenario with negative-exponen-

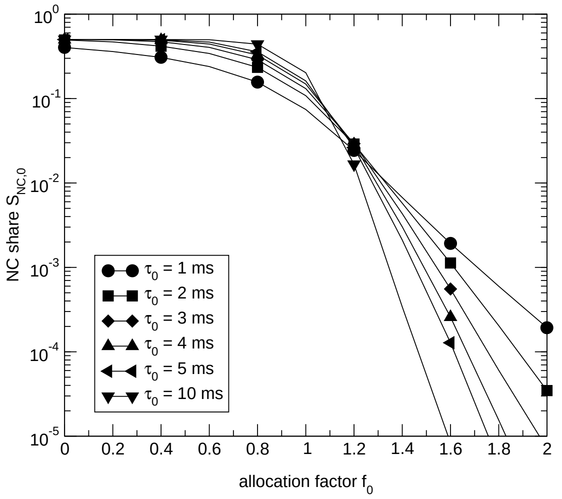


Figure 7.10: NC share $S_{\text{NC},\,0}$ against allocation factor f_0

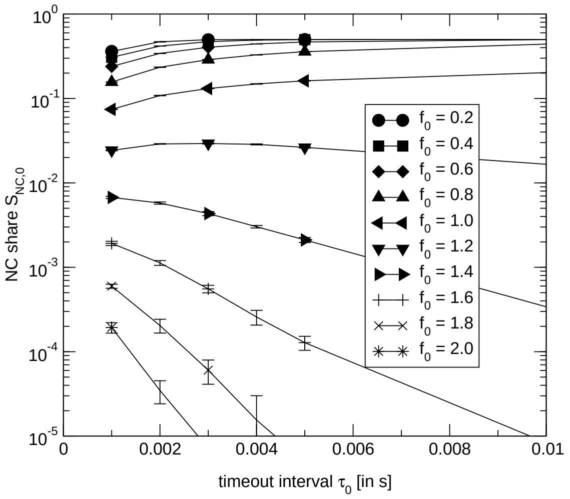


Figure 7.11: NC share $S_{\text{NC},\,0}$ against timeout interval τ_0

tially distributed file sizes yields an exponentially decrease of $S_{\text{NC},\,0}$. Hereby, f_0 determines the slope. Additionally, self-similar traffic generated from a PBP with Pareto distributed file sizes is also depicted in this figure. This traffic results in a decrease of $S_{\text{NC},\,0}$ which is much smaller than the one obtained for negative-exponentially distributed file sizes. Besides, the difference between different values of f_0 is small. This behavior is mainly caused by the fact that with $r_{\text{access}} = \infty$, an infinite amount of Bytes can reach an assembly buffer per time unit and the probability of very large files cannot be neglected. However, a greater σ_i still yields a lower $S_{\text{NC},\,0}$.

A direct consequence of the evaluations in this section is that applications which are sensitive to delay (variation) should be assembled in an assembly queue with smaller σ_i, whereas applications which are not delay sensitive can be assembled in an assembly queue with a greater σ_i in order to obtain lower burst losses. The impact on the waiting time in an assembly buffer is discussed in Section 7.3.1.6.

7.3.1.3 Impact of the Timeout Interval

In Figure 7.10 and Figure 7.11, the impact of the timeout interval τ_0 on $S_{\text{NC},\,0}$ is evaluated in a scenario with $\sigma_0/(m_0 \cdot \tau_0)$, negative-exponentially distributed file sizes and $r_{\text{access}} = \infty$. In Figure 7.10, $S_{\text{NC},\,0}$ is depicted against f_0 with τ_0 as additional parameter. From this figure, it can be seen that an increase in τ_0 only leads to a decrease in $S_{\text{NC},\,0}$ if $f_0 > 1$. If $f_0 < 1$, a greater τ_0 even leads to an increase in $S_{\text{NC},\,0}$. This can be explained from (7.3). In average, the amount of Bytes remaining in an assembly buffer after timeout is $(m_i - r_i) \cdot \tau_i$. Thus, if $r_i < m_i$, the amount increases with τ_i and thus also the probability that σ_i is exceeded. In the case where $r_i > m_i$, an additional amount of information is taken out of the assembly buffer at every timeout interval which decreases the probability that σ_i is exceeded in a greater number of timeout intervals.

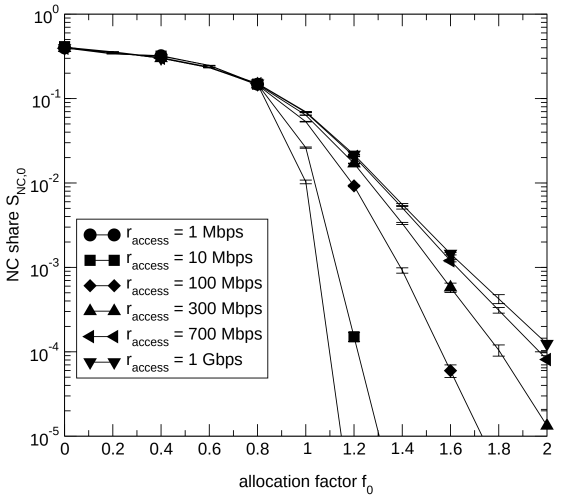


Figure 7.12: NC share against allocation factor f_0 – access link rate as parameter – neg.-exp. distributed file length

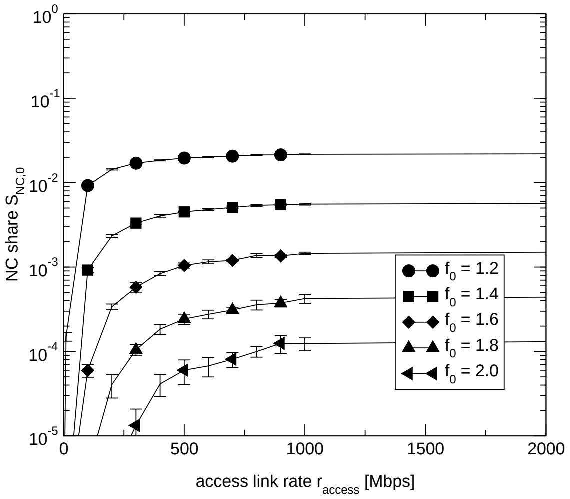


Figure 7.13: NC share against access link rate – neg.-exp. distributed file length

This characteristic is also visible from Figure 7.11 where $S_{\mathrm{NC},\,0}$ is depicted against τ_0 with f_0 as additional parameter. It can be seen that for smaller f_0, $S_{\mathrm{NC},}$ increases for increasing timeout interval and for greater f_0, $S_{\mathrm{NC},\,0}$ decreases. In between ($f_0 = 1.2$), S_{NC} has a flat maximum. With this figure, the impact of the first approximation can be now explained. As the heavy traffic approximation assumes that a new timeout interval starts immediately after the old one finished, timeout intervals are approximated too small, resulting in underestimation of $S_{\mathrm{NC},\,0}$ for underallocation and overestimation for overallocation.

Summarizing, τ_i can only be used to decrease $S_{\mathrm{NC},\,i}$ if $r_i > m_i$. This result has especially to be considered for dimensioning of τ_i for lower priority classes. Furthermore, it has to be considered that an increase in τ_i also increases the maximum waiting time in the assembly buffer according to (7.11), see also Section 7.3.1.6.

7.3.1.4 Impact of the Access Bandwidth

In all evaluations so far, results are presented for $r_{\mathrm{access}} = \infty$. In order to show how much $S_{\mathrm{NC},\,0}$ is reduced by a smaller r_{access}, the impact of the access link bandwidth on marking in an assembly buffer is evaluated. Reducing r_{access} results in shaping of packets belonging to a file. Especially for file size distributions with a large coefficient of variation, shaping has a significant impact as in case of no shaping the amount of Bytes which can possibly arrive to an assembly buffer within a time interval is not bounded.

In Figure 7.12 - Figure 7.15, the dependencies of the access link speed on $S_{\mathrm{NC},\,0}$ is depicted. In Figure 7.12 and Figure 7.13 files have negative-exponentially distributed lengths whereas in Figure 7.14 and Figure 7.15 files lengths follow a Pareto distribution.

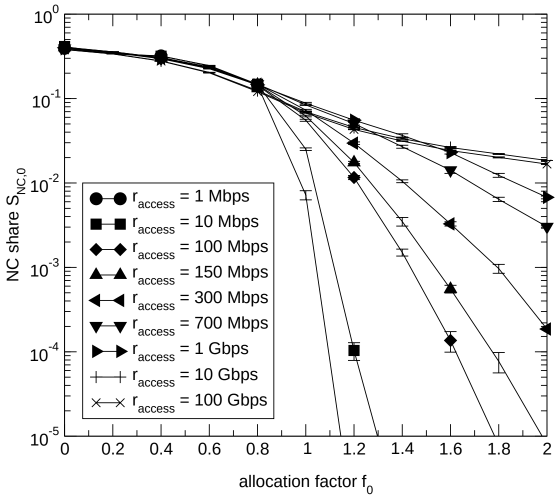


Figure 7.14: NC share $S_{NC,\,0}$ against allocation factor f_0 – access link rate as parameter – Pareto distributed file length

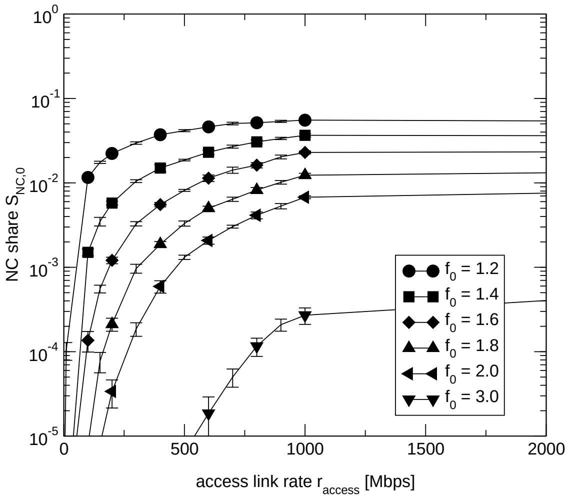


Figure 7.15: NC share $S_{NC,\,0}$ against access link rate – Pareto distributed file length

In Figure 7.12, $S_{NC,\,0}$ is depicted against f_0 with the link access rate r_{access} as parameter. In this graph, r_{access} is varied from 1 Mbps to 1 Gbps. It can be seen that the bandwidth of the access link has no impact in case of underallocation, but strongly influences the slope of the decrease of $S_{NC,\,0}$ for increasing f_0 as smaller access bandwidth strongly shapes the traffic which arrives at an assembly queue.

In Figure 7.13, the same scenario is depicted in a different way. Here, $S_{NC,\,0}$ is depicted against r_{access} and f_0 is an additional parameter. From this graph, the access link rate where saturation starts can be determined, e. g., for $f_0 = 1.6$ $S_{NC,\,0}$ is hardly influenced if the access link bandwidth is greater than 500 Mbps.

In Figure 7.14 and Figure 7.15, the file length is Pareto distributed with $\alpha = 1.6$, i. e., it has an infinite variance and thus the traffic is self-similar. Here, the influence of r_{access} is even more visible as the probability that a very large file is generated – and thus immediately contained in the assembly buffer – cannot be neglected. In comparison to Figure 7.12, it can be seen that a small access link bandwidth, e. g., 150 Mbps, results in roughly the same shape of $S_{NC,\,0}$. However, for greater r_{access} the possible decrease of $S_{NC,\,0}$ is only small. It can be seen that even for $f_0 = 2$, $S_{NC,\,0}$ is hardly reduced for great r_{access}.

This fact is depicted more clearly in Figure 7.15 where $S_{NC,\,0}$ is depicted against r_{access}. Here, for the aformentioned overallocation of $f_0 = 1.6$, $S_{NC,\,0}$ increases until 1 Gbps and remains one order of magnitude greater compared to negative-exponentially distributed files.

