Budapest University of Technology and Economics
Department of Telecommunications

Zoltán Zsóka

Performance modelling and analysis of IP over WDM networks

Scientific supervisors:
Dr. László Jereb and Dr. Renato Lo Cigno

The reviews of the dissertation and the report of the thesis discussion are available at the Dean's Office of the Faculty of Electrical Engineering and Informatics of the Budapest University of Technology and Economics.

Zoltán Zsóka, 2005
zsoka@hit.bme.hu

Budapesti Műszaki és Gazdaságtudományi Egyetem
Híradástechnikai Tanszék

Zsóka Zoltán

Teljesítmény-modellezés és elemzés IP over WDM hálózatokban

Konzulensek:
Dr. Jereb László és Dr. Renato Lo Cigno

Az értekezés bírálatai és a védésen készült jegyzőkönyv elérhető a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karának Dékáni Hivatalában.

Zsóka Zoltán, 2005
zsoka@hit.bme.hu

Contents

List of Acronyms

ASON	Automatic Switched Optical Network
CAC	Connection Admission Control
CBR	Constant BitRate traffic
DB	Data-Based traffic model
FSP	Fixed Shortest Path routing
G-OXC	Grooming OXC
HGGM	Homogenous Guaranteed-traffic Grooming Model
IP	Internet Protocol (Network)
ISP	Internet Service Provider
LD	Load Dependent routing
MD	Minimum Distance routing
MLLC	Multifiber Link-Load Correlation
MPLS	Multi-Protocol Label Switching
MSF	Multimetric Sequential Filtering
NGR	Network Graph Reduction
OADM	Optical Add-Drop Multiplexer
OXC	Optical Crossconnect
QoS	Quality of Service
ROADM	Reconfigurable Optical Add-Drop Multiplexer
RWA	Routing and Wavelength Assignment
SLA	Service Level Agreement
SW	Shortest-Widest routing
TB	Time-Based traffic model
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
WDM	Wavelength Division Multiplexing
WDMM	Wavelength Dependent Multifiber Model
WS	Widest-Shortest routing

Chapter 1

Introduction

1.1 Motivation

Networking is one of the most important challenges of the information society evolving in our days. The fast and safe transfer of sometimes huge amount of data is at the heart of current applications. The communication among the members of a population of users is indispensable in the modern working processes based on knowledge shared by users often geographically scattered.

A very important objective in the design and operation of telecommunication networks is the effective use of the available resources. A characterising factor of this efficiency is the *general routing problem* both in virtual circuit switched or in packet switched networks.

A very relevant decision of the network provider is the selection of the most suitable solution and its operation in his system. The support of the performance analysis is essential both in the network design and in monitoring tasks. Its role becomes even more important if any component of the system works in a dynamic manner.

The analysis of the *general routing problem* including the route selection on a given topology and some technology-dependent tasks like wavelength assignment, grooming etc., needs rather complex performance studies. Several previous works on these issues have been presented in recent years, introducing new algorithms, analysing and comparing them with different tools and methods, e.g., [1, 2, 3, 4, 5].

The main objective of the dissertation is to provide new methods and present results regarding the routing-dependent performance capabilities of optical and IP-based networks, that play determining role in telecommunication services of today and of tomorrow.

1.2 Main issues

The general routing includes mechanisms that may work in different network layers and consider different aspects in their decisions. We worked on three specific fields; each of them is strictly joint to the general problem of routing analysis in IP over WDM networks:

- performance analysis of dynamically switched optical networks,
- comparison of existing, and development of new routing algorithms used in IP networks,
- studies on the cooperation of the *optical* and the *electronic* layers, i.e., traffic grooming issues.

In different network scenarios technology and traffic characteristics pose limits to the methods that can be applied.

1.3 Organisation and content of this dissertation

The dissertation consists of five more chapters. Chapter 2 specifies the problems that we studied. Starting from a general description of the network environment it presents the related questions that arise in the performance analysis task. The approaches that can be applied in the studies are briefly presented. Their benefits and drawbacks is discussed from the point of view of the specific scenarios.

The results are organised according to the studied subproblems listed above. For all discussed