7.3.1.5 Impact of the File Length Distribution

From Figure 7.12 and Figure 7.14 it can be already assumed that the slope of $S_{NC,\,0}$ is about the same for both file length distributions in a scenario with $r_{access} = 150\text{ Mbps}$. In order to

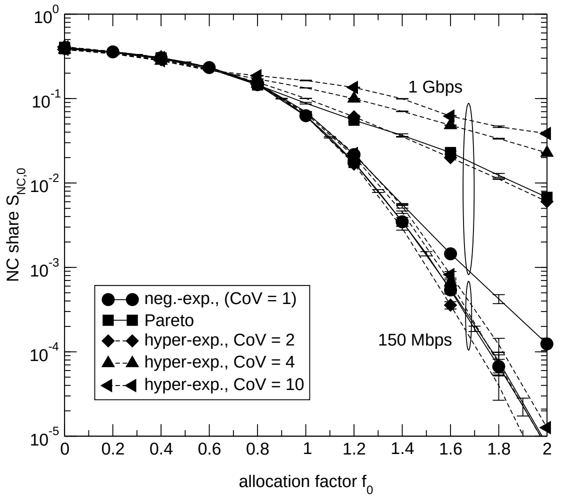


Figure 7.16: NC share against allocation factor – file length distribution

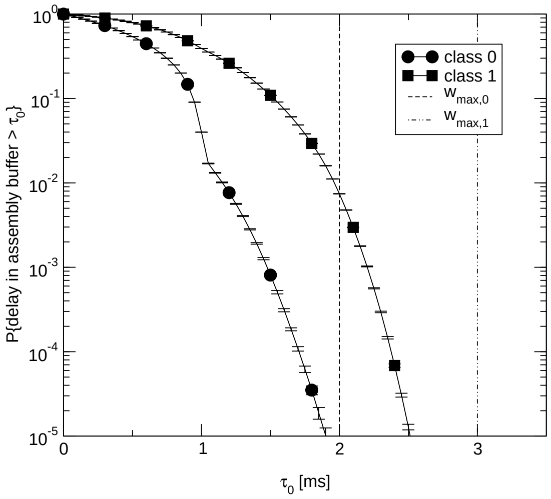


Figure 7.17: Complementary distribution function of the delay in an assembly buffer

confirm this assumption, Figure 7.16 depicts $S_{\mathrm{NC},\,0}$ against f_0 for different file length distributions where r_{access} equals 150 Mbps as well as 1 Gbps.

From Figure 7.16, it can be seen that in case of shaped traffic in the access link, the file length distribution has only small impact on $S_{\mathrm{NC},\,0}$. Even self-similar traffic and hyperexponentially distributed files with varying coefficient of variance, CoV, have very similar shape compared to negative-exponentially distributed files. On the contrary, $r_{\mathrm{access}} = 1\ \mathrm{Gbps}$ results in significantly changed marking behavior. Whereas $S_{\mathrm{NC},\,0}$ of traffic with negative-exponentially distributed files still drops down quickly, Pareto and hyperexponentially distributed files yield a much higher share of NC bursts.

7.3.1.6 Waiting Time in an Assembly Buffer

In Figure 7.17, the complementary distribution function of the waiting time in an assembly buffer is depicted. Hereby, the scenario of $f_0 = 1.6$ and thus, according to (7.19), $f_1 = 0.74$ is depicted. For class 0 bursts, a knee is visible indicating bursts leave the assembly buffer within one timeout interval ($\tau_0 = 1\ \mathrm{ms}$) and burst that have to wait another timeout interval. For class 1 bursts, this knee is not visible as less bandwidth than the mean bandwidth is allocated and thus a strong decrease at the end of a timeout interval does not take place.

The maximum waiting time for both classes, $w_{\mathrm{max},\,i}$ according to (7.11) is also depicted in this figure. In an assembly buffer of a class 0 burst, the maximum waiting time is 2 ms whereas it is 3 ms in a class 1 assembly buffer. However, as already discussed in the context of the difference between analysis and simulation, the simulation applies exponential weighting in order to be able to use some part of bandwidth which is reserved but was not used in the past. Therefore, the maximum waiting time obtained by analysis in (7.11) may be exceeded with a small probability.

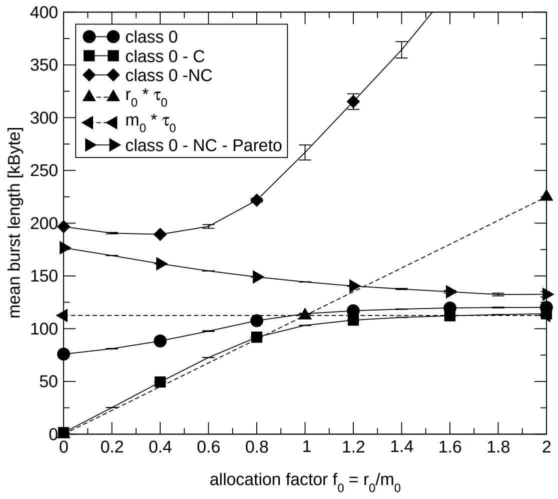


Figure 7.18: Mean burst length l_i against allocation factor f_0

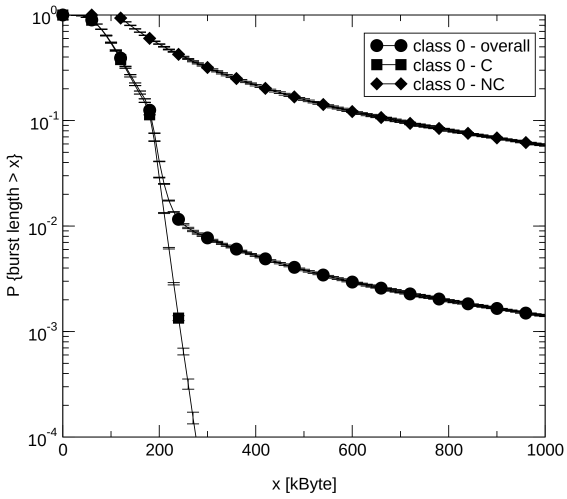


Figure 7.19: Burst length distribution of class 0

7.3.2 Resulting Burst Characteristics

All following evaluations in this section are carried out in a scenario with Pareto distributed file sizes and infinite access rate in order to emphasize the impact of system parameters on burst characteristics. Thus, the results depicted in this section are extreme values which are exaggerating the burst characteristics of traffic which is not self-similar or which is shaped by a lower access rate.

Like in previous sections, in all following evaluations, only results of class 0 are depicted and $\sigma_0 = (m_0 \cdot \tau_0)$ and $\tau_0 = 1 \text{ ms}$ are chosen as parameters. As the burst assembly mechanism is carried out independently for different FECs, burst characteristics of every FEC can be influenced independently according to the influence of the parameters described in the following sections.

7.3.2.1 Impact of the Reservation Envelope

Figure 7.18 depicts the resulting overall mean burst length of class 0, l_0, as well as mean burst length of compliant bursts of class 0 bursts, $l_{0,\,\mathrm{C}}$ and the mean burst length of non-compliant bursts of class 0 bursts, $l_{0,\,\mathrm{NC}}$, against f_0. Additionally, $r_0 \cdot \tau_0$ and $m_0 \cdot \tau_0$ are shown.

It can be seen that, $l_{0,\,\mathrm{C}}$ follows

$$l_{0,\,\mathrm{C}} = \begin{cases} r_0 \cdot \tau_0 & f_0 < 1 \\ m_0 \cdot \tau_0 & f_0 > 1 \end{cases}. \tag{7.21}$$

For $f_0 < 1$ this is motivated by the assembly strategy described in (6.1) which says that C bursts with maximum length $r_0 \cdot \tau_0$ are assembled. For $f_0 > 1$, $l_{0,\,\mathrm{C}}$ flattens out and reaches

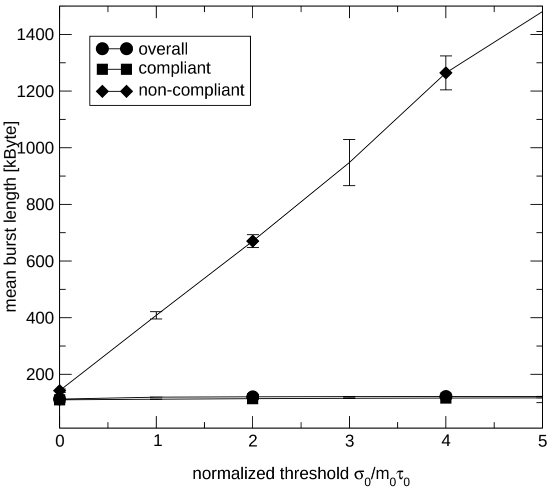


Figure 7.20: Mean burst length l_0 against normalized threshold

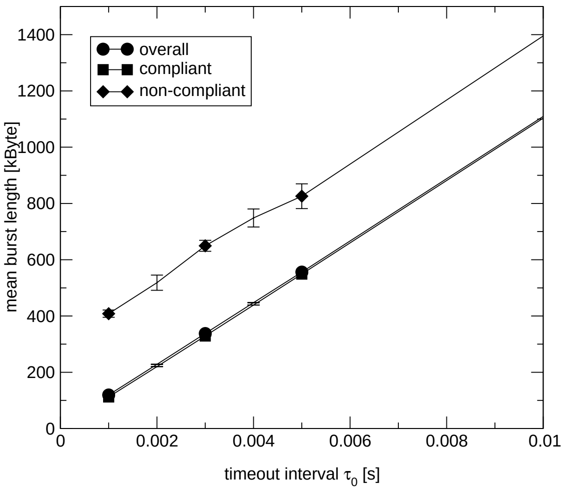


Figure 7.21: Mean burst length l_0 against timeout interval τ_0

the boundary value $m_0 \cdot \tau_0$ which corresponds to the mean amount of Bytes that arrives within timeout interval τ_0. Not depicted here, this behavior is rather independent of the IP traffic characteristics. Hence, $l_{0,\,\mathrm{C}}$ is almost the same also for negative-exponentially distributed files. As the share of NC burst gets smaller for increasing f_0, see also Figure 7.6, the difference between l_0 and $l_{0,\,\mathrm{C}}$ decreases. In the contrary to $l_{0,\,\mathrm{C}}$, $l_{0,\,\mathrm{NC}}$ strongly depends on the characteristics of the traffic that is assembled. In case of Pareto distributed file sizes, $l_{0,\,\mathrm{NC}}$ increases linearly for increasing f_0 whereas in case of negative-exponentially distributed file sizes, $l_{0,\,\mathrm{NC}}$ decreases for increasing f_0. This can be explained as NC bursts transport the amount of information that exceeds the reserved bandwidth envelope and thus contains peaks of IP traffic.

This behavior is also confirmed by the resulting burst length distributions which are depicted in Figure 7.19 for $f_0 = 1.6$. Whereas the burst length distribution of C bursts drops quickly and is very similar for different IP traffic characteristics, the variance of the IP file size distribution is captured by NC bursts. This behavior is especially advantageous as C bursts (and hence the smoothed burst traffic) are always admitted to the wavelengths reservation process whereas NC burst are only admitted if the carried traffic is low. Thus, the undesirable impact of long lasting congestions caused by very long NC bursts is moderated. For completeness, the overall burst length distribution of class 0 is also depicted. It can be seen that the impact of NC bursts on the overall burst length distribution is significant, although the share of NC bursts is very small.

7.3.2.2 Impact of the Threshold

In Figure 7.20, the impact of the threshold in the assembly buffer σ_0 on the mean burst length is depicted. As it is obvious from the functionality of the burst assembly described in Section 6.3, σ_0 has no impact on $l_{0,\,\mathrm{C}}$ whereas $l_{0,\,\mathrm{NC}}$ increases linearly for increasing σ_0. Because of the small share of NC bursts, l_0 almost equals $l_{0,\,\mathrm{C}}$.

7.3.2.3 Impact of the Timeout Interval

The impact of an increase of the timeout interval τ_0 on the mean burst length is evaluated in this section. As expected from the description of the assembly mechanism in Section 6.3, both, $l_{0,\,\mathrm{C}}$ and $l_{0,\,\mathrm{NC}}$ increase linearly from increasing τ_0. Like in the previous section, l_0 is very similar to $l_{0,\,\mathrm{C}}$ because of the small share of NC bursts.

7.3.3 Enforcement at the Network Core

In this section, the focus is on the probability that a burst is dropped and thus is not admitted to take part in the wavelengths reservation process. Furthermore, the probability that a burst is lost because the wavelengths reservation process fails is considered. Hence, in contrast to previous evaluations, not only the assembly at the ingress is considered but the focus is on the superposition of assembled traffic at a core node. Herefore, a number of edge nodes send bursts to one core node which applies trunk reservation admission control, has a certain number of wavelengths and carries out the wavelengths reservation process according to the Horizon reservation mechanism, see Section 3.3.2. If not denoted differently, trunk reservation threshold $\theta_{\mathrm{NC}} = 0.75$ is assumed for all evaluations. In this section, only the one-node scenarios are considered whereas network scenarios are considered in Appendix A.

The chosen scenario, i. e., the number of edge nodes that sends traffic to a core node, will influence the resulting performance of a system as a greater number of edge nodes yields a greater number of bursts that may arrive within the transmission time of a burst at a core node and thus may cause a collision or at least the rejection of NC bursts due to a greater number of simultaneously allocated wavelengths. The approximative analysis represents the superposition of an infinite number of edge nodes which send traffic with a negative-exponentially distributed interarrival time. This is a very rough approximation as – according to the burst assembly mechanisms introduced in Section 6.3 – the burst interarrival time of one burst assembly node is approximative constant.

In order to highlight these differences, two scenarios will be considered in the following. The first scenario is a *great network* where traffic of 50 assembly nodes with two different FECs each representing a high and a low priority class is generated. Hence, a core nodes receives traffic of 100 different FECs. This scenario can be applied as upper bound and also to compare simulation results with results obtained analytically. The second scenario represents a *small network*, e. g. the size of Germany see [64], where 10 edge nodes which distinguish 2 FECs each sent traffic to a core node, i. e. the core nodes receives traffic of 20 FECs.

If the scenario is changed, i. e., from a higher number of assembly nodes to a lower number of assembly nodes, the file interarrival time is reduced according to the ratio of assembly nodes which results in a reduced offered traffic. As a consequence, the reserved bandwidth envelope

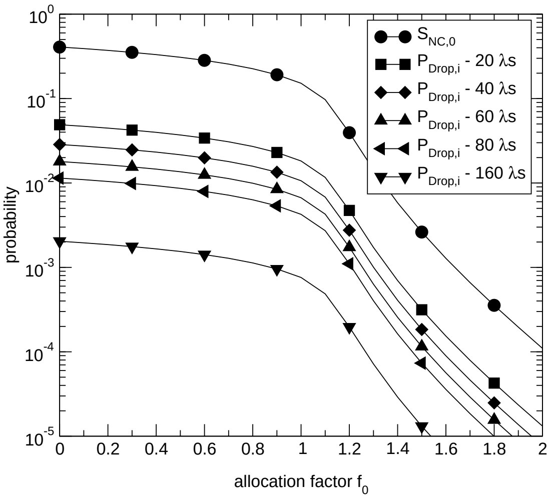


Figure 7.22: $P_{\text{Drop, 0}}$ for different number of wavelengths in the core

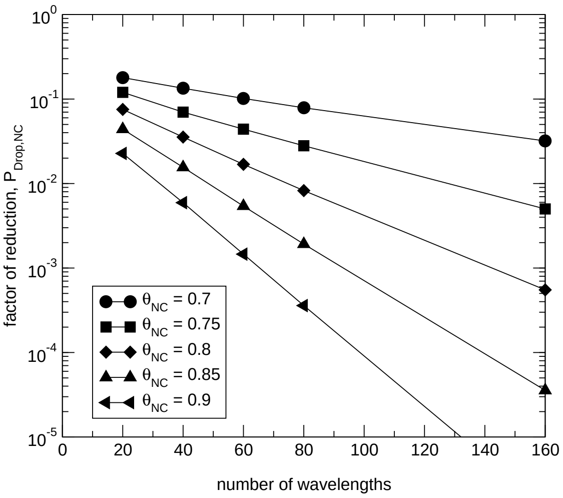


Figure 7.23: Factor of reduction against number of wavelengths

r_i is reduced accordingly. Additionally, in order to obtain the same burst characteristics, the timeout interval τ_i is increased (also according to the ratio of assembly nodes). Following $\sigma_0 = m_0 \cdot \tau_0$, the threshold in an assembly queues remains unchanged.

In case the scenario is kept, but the number of wavelengths in the core is changed, the adaption is carried out accordingly in order to remain the burst characteristics.

7.3.3.1 Impact of Number of Wavelengths on the Burst Drop Probability

In Figure 7.22, $S_{\text{NC, 0}}$ as well as $P_{\text{Drop, 0}}$ according to (7.10) are depicted. Thus, $P_{\text{Drop, 0}}$ is obtained by duplication of $S_{\text{NC, 0}}$ and reduction by a constant factor which results from the trunk reservation admission control. For the calculation of the factor of reduction $P_{\text{Drop, NC}}$, a constant $S_{\text{NC, 0}}$ is assumed for the solution of (4.13) which is an approximation. For all curves in Figure 7.22, $\theta_{\text{NC}} = 0.75$ is applied. It can be seen that a higher number of wavelengths in the core yields a lower drop probability. According to the discussion in Chapter 4, this can be explained by the increased multiplexing gain.

The factor of reduction is depicted in Figure 7.23 for different values of θ_{NC}. This graph corresponds to the loss probability of a request of a low priority class depicted in Figure 4.10 and Figure 4.11. Like already generally discussed in Section 4.2, the decrease of the drop probability is exponential with increasing number of wavelengths and the slope of the decease depends on θ_{NC} and thus on the grade of differentiation. A smaller θ_{NC} yields a smaller decrease, but also greater differentiation. Hence, as can be seen from Figure 7.23, in case of great differentiation, a higher number of wavelengths yields a significantly lower burst drop probability.

In Figure 7.24, results obtained by analysis and simulation are compared for 20 and 60 wavelengths, respectively in the large network scenario. The analysis matches the behavior quite well, however, it can be seen that for greater overallocation the analysis underestimates

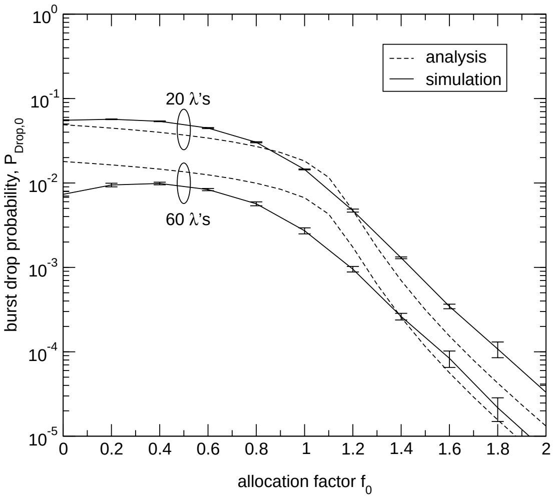


Figure 7.24: Comparison of $P_{\text{Drop, 0}}$ by simulation and analysis in a large network

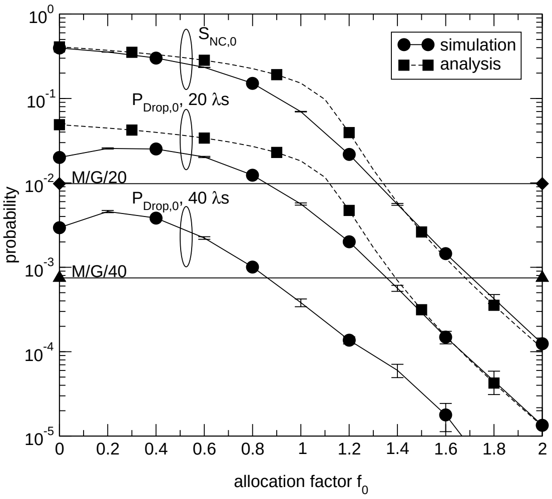


Figure 7.25: Comparison of $P_{\text{Drop, 0}}$ by simulation and analysis in a small network

$P_{\text{Drop, 0}}$. This behavior can be explained by the approximation already discussed in the context of the analysis. Another very important evidence of Figure 7.24 is that all results obtained individually in Section 7.3.1 for each assemble queue are the basis for the drop probability in the core.

In Figure 7.25, $S_{\text{NC, 0}}$ as well as $P_{\text{Drop, 0}}$ according to (7.10) are obtained by analysis and simulation. In contrary to the just discussed results, the small network scenario is applied. It can be seen that for a small number of wavelengths, e. g. 20, the analysis matches the simulation very well. For clarity, only a simulated curve is depicted for 40 wavelengths whereas the analyzed cures can be taken from Figure 7.22 and Figure 7.24. It can be seen that for a greater number of wavelengths, the analysis strongly overestimates $P_{\text{Drop, 0}}$. However, the principal shape is still reasonably approximated with a factor of reduction that is smaller than the one obtained from trunk reservation admission control. The much smaller drop probability in the small network scenario stems from the fact that only a small number of bursts may arrive at the same time and cause congestions. The greater the number of wavelengths, the smaller the probability that θ_{NC} is exceeded and as a consequence an NC burst is dropped. Thus, this effect can be also explained by the theory of finite source loss systems, see, e. g., [91] where the number of sources is small compared to the number of servers.

Also depicted in Figure 7.25 are loss probabilities of the M/G/n loss systems with $A/n = 0.6$, $E_{1,20}(12)$ and $E_{1,40}(24)$, respectively, obtained by (4.3). As expected, $P_{\text{Drop, 0}}$ exceeds the loss probability of an M/G/n loss system for underallocation. However, if the reserved bandwidth envelope exceeds the mean rate, $P_{\text{Drop, 0}}$ is smaller than $E_{1,x}(A)$. Hereby, $E_{1,x}(A)$ is intersected roughly for $f_0 = 1$. This is also an important result confirming the reasonable performance of the Assured Horizon framework.

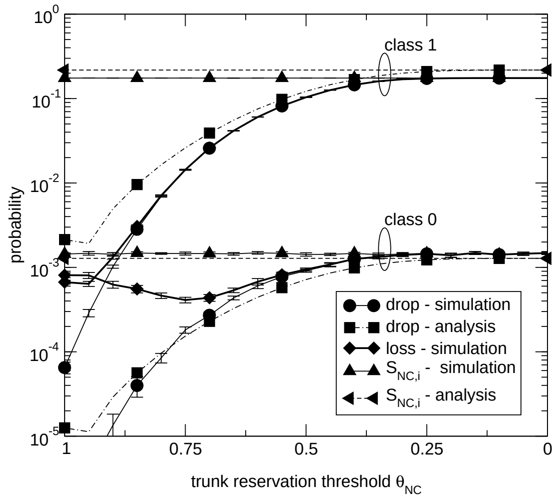


Figure 7.26: $P_{\text{Drop}, i}$ and $P_{\text{Loss}, i}$ against trunk reservation threshold θ_{NC}

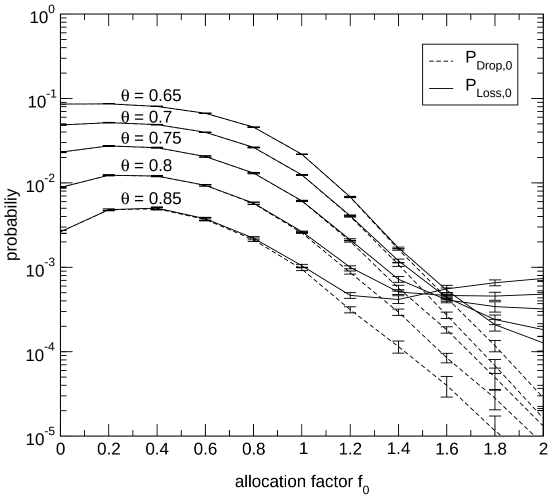


Figure 7.27: $P_{\text{Drop}, i}$ and $P_{\text{Loss}, i}$ against allocation factor f_0

7.3.3.2 Impact of Trunk Reservation Threshold

All evaluations in this section are carried out for the small network scenario. As already indicated generally in Section 4.2 as well as for the Assured Horizon framework in Figure 7.23, the trunk reservation threshold θ_{NC} strongly influences the burst drop probability as it is the only parameter in the core. A greater θ_{NC} yields greater service differentiation, but also increases the overall burst drop probability. Therefore, the impact of θ_{NC} comparable to the general shape depicted in Figure 4.13 is shown in Figure 7.26.

For $S_{\text{NC}, i}$ as well as $P_{\text{Drop}, i}$ analyzed and simulated curves are depicted for $f_0 = 1.6$. Additionally, in order to underpin the negligence of $P_{\text{Loss, wl}}$ in (7.10), simulated curves of $P_{\text{Loss}, i}$ are depicted. It can be seen that $P_{\text{Drop}, i}$ is increased with decreasing θ_{NC} and approaches $S_{\text{NC}, i}$ as – for the limit of $\theta_{\text{NC}} = 0$ – all NC bursts of all classes are dropped by TR. In case of no service differentiation ($\theta_{\text{NC}} = 1$), $P_{\text{Loss}, i}$ and $P_{\text{Drop}, i}$ differ significantly. However, this is an operation point which is not meaningful. Instead, for $\theta_{\text{NC}} = 0.75$, $P_{\text{Loss}, i}$ approaches $P_{\text{Drop}, i}$, even without additional internal buffering. It can be seen that the analysis captures the simulated curves quite well for meaningful values of θ_{NC}.

Finally, in Figure 7.27, the impact of θ_{NC} is discussed in a different context. Here, $P_{\text{Drop}, 0}$ as well as $P_{\text{Loss}, 0}$ for varying θ_{NC} are depicted against the allocation factor f_0 in the scenario of a small network. Again, it can be seen that a greater θ_{NC} yields a lower $P_{\text{Drop}, 0}$. However, for greater overallocation, $P_{\text{Loss}, 0}$ and $P_{\text{Drop}, 0}$ differ significantly as too many bursts are admitted to the wavelengths reservation process. This results in a $P_{\text{Loss}, 0}$ which is even greater for greater θ_{NC}. In case of internal buffering by FDLs, $P_{\text{Loss}, 0}$ can be reduced to follow $P_{\text{Drop}, 0}$ again.

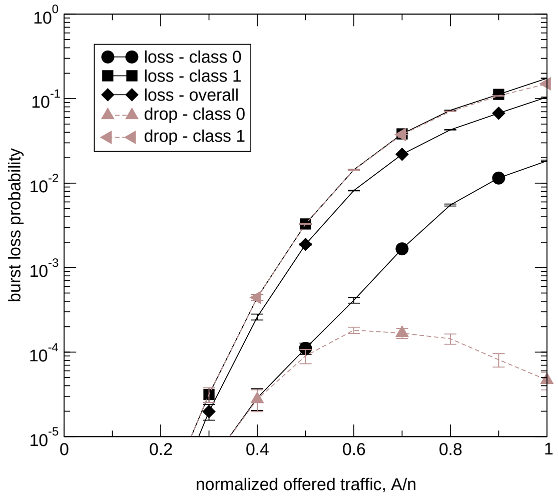


Figure 7.28: $P_{\mathrm{Drop},\,i}$ and $P_{\mathrm{Loss},\,i}$ against normalized offered traffic

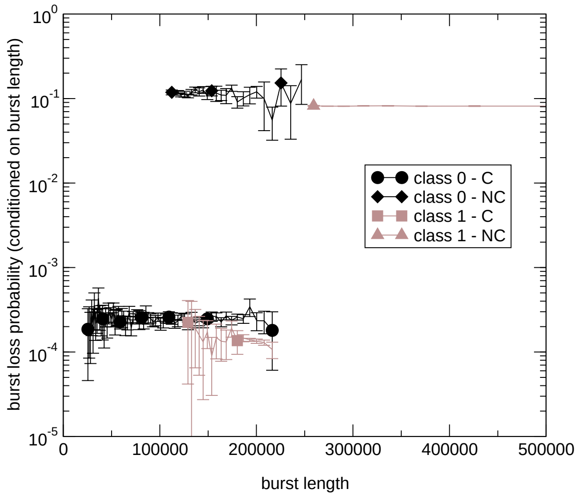


Figure 7.29: $P_{\mathrm{Loss},\,i}$ conditioned on the actual burst length

7.3.3.3 Impact of Offered Traffic

In Figure 7.28, $P_{\mathrm{Loss},\,i}$ as well as $P_{\mathrm{Drop},\,i}$ are depicted against the normalized offered traffic A/n comparable to Figure 4.14 and Figure 4.15. It can be seen that a good grade of differentiation with respect to the burst loss probability is kept over the whole range of offered traffic. Furthermore, the difference between $P_{\mathrm{Loss},\,i}$ and $P_{\mathrm{Drop},\,i}$ which was already discussed in the context of Figure 7.26 is visible. Whereas $P_{\mathrm{Loss},\,1}$ follows $P_{\mathrm{Drop},\,1}$ directly, there is a significant difference between $P_{\mathrm{Loss},\,0}$ and $P_{\mathrm{Drop},0}$ for $A/n > 0.6$. Due to the increased number of NC burst which are not admitted to the wavelengths reservation process for greater offered traffic, the probability that a burst of class 0 is dropped is even decreased. If internal buffering is applied, $P_{\mathrm{Loss},\,0}$ can be reduced in order to follow $P_{\mathrm{Drop},0}$. By doing so, the loss probability of class 0 can be kept reasonably constant for a great interval of offered traffic, even if A/n approaches 1.

7.3.3.4 Impact of Burst Length

In order to show that the performance of Assured Horizon does not depend on the actual burst length, the burst loss probability conditioned on the actual burst length is depicted in Figure 7.29. This figure is comparable to Figure 5.5 which is discussed in Section 5.3.3. It can be seen that the conditioned burst loss probability does neither depend on the class of a burst nor on its actual length. Also visible in this figure is the different loss probability for C and NC bursts. Thus, one major design requirement of Assured Horizon which is derived from a shortcoming of offset-based QoS is satisfied.

7.4 Summary

In the first part of this chapter, an approximative analysis of the burst loss probability of the Assured Horizon framework was presented which consists of two steps, comparable to policing of bursts at the edge and enforcement of bursts at a core node. In this approximation, the probability that a certain threshold in an assembly buffer is exceeded and thus a NC burst is generated is obtained independently of the IP traffic characteristics which are considered in a later step of the approximation. A major advantage of this approximation is that the excess probability can be calculated separately for each assembly buffer. The second step of this approximation considers the admission control to the wavelengths reservation process in front of every core node which is carried out by a burst dropper. The dropping probability follows the theory of trunk reservation admission control. In addition to the approximation of the burst loss probability, an upper bound of the waiting time in an assembly buffer was presented.

Dependencies of the NC share per assembly buffer on the threshold in an assembly buffer σ, the timeout interval τ, the rate of the access link r_{access} and the file lengths distribution, respectively are highlighted. It is shown that the NC share can be reduced by an increase in σ, an increase in τ in case of overallocation and by a smaller r_{access}, respectively. The file length distribution has a strong impact in case of great r_{access} and is almost insignificant in case of small r_{access}. As a consequence of the presented evaluations, a reasonable parameter setting of burst assembly queues is possible.

In order to get a deeper understanding of the effects, the impact of the above introduced parameters on the burst characteristics are presented. It is shown that the mean length of C bursts follows $r_0 \cdot \tau_0$ in case of underallocation and $m_0 \cdot \tau_0$ in case of overallocation and their burst length distribution drops down very quickly, almost independent from the IP traffic characteristics. In the contrary, both, the mean burst length as well as the burst length distribution of NC bursts strongly depend on the IP traffic characteristics.

The third part of this chapter focuses on the comparison of the presented approximative analysis and simulations as well as on the dependence of the results on various parameters in the core. It is shown that the approximation matches the simulation reasonably well and thus, the major characteristics of the Assured Horizon framework are captured by this analysis.

For the scenario of a small network as well as the scenario of a great network, the burst drop probability $P_{\text{Drop},\,i}$ is depicted depending on the number of wavelengths. It is shown that the approximation with the theory of trunk reservation admission control captures the results reasonable well for greater networks. However, for the small network scenario and a greater number of wavelengths, $P_{\text{Drop},\,i}$ is overestimated by the approximation.

As already indicated generally in the context of the trunk reservation theory in Chapter 4, it is also shown here that a greater trunk reservation threshold θ_{NC} yields better differentiation but also a greater burst loss probability.

Finally, the difference between $P_{\mathrm{Loss},\,i}$ and $P_{\mathrm{Drop},\,i}$ are outlined in a scenario without buffering in a core node. For reasonable dimensioning and controlled offered traffic, $P_{\mathrm{Loss},\,i}$ and $P_{\mathrm{Drop},\,i}$ are similar. However, for small service differentiation (great θ_{NC}) or normalized offered traffic greater than 0.6, $P_{\mathrm{Loss},\,i}$ an $P_{\mathrm{Drop},\,i}$ differ significantly.

Thus, QoS-differentiation with respect to burst loss probability is achievable comparable to a weighted scheduler. Because the burst admission control at every core node only distinguishes between C and NC bursts, bursts of any class are subject to be dropped. However, if a class sends only traffic within the reserved bandwidth envelope, no burst is intentionally dropped. Furthermore, a stateless core is realized which is also an indicator that the presented framework is simple and efficient.

7.4.1 Comparison to the Offset-Based OBS-QoS Mechanism

In comparison to the offset-based OBS-QoS mechanism which is evaluated in Chapter 5, the following conclusions can be drawn:

As it was a requirement for the design of Assured Horizon, the burst loss probability does neither depend on traffic characteristics of a different class nor on the actual length of a burst. Thus the unpleasant characteristics of the offset-based OBS-QoS mechanism are overcome. Additionally, isolation between FECs is achieved which may allow to operate an OBS system with several FECs having all the same priority, but are protected from each other.

A very important milestone that is achieved by Assured Horizon is the control of the offered traffic which is the basis for low and controlled burst loss probabilities as well as the basis for a guaranteed service.

However, the very low burst loss probabilities obtained by the offset-based OBS-QoS mechanism for the highest priority class achieved in the Assured Horizon framework, as no class can exclusively reserve all wavelengths.

Chapter 8

Conclusions and Further Work

In this dissertation, mechanisms for QoS differentiation in optical burst switched networks are evaluated and compared. Chapter 2 surveys on transport network architectures and motives the evolution towards an IP-over-WDM-based architecture also by current work of different standardization bodies.

In Chapter 3, optical burst switching, OBS, as a promising representative of an IP-over-WDM-based transport network architecture is introduced and burst assembly mechanisms as well as burst reservation mechanisms are outlined. Furthermore, currently in literature reported OBS-QoS mechanisms are classified and discussed. As major OBS-QoS mechanisms, offset-based, segmentation-based and active-dropping based mechanisms are distinguished. In the evaluations, segmentation-based mechanisms are not considered any more as they give up the assumption that a burst is an atomic unit which leads to highly increased complexity in the core.

In Chapter 4, the theory of loss systems is presented as this theory provides a deeper understanding of possible OBS-QoS mechanisms – which, in this thesis, work without mandatory buffering in the core. The major conclusions of this chapter are also requirements for the design of a new OBS-QoS framework in Chapter 6, namely: (i) the normalized offered traffic has to be controlled far below 1 in order to achieve a reasonable multiplexing gain. (ii) For multi-class loss systems, trunk reservation admission control is a simple, efficient and robust mechanism to control the number of requests of different classes. However, the price for service differentiation that has to be paid is an increased overall loss probability as requests may not be admitted to the system in order to leave space for higher priority classes and no higher priority request arrives.

Chapter 5 compares the performance of one-class OBS reservation mechanisms JIT, Horizon and JET and shows that added complexity in the reservation protocols of Horizon and JET leads to decreased burst loss probability compared to JIT. However, the additional complexity of JET yields only better results compared to Horizon if the offset is varying.

In the second part of Chapter 5, an approximative analysis of the burst loss probability of offset-based OBS-QoS mechanisms is presented. A very central building block in this approximation which is based on the well-known Erlang-B formula is that in order to obtain the burst loss probability of a high priority class, the forward recurrence time of the burst length distribution of lower priority classes contributes. Thus, the performance strongly depends on burst characteristics of other (lower priority) classes and worsens if the variance is increased. Another unpleasant feature that is revealed in this section is that the burst loss probability of lower priority classes depends on the actual length of a burst which contradicts an efficient operation with longer low priority bursts and shorter high priority bursts. Longer low priority bursts are discarded with a higher probability that shorter low priority bursts.

As a consequence of these shortcomings revealed in Chapter 5 and based on the requirements of multi-class loss systems discussed in Chapter 4, a new OBS-QoS framework called Assured Horizon is introduced in Chapter 6. This framework consists of a new burst assembly mechanism, a new burst reservation mechanism as well as the communication between them. The major building blocks of Assured Horizon are (i) a coarse-grained or static bandwidth reservation envelope, (ii) policing including marking of bursts at the network ingress combined with the burst assembly mechanism and (iii) enforcement of the marked bursts in case of congestion by a burst dropper in front of every core node. Hereby, the burst reservation mechanism is an active-dropping based mechanism where the access to the wavelengths reservation mechanism is controlled by trunk reservation admission control. This allows to realize a very simple but efficient solution for QoS differentiation.

Finally, in Chapter 7, the performance of Assured Horizon is evaluated. In the first part of this chapter, an approximative analysis of the burst loss probability is presented. After (mathematical) formulation of a traffic model and system parameters, results obtained by the analysis are compared to results obtained by simulation. Herefore, the impact of major parameters of the Assured Horizon framework on the burst loss probability are discussed.

In Appendix A, the performance analysis of OBS-QoS mechanisms is extended from the one-node case to a network scenario. In the first part, it is shown that the offset-based OBS-QoS mechanism leads to an increased number of classes in case the basic offset in not compensated by an FDL in front of every core node. The second part evaluates Assured Horizon and shows that, in a networking scenario, the burst loss probability of NC bursts increases with increasing number of hops, however, S_{NC} is an upper bound for the overall burst loss probability.

Further work could improve the second step of the approximative analysis by considering the real number of assembly nodes. Thus, the burst drop probability obtained by the trunk reservation admission control could be obtained by an exact formula considering the real finite source model and not only an approximation by Markovian traffic. By doing so, the decomposition approach which yields the burst loss probability would be more accurate. Finally, a verification by a prototypical realization/demonstration would round up the evaluation of the new OBS-QoS framework Assured Horizon.

Bibliography

[1] P. ABRY, D. VEITCH: "Wavelet analysis of long-range-dependent traffic." *IEEE Transactions on Information Theory,* Vol. 44, No. 1, Jan. 1998, pp. 2-15.

[2] S. R. AMSTUTZ: "Burst Switching – An Introduction." *IEEE Communications Magazine*, Vol. 21, 1983, pp. 36-42.

[3] S. R. AMSTUTZ: "Burst Switching – An Update." *IEEE Communications Magazine*, Vol. 27, No. 6, 1989, pp. 50-57.

[4] M. A. ALI, A. SHAMI, C. ASSI, Y. YINGHUA, R. KURTZ: "Architectural options for the next-generation networking paradigm: is optical Internet the answer?" *Photonic Network Communications*, Vol. 3, Nos. 1/2, July 2001, pp. 7-21.

[5] J. ANDERSON, J. S. MANCHESTER, A. RODRIGUEZ-MORAL, M. VEERARAGHAVAN: "Protocols and Architectures for IP Optical Networking." *Bell Labs Technical Journal*, Vol. 4, No. 1, Jan. 1999, pp. 105-124.

[6] American Nationale Standards Institute, T1.105: SONET. Basic Description including Multiplex Structure, Rates and Formats, 1991.

[7] D. AWDUCHE, Y. REKHTER: "Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects." *IEEE Communications Magazine,* Vol. 39, No. 3, March 2001, pp. 111-116.

[8] I. BALDINE, G. ROUSKAS, H. PERROS, D. STEVENSON: "JumpStart: A just-in-time signalling architecture for WDM burst switched networks." *IEEE Communications Magazine,* Vol. 40, No. 2, Feb. 2002, pp. 82-89.

[9] R. BALLART, Y.-C. CHING: "SONET: Now It's the Standard Optical Network." *IEEE Communications Magazine*, Vol. 27, No. 6, 1989, pp. 8-15.

[10] A. BANERJEE, J. DRAKE, J. P. LANG, B. TURNER, K. KOMPELLA, Y. REKHTER: "Generalized multiprotocol label switching: An overview of routing and management enhancements." *IEEE Communications Magazine*, Vol. 39, No. 1, Jan. 2001, pp. 144-150.

[11] A. BANERJEE, J. DRAKE, J. LANG, B. TURNER, D. AWDUCHE, L. BERGER, K. KOMPELLA, Y. REKHTER: "Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques." *IEEE Communications Magazine*, Vol. 39, No. 7, July 2001, pp. 144-151.

[12] G. BERNSTEIN, B. RAJAGOPALAN, D.. SPEARS, "OIF UNI 1.0 – controlling optical networks." *OIF white paper,* available at http://www.oiforum.com.

[13] G. M. BERNSTEIN, J. YATES, D. SAHA: “IP-centric control and management of optical transport networks.” *IEEE Communications Magazine*, Vol. 38, No. 10, Oct. 2000, pp. 161-167.

[14] S. BLAKE, D. BLACK, M. CARLSON, E. DAVIES, Z. WANG, W. WEISS: *An architecture for differentiated services*, IETF, RFC 2475, Dec. 1998.

[15] F. BONOMI, W. K. FENDICK: “The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service.” *IEEE Network*, Vol. 9, No. 2, March 1995, pp. 25-39.

[16] S. BODAMER, K. DOLZER, M. LORANG, W. PAYER, R. SIGLE: *Scheduling and bandwidth allocation for flow aggregates in multi-service networks,* Internal Report No. 30, IND, University of Stuttgart, Aug. 1999.

[17] S. BODAMER, J. CHARZINSKI, K. DOLZER: „Dienstgütemetriken für elastischen Internetverkehr.“ *Fachzeitschrift Praxis der Informationsverarbeitung und Kommunikation (PIK),* Bd. 25, Nr. 2, April 2002, S. 82-89.

[18] S. Bodamer: „*Verfahren zur relativen Dienstgütedifferenzierung in IP-Netzknoten*“ Monographie, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 2002.

[19] P. BONENFANT, A. RODRIGUEZ-MORAL: “Framing techniques for IP over fiber.” *IEEE Network*, Vol. 15, No. 5, July 2001, pp. 12-18.

[20] S. C. BORST, D. MITRA: “Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic.” *IEEE Journal on Selected Areas in Communications,* Vol. 16, No. 5, June 1998, pp. 668-678.

[21] P. E. BOYER, D. P. TRANCHIER: “A Reservation Principle with Applications to the ATM Traffic control.” *Computer Networks and ISDN Systems*, Vol. 24, 1992, pp. 321-334.

[22] C. A. BRACKETT: “Dense Wavelength Division Multiplexing Networks: Principles and Applications.” *IEEE Journal on Selected Areas in Communications,* Vol. 8, No. 8, Aug. 1990, pp. 948-964.

[23] R. BRADEN, D. CLARK, S. SHENKER: *Integrated services in the Internet architecture: an overview*, IETF, RFC 1633, July 1994.

[24] R. BRADEN, L. ZHANG, S. BERSON, S. HERZOG, S. JAMIN: *Resource ReSerVation Protocol (RSVP),* IETF, RFC 2205, Sep. 1997.

[25] U. BRIEM: *Verbindungslose Datenkommunikation über ATM-Weitverkehrsnetze: Architekturen, Protokolle und Verkehrsleistung – 71. Bericht über verkehrstheoretische Arbeiten*, Dissertation, Universität Stuttgart, 1998.

[26] H. BUCHTA, C. M. GAUGER, E. PATZAK, J. SANITER: “Limits of effective throughput of optical burst switches based on seminconductor optical amplifiers.” *Proceedings of the Optical Fiber Communication Conference (OFC 2003),* Atlanta, March 2003.

[27] R. Bush, D. Meyer "Some Internet architectural guidelines and philosophy", IETF, RFC 3439, Dec. 2002.

[28] F. CALLEGATI: "Optical buffers for variable length packets." *IEEE Communications Letters*, Vol. 4, No. 9, Sep. 2000, pp. 292-294.

[29] F. CALLEGATI, H. C. CANKAYA, Y. XIONG, M. VANDENHOUTE: "Design issues of optical IP routers for Internet backbone applications." *IEEE Communications Magazine*, Vol. 37, No. 12, Dec. 1999, pp. 124-128.

[30] V. CERF, R. KAHN: "A protocol for packet network interconnection." *IEEE Transactions on Communications*, Vol. 22, No. 10, May 1974, pp. 637-648.

[31] J. CHARZINSKI: "Measured HTTP performance and fun factors." *Proceedings of the 17th International Teletraffic Congress (ITC 17)*, Salvador da Bahia, Brazil, Dec. 2001, pp. 1063-1074.

[32] Charzinski, J., Dolzer, K., Färber, J., Koehler, S., Krieger, U., Macfayden, R., Markovitch, N., Tutschku, K., Vicari, N., Vidacs, A., Virtamo, J.T.: Traffic Measurement and Data Analysis. COST-257 Final Report: Impacts of new services on the architecture and performance of broadband networks (Ed. Tran-Gia, P., Vicari, N.), compuTEAM 2000, pp. 33-42, ISBN 3-930111-10-1.

[33] H. M. CHASKAR, S. VERMA, R. RAVIKANTH: "A framework to support IP over WDM using optical burst switching." *Proceedings of the Optical Networks Workshop*, Richardson, TX, Jan. 2000.

[34] H. M. CHASKAR, S. VERMA, R. RAVIKANTH: "Robust transport of IP traffic over WDM using optical burst switching." *Optical Networks Magazine*, July 2002.

[35] Y. CHEN, M. HAMDI, D. H. K. TSANG: "Proportional QoS over OBS Networks." *Proceedings of IEEE GLOBECOM 2001*, San Antonio, Dec. 2001.

[36] M. E. CROVELLA, A. BESTAVROS: "Self-similarity in World Wide Web traffic: evidence and possible causes." *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, Dec. 1997, pp. 835-846

[37] I. DE MIGUEL, M. DUESER, P. BAYVEL: "Traffic Load Bounds for Optical Burst-Switched Networks with Dynamic Wavelength Allocation." *Proceedings of the IFIP TC6 5th International Working Conference on Optical Network Design and Modelling (ONDM 2001)*, Vienna, Feb. 2001.

[38] A. DETTI, M. LISTANTI: "Application of tell & go and tell & wait reservation strategies in a optical burst switching network: a performance comparison." *Proceedings of the 8th IEEE International Conference on Telecommunications (ICT 2001)*, Bucharest, June 2001.

[39] K. DOLZER: "QoS in optical burst switching networks – a new performance analysis of the assured horizon framework." *Proceedings of the 18th International Teletraffic Congress (ITC 18)*, Berlin, Sep. 2003, pp. 881-900.

[40] K. DOLZER: “Assured Horizon – An efficient framework for service differentiation in optical burst switched networks.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002)*, Boston, July 2002.

[41] K. DOLZER: “Assured Horizon – A new combined framework for burst assembly and reservation in optical burst switched networks.” *Proceedings of the European Conference on Networks and Optical Communications (NOC 2002)*, Darmstadt, June 2002.

[42] K. DOLZER, C. GAUGER: “On burst assembly in optical burst switching networks – a performance evaluation of Just-Enough-Time.” *Proceedings of the 17th International Teletraffic Congress (ITC 17)*, Salvador da Bahia, Brazil, Sep. 2001, pp. 149-160.

[43] K. DOLZER, C. GAUGER, J. SPÄTH, S. BODAMER: “Evaluation of reservation mechanisms for optical burst switching.” *AEÜ International Journal of Electronics and Communications*, Vol. 55, No. 1, Jan. 2001.

[44] K. DOLZER, W. PAYER: “On aggregation strategies for multimedia traffic.” *Proceedings of the 1st Polish-German Teletraffic Symposium (PGTS 2000)*, Dresden, Sep. 2000.

[45] K. DOLZER, W. PAYER, M. EBERSPÄCHER: “A simulation study on traffic aggregation in multi-service networks.” *Proceedings of the IEEE Conference on High Performance Switching and Routing (ATM 2000)*, Heidelberg, June 2000, pp. 157-165.

[46] K. DOLZER, S. KÖHLER, C. SCOGLIO: “Internet performance.” *COST-257 Final Report: Impacts of new services on the architecture and performance of broadband networks*, P. Tran-Gia, N. Vicari (Eds.), compuTEAM, Würzburg, Sep. 2000, pp. 53-64.

[47] M. DUESER, P. BAYVEL: “Bandwidth Utilisation and Wavelength Re-Use in WDM Optical Burst-Switched Packet Networks.” *Proceedings of the IFIP TC6 5th International Working Conference on Optical Network Design and Modelling (ONDM 2001)*, Vienna, Feb. 2001.

[48] G. EILENBERGER: „Optische Paketnetze – Alles optisch, oder?“ *Beiträge zur 2. ITG Fachtagung Photonische Netze*, Dresden, März 2001, S. 109-114.

[49] A. K. ERLANG: „Lösung einiger Probleme der Wahrscheinlichkeitsrechnungvon Bedeutung für die selbsttätigen Fernsprechämter.“ *Elektrotechnische Zeitschrift (ETZ)*, 1918, S. 189-197.

[50] A. K. ERLANG: “Losning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler.” *Elektroteknikeren*, Vol. 13, 1917.

[51] S. FLOYD, V. JACOBSON: “Random early detection gateways for congestion avoidance.” *IEEE/ACM Transactions on Networking*, Vol. 1, No. 4, Aug. 1993, pp. 397-413.

[52] P. GAMBINI, M. RENAUD, C. GUILLEMOT, F. CALLEGATI, I. ANDONOVIC, B. BOSTICA, D. CHIARONI, G. CORAZZA, S. L. DANIELSEN, P. GRAVEY, P. B. HANSEN, M. HENRY, C. JANZ, A. KLOCH, R. KRAHENBUHL, C. RAFFAELLI, M. SCHILLING, A. TALNEAU,

L. Zucchelli: “Transparent optical packet switching: network architecture and demonstrators in the KEOPS project.” *IEEE Journal on Selected Areas in Communications,* Vol. 16, No. 7, Sep. 1998, pp. 1245-1259.

[53] C. Gauger: “Performance of converter pools for contention resolution in optical burst switching.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002),* Boston, July 2002.

[54] C. Gauger: “Dimensioning of FDL buffers for optical burst switching nodes.” *Proceedings of the 6th IFIP Working Conference on Optical Network Design and Modeling (ONDM 2002),* Torino, Feb. 2002.

[55] C. Gauger, K. Dolzer, J. Späth, S. Bodamer: “Service differentiation in optical burst switching networks.” *Beiträge zur 2. ITG Fachtagung Photonische Netze,* Dresden, March 2001, pp. 124-132.

[56] C. Gauger: *Untersuchung von Reservierungsverfahren für Optical Burst Switching,* Diplomarbeit Nr. 1681, Uni Stuttgart, Sep. 2000.

[57] A. Ge, F. Callegati, L. S. Tamil: “On optical burst switching and self-similar traffic.” *IEEE Communications Letters*, Vol. 4, No. 3, March 2000, pp. 98-100.

[58] N. Ghani: “Lambda-labeling: A framework for IP-over-WDM using MPLS.” *Optical Networks Magazine,* Vol. 1, No. 2, April 2000, pp. 45-58.

[59] N. Ghani: “Integration strategies for IP over WDM.” *Proceedings of the Optical Networks Workshop*, Richardson, TX, Jan. 2000.

[60] N. Ghani, S. Dixit, T.-S. Wang: “On IP-over-WDM integration.” *IEEE Communications Magazine,* Vol. 38, No. 3, March 2000, pp. 72-84.

[61] A. M. glass, D. J. DiGiovanni, T. A. Strasser, A. J. Stentz, R. E. Slusher, A. E. White, A. R. Kortan, B. J. Eggleton: “Advances in fiber optics.” *Bell Labs Technical Journal*, Vol. 5, No. 1, Jan. 2000, pp. 168-187.

[62] P. Green: “Progress in optical networking.” *IEEE Communications Magazine*, Vol. 39, No. 1, Jan. 2001, pp. 54-61.

[63] P. G. Harrison, N. M. Patel: *Performance Modelling of Communication Networks and Computer Architectures*, Addison-Wesley, Workingham, 1993.

[64] G. Hoffmann: “G-WiN – the Gbit/s infrastructure for the German scientific community.” *Computer Networks*, Vol. 34, No. 6, Dec. 2000, pp. 959-964.

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

[66] G. Hu: *Charakterisierung von Internet-Verkehr mittels Wavelets*, Diplomarbeit Nr. 1726, Uni Stuttgart, Mai 2002.

[67] M. N. HUBER: „Ein Netzknotenkonzept für integrierte Durchschalte- und Paketvermittlung“ Dissertation, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1990.

[68] J. HUI: “Resource allocation for broadband networks.” *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, Dec. 1988, pp. 1598-1608.

[69] P. J. HUNT, C. N. LAWS: “Optimization via trunk reservation in single resource loss systems under heavy traffic.” *Annals of Applied Probability*, No. 4, 1997.

[70] D. K. HUNTER, M. C. CHIA, I. ANDONOVIC: “Buffering in Optical Packet Switches.” *IEEE Journal of Lightwave Technology*, Vol. 16, No. 12, Dec. 1998, pp. 2081-2094.

[71] D. K. HUNTER, M. H. M. NIZAM, M. C. CHIA, I. ANDONOVIC, K. M. GUILD, A. TZANAKAKI, M. J. O‘MAHONY, ET AL: “WASPNET: A Wavelength Switched Packet Network.” *IEEE Communications Magazine*, Vol. 37, No. 3, March 1999, pp. 120-129.

[72] ISO: *Information technology – Open systems interconnection – Basic reference model: The basic model*, International Organisation for Standardization (ISO), 1994.

[73] ITU-T: *ITU-T Recommendation G.705 (03/93) – Characteristics of plesiochronous digital hierarchy (PDH) equipment functional block*S, International Telecommunication Union, October 2000.

[74] ITU-T: ITU-T Recommendation G.7713.1/Y.1704 Distributed Call and Connection Management - PNNI Implementation, International Telecommunication Union.

[75] ITU-T: ITU-T Recommendation G.7713.2/Y.1704 Distributed Call and Connection Management - GMPLS RSVP-TE Implementation, International Telecommunication Union.

[76] ITU-T: ITU-T Recommendation G.7713.3/Y.1704 Distributed Call and Connection Management - GMPLS CR-LDP Implementation, International Telecommunication Union.

[77] ITU-T: *ITU-T Recommendation G.803 (03/93) – Architectures of transport networks based on the synchronous digital hierarchy (SDH)*, International Telecommunication Union, March 2000.

[78] ITU-T: *ITU-T Recommendation G.805 (03/00) – General functional architecture for transport networks*, International Telecommunication Union, March 2000.

[79] ITU-T: *ITU-T Recommendation G.8080 (10/01) – Architecture for Automatically Switched Optical Network (ASON)*, International Telecommunication Union, October 2001.

[80] ITU-T: *ITU-T Recommendation G.872 (02/99) – Architecture of optical transport networks*, International Telecommunication Union, February 1999.

[81] ITU-T: *ITU-T Recommendation I.121 (04/91) – Broadband aspects of ISDN*, International Telecommunication Union, April 1991.

[82] ITU-T: *ITU-T Recommendation X.200 (07/94) – Information technology – Open Systems Interconnection – Basic reference model: The basic model*, International Telecommunication Union, July 1994.

[83] S. IYER, A. AWADALLAH, N. MCKEOWN: “Analysis of a packet switch with memories running slower than the line-rate.” *Proceedings of IEEE INFOCOM 2001*, Anchorage, AK, March 2001, pp. 529-537.

[84] R. HÄNDEL, M. HUBER: *Integrated Broadband Networks: An Introduction to ATM-Based Networks*, Addison-Wesley, Wokingham, 1991.

[85] Y. Katsube, K. Nagami, H. Esaki: *Toshiba’s Router Architecture Extensions for ATM*, RFC 2098, IETF, February 1997.

[86] J. S. KAUFMAN: “Blocking in a shared resource environment.” *IEEE Transactions on Communications*, Vol. COM29, No. 10, 1981, pp. 1474-1481.

[87] F. P. KELLY: “Notes on effective bandwidths.” *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, I. B. Ziedins (Eds.), Oxford University Press, Oxford, 1996, pp. 141-168.

[88] L. KLEINROCK: *Queueing systems – Volume I: theory*, John Wiley & Sons, New York, NY, 1975.

[89] H. KRÖNER: *Verkehrssteuerung in ATM-Netzen: Verfahren und verkehrstheoretische Analysen zur Zellpriorisierung und Verbindungsannahme – 62. Bericht über verkehrstheoretische Arbeiten*, Dissertation, Uni Stuttgart, 1995.

[90] P. J. Kühn, Communication Networks I, Lecture at University, Stuttgart, 2002/2003.

[91] P. J. Kühn, Teletraffic Theory and Engineering, Lecture at University, Stuttgart, 2002/2003.

[92] H. T. KUNG, R. MORRIS: “Credit-Based Flow Control for ATM Networks.” *IEEE Network*, Vol. 9, No. 2, March 1995, pp. 40-48.

[93] M. KUTTIG: *Untersuchung von Optical Burst Switching für geslottete WDM-Kanäle*, Diplomarbeit, Uni Stuttgart, 2001.

[94] K. LAEVENS: “Traffic characteristics inside optical burst switched networks.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002)*, Boston, July 2002, pp. 137-148.

[95] S. LIN, N. MCKEOWN: “A simulation study of IP Switching.” *Proceedings of ACM SIGCOMM ‘97*, Cannes, Oct. 1997, pp. 15-24.

[96] M. LORANG: *„Skalierbares Verkehrsmanagement für diensteintegrierende IP-Netze mit virtuellen Verbindungen und verbindungslosen Routen von Datagrammen*, Monographie, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 2002.

[97] F. Masetti, J. Benoit, J. M. Gabriagues, D. Böttle, G. Eilenberger, K. Wünstel, H. Melchior, et al: “High Speed, High Capacity ATM Optical Switches for Future Telecommunication Transport Networks.” *IEEE Journal on Selected Areas in Communications,* Vol. 14, No. 5, June 1996, pp. 979-998.

[98] D. Mitra, M. I. Reiman, J. Wang: “Robust dynamic admission control for unified cell and call QoS in statistical multiplexing.” *IEEE Journal on Selected Areas in Communications,* Vol. 16, No. 5, June 1998, pp. 692-707.

[99] K. Nagami, H. Esaki, Y. Katsube, O. Nakamura: “Flow Aggregated, Traffic Driven Label Mapping in Label-Switching Networks.” *IEEE Journal on Selected Areas in Communications*, Vol. 17, No. 6, June 1999, pp. 1170-1177.

[100] T. D. Neame, M. Zuckerman, R. G. Addie: “Modelling broadband traffic streams.” *Proceedings of IEEE GLOBECOM ‘99,* Rio de Janeiro, Dec. 1999.

[101] P. Newman, G. Mishall, T. L. Lyon: “IP Switching-ATM Under IP.” *IEEE/ACM Transactions on Networking*, Vol. 6, No. 2, April 1998, pp. 117-129.

[102] H. Ohnishi, T. Okada, K. Noguchi: “Flow control schemes and delay/loss tradeoff in ATM networks.” *IEEE Journal on Selected Areas in Communications*, Vol. SAC-6, No. 4, 1988, pp. 1609-1616.

[103] OIF: *User Network Interface (UNI) 1.0 Signaling Specification*, Optical Internetworking Forum, October 2001.

[104] C. Qiao: “Labeled optical burst switching for IP-over-WDM integration.” *IEEE Communications Magazine*, Vol. 38, No. 9, Sep. 2000, pp. 104-114.

[105] C. Qiao, M. Yoo: “Choices, features and issues in optical burst switching.” *Optical Networks Magazine*, Vol. 1, No. 2, April 2000, pp. 37-44.

[106] C. Qiao, M. Yoo: “Optical burst switching (OBS) – a new paradigm for an optical Internet.” *Journal of High Speed Networks*, Vol. 8, No. 1, Jan. 1999, pp. 69-84.

[107] B. Rajagopalan, D. Pendarakis, D. Sha, R. S. Ramamoorthy, K. Bala: “IP over optical networks: architectural aspects.” *IEEE Communications Magazine*, Vol. 38, No. 9, Sep. 2000, pp. 94-102.

[108] K. K. Ramakrishna, R. Jain: “A Binary Feedback Scheme for Congestion Avoidance in Computer Networks.” *ACM Transaction on Computer Systems*, Vol. 8, No. 2, May 1990, pp. 158-181.

[109] R. Ramaswami, K. Sivarajan: *Optical networks: a practical perspective*, Morgan Kaufmann Publishers Inc., San Francisco, March 1998.

[110] R. Ramaswami: “Optical fiber communication: from transmission to networking.” *IEEE Communications Magazine*, Vol. 40, No. 6, May 2002, pp. 138-147.

[111] Y. Rekhter, B. Davie, D. Katz, E. Rosen, G. Swallow: *Cisco Systems’ Tag Switching Architecture Overview*, RFC 2105 , IETF, February 1997.

[112] M. RENAUD, M. BACHMANN, M. ERMAN: “Semiconductor Optical Space Switches.” *IEEE Journal on Selected Topics in Quantum Electronics*, Vol. 2, No. 2, June 1996, pp. 277-288.

[113] M. RENAUD, F. MASETTI, C. GUILLEMOT, B. BOSTICA: “Network and System Concepts for Optical Packet Switching.” *IEEE Communications Magazine*, Vol. 35, No. 4, April 1997, pp. 96-102.

[114] J. ROBERTS, U. MOCCI, J. VIRTAMO: *Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks; Final Report of Action COST 242*, Springer, Berlin, 1996.

[115] J. W. ROBERTS: “Teletraffic models for the TELECOM 1 integrated services network.” *Proceedings of the 10th International Teletraffic Congress (ITC 10)*, Montreal, June 1983.

[116] J. W. ROBERTS: “A Service System with Heterogeneous User Requirements -Application to Multi-Services Telecommunications Systems.” *Performance of Data Communication Systems and their Applications, North Holland*, 1981, pp. 423-431.

[117] E. Rosen, A. Viswanathan, R. Callon: *Multiprotocaol Label Switching Architecture*, RFC 3031, IETF, January 2001.

[118] K. W. ROSS: *Multiservice loss models for broadband telecommunication networks*, Springer, Berlin, 1995.

[119] D. SADOT, E. BOIMOVICH: “Tunable Optical Filters for Dense WDM Networks.” *IEEE Communications Magazine*, Vol. 36, No. 12, Dec. 1998, pp. 50-55.

[120] J. SANITER: "Heinrich-Hertz-Institut Report 2001, Selected Contributions, Photonic Networks, p. 49".

[121] K. SATO: *Advances in Transport Network Technologies: Photonic Networks, ATM, and SDH*, Artech House, Boston, 1996.

[122] M. SCHARF: *Entwurf und Bewertung von Puffermechanismen für Optical Burst Switching*, Student thesis project Nr. 1700, Uni Stuttgart, Juli 2001.

[123] L. B. SOFTMAN, T. S. EL-BAWAB, K. LAEVENS: “Segmentation overhead in optical burst switching.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002)*, Boston, July 2002, pp. 101-108.

[124] J. SPÄTH, S. BODAMER: “Routing of dynamic Poisson and non-Poisson traffic in WDM networks with limited wavelength conversion.” *Proceedings of the 24th European Conference on Optical Communication (ECOC ‘98): Regular and Invited Papers*, Madrid, Sep. 1998, pp. 359-360.

[125] J. SPÄTH: “Dynamic routing and resource allocation in WDM transport networks.” *Computer Networks*, Vol. 32, No. 5, May 2000, pp. 519-538.

[126] J. SPÄTH: „Entwurf und Bewertung von Verfahren zur Verkehrlenkung in WDM-Netzen“ Dissertation, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 2002.

[127] T. E. STERN, K. BALA: *Multiwavelength optical networks: a layered approach,* Addison-Wesley, Reading, MA, 1999.

[128] I. STOICA, S. SHENKER, H. ZHANG: “Core-stateless fair queueing: achieving approximately fair bandwidth allocations in high speed networks.” *Proceedings of ACM SIGCOMM ‘98,* Vancouver, Sep. 1998, pp. 118-130.

[129] S. K. TAN, G. MOHAN, K. C. CHUA: “Algorithms for burst rescheduling in WDM optical burst switched networks.” *Computer Networks,* Vol. 41, No. 1, Jan. 2003, pp. 41-44.

[130] L. TANCEVSKI, S. YEGNANARAYANAN, G. CASTANON, L. TAMIL, F. MASETTI, T. MCDERMOTT: “Optical routing of asynchronous, variable length packets.” *IEEE Journal on Selected Areas in Communications,* Vol. 18, No. 10, Oct. 2000, pp. 2084-2093.

[131] A. S. TANENBAUM: *Computer networks,* Prentice-Hall, Upper Saddle River , 1996.

[132] P. TRAN-GIA, F. HÜBNER: “An analysis of trunk reservation and grade of service balancing mechanisms in multiservice broadband networks.” *IFIP Workshop TC 6 modelling and performance evaluation of ATM technology,* Jan. 1993, pp. 83-97.

[133] R. S. TUCKER, W. D. ZHONG: “Photonic Packet Switching: An Overview.” *IEICE Transactions on Communications,* Vol. 82, No. 2, Feb. 1999, pp. 254-264.

[134] J. S. TURNER: “Terabit burst switching.” *Journal of High Speed Networks*, Vol. 8, No. 1, Jan. 1999, pp. 3-16.

[135] S. VERMA, H. CHASKAR, R. RAVIKANTH: “Optical burst switching: a viable solution for terabit IP backbone.” *IEEE Network,* Vol. 14, No. 6, Nov. 2000, pp. 48-53.

[136] V. M. VOKKARANE, J. P. JUE: “Prioritized routing and burst segmentation for QoS in optical burst switched networks.” *Proceedings of the Optical Fiber Communication Conference (OFC 2002),* Anaheim, USA, March 2002, pp. 221-222.

[137] V. M. VOKKARANE, J. P. JUE: “Burst segmentation: an approach for reducing packet loss in optical burst switched networks.” *Proceedings of the IEEE International Conference on Communications (ICC 2002),* New York City, April 2002, pp. 2673-2677.

[138] V. M. VOKKARANE, K. HARIDOSS, J. P. JUE: “Threshold-based burst assembly policies for QoS support in optical burst-switched networks.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002),* Boston, July 2002, pp. 125-136.

[139] T.-S. WANG: “Architectural evolution and principles of optical terabit packet switches.” *Computer Communications,* Vol. 25, No. 5, April 2002, pp. 557-576.

[140] J. Y. WEI, C.-D. LIU, S.-Y. PARK, K. H. LIU, R. S. RAMAMURTHY, H. KIM, M. W. MAEDA: “Network control and management for the next generation Internet.” *IEICE Transactions on Communications*, Vol. 83-B, No. 10, Oct. 2000, pp. 2191-2209.

[141] J. Y. WEI, J. L. PASTOR, R. S. RAMAMURTHY, Y. TSAI: “Just-in-time optical burst switching for multiwavelength networks.” *Proceedings of the 5th IFIP TC6 International Conference on Broadband Communications (BC ‘99)*, Hong Kong, Nov. 1999, pp. 339-352.

[142] I. M. WHITE, D. WONGLUMSON, K. SHRIKHANDLE, S. M. GEMELOS, M. S. ROGGE, L. G. KAZOVSKY: “The architecture of HORNET: A packet-over-WDM multiple-access optical metropolitan area ring network.” *Computer Networks*, Vol. 32, No. 5, May 2000, pp. 587-598.

[143] C. XIN, C. QIAO: “A comparative study of OBS and OFS.” *Proceedings of the Optical Fiber Communication Conference (OFC 2001)*, Anaheim, CA, March 2001.

[144] Y. XIONG, M. VANDERHOUTE, C. C. CANKAYA: “Control architecture in optical burst-switched WDM networks.” *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 10, Oct. 2000, pp. 1838-1851.

[145] Y. XU, P. N. LAMY, E. L. VARMA, R. NAGARAJAN: “Generalized MPLS-based distributed control architecture for automatically switched tramsport networks.” *Bell Labs Technical Journal*, Jan. 2001, pp. 33-49.

[146] L. XU, H. PERROS, G. ROUSKAS: *Transporting IP packets over light: a survey*, TR-2000-3, North Caronlina State University, 2000.

[147] L. XU, H. G. PERROS, G. ROUSKAS: “Techniques for optical packet switching and optical burst switching.” *IEEE Communications Magazine*, Vol. 39, No. 1, Jan. 2001, pp. 136-142.

[148] C. XIN, Y. YE, W. TI-SHIANG, S. DIXIT, C. QIAO, M. YOO: “On an IP-centric optical control plane.” *IEEE Communications Magazine*, Vol. 39, No. 9, Sep. 2001, pp. 88-93.

[149] S. YAO, B. MUKHERJEE, S. DIXIT: “Advances in photonic packet switching: an overview.” *IEEE Communications Magazine*, Vol. 38, No. 2, Feb. 2000, pp. 84-94.

[150] M. YOO, M. JEONG, C. QIAO: “A high speed protocol for bursty traffic in optical networks.” *Proceedings of the 3rd SPIE Conference on All-Optical Communication Systems*, Dallas, Nov. 1997, pp. 79-90.

[151] M. YOO, C. QIAO, S. DIXIT: “QoS performance in IP over WDM networks.” *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 10, Oct. 2000, pp. 2062-2071.

[152] M. YOO, C. QIAO: “Supporting multiple classes of services in IP over WDM networks.” *Proceedings of IEEE GLOBECOM ‘99*, Rio de Janeiro, Dec. 1999, pp. 1023-1027.

[153] M. YOO, C. QIAO, S. DIXIT: “Optical burst switching for service differentiation in the next-generation optical internet.” *IEEE Communications Magazine*, Vol. 39, No. 2, Feb. 2001, pp. 98-104.

[154] X. YU, Y. CHEN, C. QIAO: “A study of traffic statistics of assembled burst traffic in optical burst switched networks.” *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002)*, Boston, Jul. 2002, pp. 149-159.

[155] H. ZHANG: “Service disciplines for guaranteed performance service in packet-switching networks.” *Proceedings of the IEEE*, Vol. 83, No. 10, Oct. 1995.

Drafts

[156] E. MANNIE, ET. AL. “Generalized Multi-Protocol Label Switching (GMPLS) Architecture“ draft-ietf-ccamp-gmpls-architecture-04.txt, *February 2003, work in Progress.*

Talks

[157] M. VISSERS: *Automatic switched optical network (ASON) and Generalized MPLS (GMPLS)*, Proceedings of the 52 Internet Engineering Task Force, Salt Lake City, December 2001.

WWW-Links

[158] http://www.atmforum.com/

[159] http://www.cisco.com/

[160] http://www.caida.org/

[161] http://www.ikr.uni-stuttgart.de/INDSimLib

[162] Expired draft available at http://www.networking.ibm.com/isr/arisspec.html
N. FELDMAN, A. VISWANATHAN “ARIS Specification“ draft-feldman-aris-spec-00.txt, September 1997, work in Progress.

[163] http://www.oiforum.com/

[164] http://www.omminc.com/technology/whitepapers

Appendix A

OBS in a Networking Scenario

So far, all evaluations presented in this thesis focus on the performance of one node. However, especially as the focus of OBS is on core networks, the performance in a network scenario is of interest. Therefore, a brief discussion on how to extend the results obtained in the single-node case to a scenario of a whole network is presented here.

Section A.1 presents general formulæ and conditions in order to obtain the burst loss probability in a network scenario from burst loss probabilities obtained individually for each (isolated) node. In Section A.2, specific characteristics of the burst loss probability in an OBS network with offset-based QoS mechanism is evaluated which are published first in [42]. Section A.3 focuses on Assured Horizon in a network scenario.

A.1 General Formulæ

One way to calculate the burst loss probability between source and destination without considering the whole network is the approach of a reference path, see also Figure A.1. Here, a reference path through a network from a source node to a destination node traversing n core nodes is depicted. Under the assumption of independence, i. e., traffic seen by a core node is approximately the same for each core node, the well-known stream analysis can be applied. In this analysis, the solution of the loss probability on a path is decomposed in the solution of loss

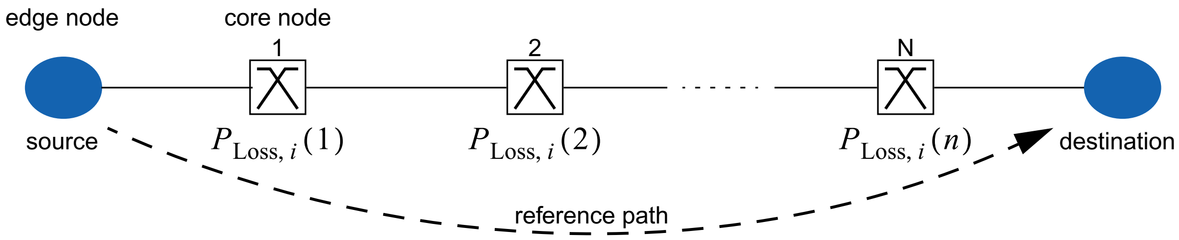


Figure A.1: Network scenario with reference path

probabilities for each node. Afterwards, the end-to-end loss probability is obtained from the loss probabilities at each core node.

The derivation of the burst drop probability on a reference path, $P_{\text{Loss, ref, } i}$ concentrates on the burst drop probability, $P_{\text{Drop, } i}$ of class i which can be obtained individually for each core node v according to the solution of (5.9) in case of offset-based OBS-QoS and according to (7.8) in case of Assured Horizon. Moving along the reference path from source to destination, the offered traffic A_i of class i reduces to $A_i \cdot (1 - P_{\text{Loss, } i}(1))$ after node 1, $A_i \cdot (1 - P_{\text{Loss, } i}(1)) \cdot (1 - P_{\text{Loss, } i}(2))$ after node 2 etc. Hence, after node n, it follows

$$A_i \cdot \prod_{v=1}^{n} (1 - P_{\text{Loss, } i}(v)) = A_i \cdot (1 - P_{\text{Loss, ref, } i}) \tag{A.1}$$

And the end-to-end burst drop probability $P_{\text{Loss, } i}$ for class i on the reference path as

$$P_{\text{Loss, ref, } i} = 1 - \prod_{v=1}^{n} (1 - P_{\text{Loss, } i}(v)) \approx \sum_{v=1}^{n} P_{\text{Loss, } i}(v) \qquad \text{if } P_{\text{Loss, } i}(v) \ll 1 \tag{A.2}$$

Thus, under the assumption that the burst loss probability at a core node is very small (which should be always the case in a core network), the burst loss probability on a reference path is obtained by summation of the loss probabilities of each core nodes.

A.2 Performance of Offset-based OBS-QoS in a Network Scenario

In the following, the burst loss probabilities in a simple network scenario where every destination can be reached with either one or two hops is evaluated. This is a reasonable scenario for a future national core network in a country like Germany [64]. In Section A.2.1, the focus is on effects in a single node in a network scenario while network wide effects are considered in Section A.2.2.

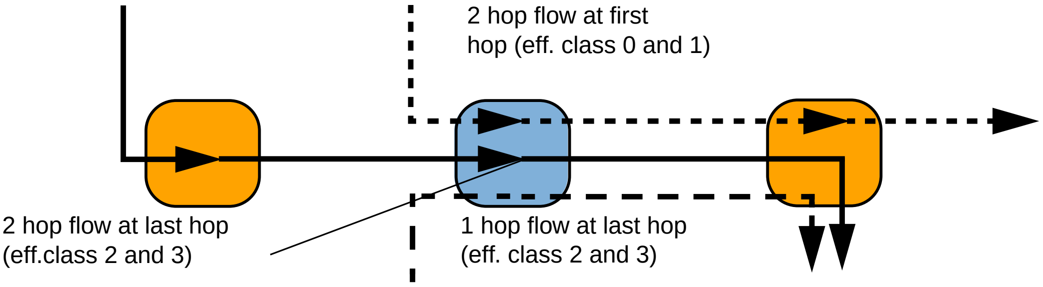


Figure A.2: Traffic flows and effective classes at the evaluated node

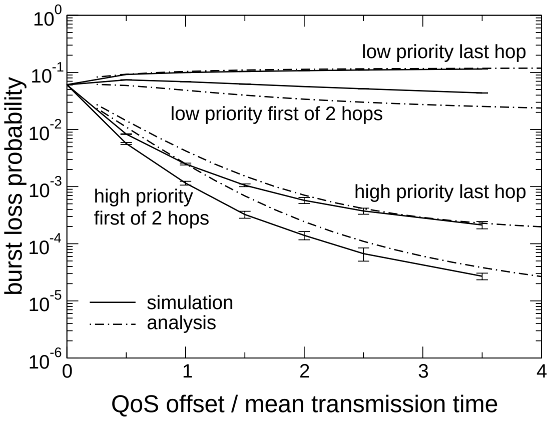


Figure A.3: Comparison of analytical and simulation results for two-class network scenario

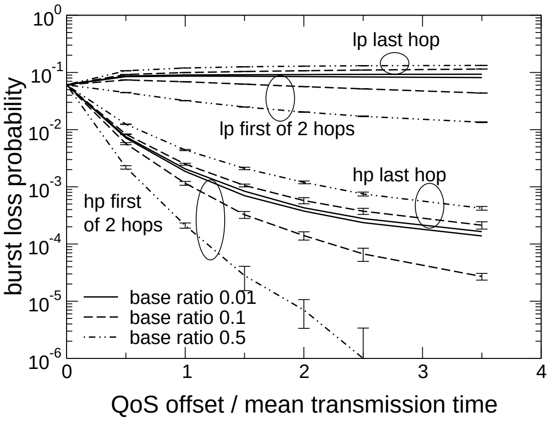


Figure A.4: Burst loss probability at the second node in a two-class network scenario

A.2.1 Multiple Effective Classes due to Basic Offset Adaptation

In a network scenario, bursts with a different number of remaining hops to their destination have different basic offsets as the offsets are decreased in every OBS node traversed. The resulting differentiation based on QoS as well as basic offset can be described by an increased number of *effective* classes. Approximations of burst loss probabilities for the effective classes can be calculated with the multi-class analysis presented in Section 5.2.2. For two service classes in a two hop network, i. e., bursts have either one or two more nodes to traverse (as in Figure A.2), four effective classes have to be considered.

In order to get an idea how basic offset δ_{basic}, QoS offset δ_{QoS}, and mean burst length should be chosen, a basic offset ratio as $r_{\mathrm{b}} = \delta_{\mathrm{basic}} / \delta_{\mathrm{QoS}}$ is introduced. While δ_{basic} is determined by the speed of processing and switching, δ_{QoS} can be chosen rather independently always keeping in mind its influence on loss probability and delay. Original traffic flows and classes are mapped to effective classes according to Table 1. In Figure A.3 and Figure A.4 burst loss probabilities are depicted for different values of r_{b} against δ_{QoS}/h_1. Herefore, the parameters listed in Table 1 and Table 2 with $\delta = \delta_{\mathrm{QoS}} + \delta_{\mathrm{basic}}$.

In Figure A.3, analytical and simulation results are compared for $r_{\mathrm{b}} = 0.1$. It can be seen, that the shapes of respective curves match rather well and that the following principle effects are described by the analysis. From Figure A.4, it can be observed that the curves diverge for both increasing δ_{QoS} and increasing r_{b}. However, an increased r_{b} significantly splits up both, the high priority class and the low priority class, which is very undesirable as bursts which already occupied resources are discriminated. For instance, high priority bursts of the two hop flow at their last hop (effective class 2), which already occupy resources on their first hop link, have a higher loss probability than any high priority burst at its first hop (effective class 0). Thus $r_b < 0.1$ must hold in order to keep the difference in loss probabilities to roughly less than one

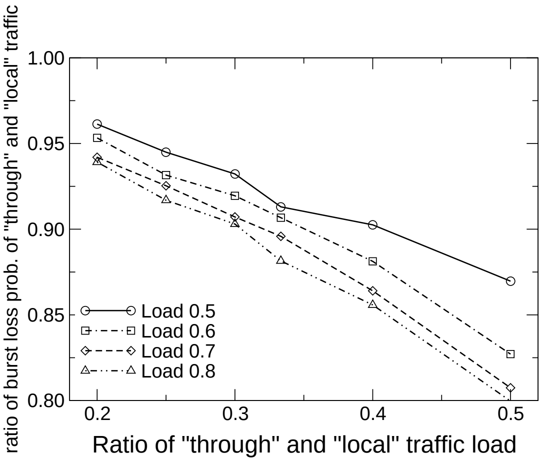


Figure A.5: Burst loss probability in a tandem model with varied „through traffic“

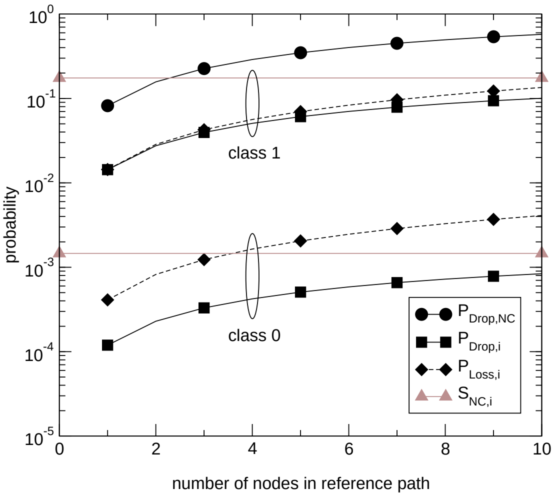


Figure A.6: $P_{\text{Drop, NC}}$, $P_{\text{Drop}, i}$ and $P_{\text{Loss}, i}$ against number of nodes in reference path

order of magnitude for QoS offsets $\delta_{\text{QoS}} < 3 \cdot h_1$ and to allow a reasonable operation in a multi-hop environment. For $r_b > 0.1$ or very large offset values, this spreading in more classes has to be avoided by placing a fiber delay line of length δ_{basic} in front of each JET-OBS node. This fiber delay line compensates processing and switching times and makes a basic offset unnecessary.

Thus, these effects which are just discussed are not caused by the offset-based OBS-QoS mechanisms, but by the compensation of the processing time by a basic offset (and not by an FDL in front of every node). One major results of this evaluation is that this should be avoided in a network scenario and instead, processing of burst header packets should be compensated by an FDL in front of every core node. This has the additional advantage that no source routing is required in order to know the number of hops and determine the respective basic offset.

A.2.2 Generalization of Single-Node Results to Networks

In this section, the assumption is studied that congestion in an OBS-node is independent of the origin of traffic streams as long as they are mixed to a certain degree. If a stream of bursts traverses a sequence of nodes without injection of any other bursts there will be no blocking but in the first node. However, if traffic leaving a node is split up among several nodes and

a.	2 hop traffic flows		1 hop traffic flows	
flow share	1/2		1/2	
QoS class	0	1	0	1
QoS class share	3/10	7/10	3/10	7/10
initial offset	δ	δ_{basic}	δ_{QoS}	0

b.	first of 2 hops		last of 1 or 2 hops	
traffic share	1/3		2/3	
QoS class	0	1	0	1
eff. class	0	2	1	3
eff. class share	3/30	7/30	6/30	14/30
eff. offset	δ	δ_{basic}	δ_{QoS}	0

Tabelle A.1: Flows and classes

Tabelle A.2: Effective classes

input traffic into a node comprises traffic from several preceding nodes, blocking is almost equal for all streams. In Figure A.5 the ratio of traffic is varied which has already undergone a reservation process in a preceding node (through traffic, e. g., solid line at second node in Figure A.2) and traffic which has not (local traffic, e. g., dashed lines at second node in Figure A.2) and plotted the ratio of loss probabilities of through and local traffic. It is shown that for a smaller traffic share of an individual traffic stream, the loss ratio increases and approaches 1.

In a meshed core network, node degrees of at least four (splitting ratio ≤ 0.33 in Figure A.5) are assumed allowing the approximation of independent loss probabilities. Due to this justification the results for the single-node evaluation can be applied also to OBS networks. The end-to-end loss probability can be estimated by the solution given in (A.2).

A.3 Performance of Assured Horizon in a Network Scenario

In Figure A.6, $P_{\text{Drop, NC}}$, $P_{\text{Drop}, i}$ as well as $P_{\text{Loss}, i}$ are depicted against the number of nodes in the reference path. According to the above discussion, it is assumed that all nodes on the reference path carry the same amount of traffic and the ratio between C and NC bursts is the same. Furthermore, all parameters are the same as in Figure 7.23, i. e. $f_0 = 1.6$, $r_{\text{access}} = \infty$, 20 wavelength and $\theta_{\text{NC}} = 0.75$. The dependence of the number of nodes in the reference path follows (A.2) whereas the burst drop probability of a class is obtained by reduction of $P_{\text{Drop, NC}}$ by the share of NC bursts of class i, $S_{\text{NC}, i}$ derived in (7.8). For the burst loss probabilities, the values for the one node case are take from simulations and are extended to multiple nodes in the reference path according to (A.2). Also depicted are $S_{\text{NC}, i}$ which are upper boundaries of $P_{\text{Drop}, i}$ as in a worst case for class i, all NC bursts are dropped in the core network and hence only C bursts reach the destination. Thus, the share of $S_{\text{NC}, i}$ discussed in detail in Section 7.3 yields an upper boundary for large and highly loaded networks. In this case, no multiplex gain between classes is possible.

Also from Figure A.6, it can be seen that $P_{\text{Drop, NC}}$ is increased with increasing number of nodes in the reference path until about 50% of NC bursts are dropped in a scenario of 10 core nodes between ingress and egress nodes. However, it should be emphasized here again, that only NC burst may be dropped and thus not admitted to the wavelengths reservation process whereas C bursts are never dropped. Hence, with given $P_{\text{Drop, NC}}$, the burst drop probability of a class is determined by the share of NC bursts. Comparable to previous discussions, $P_{\text{Drop, 1}}$ and $P_{\text{Loss, 1}}$ only hardly differ whereas the difference between, $P_{\text{Drop, 0}}$ and $P_{\text{Loss, 0}}$ is greater.

Unpleasant effects like discussed in Section A.2 where the number of effective classes is increased and the burst loss probability depends on the current hop are not faced if the basic offset is compensated at every node by, e. g., a FDL like it is proposed for Assured Horizon.

Danksagung

Ich möchte Herrn Prof. Kühn und Herrn Prof. Killat herzlich für die Betreuung dieser Arbeit und die damit verbundenen Diskussionen und Anregungen danken.

Allen meinen Kolleginnen und Kollegen am Institut für Kommunikationsnetze und Rechnersysteme möchte ich für gute und freundschaftliche Zusammenarbeit, sowie unzählige anregende Diskussionen danken.

Schließlich möchte ich mich ganz besonders bei meinen Eltern für ihre Unterstützung bedanken und bei meiner Frau Iris für ihre große Geduld während der Erstellung dieser Arbeit.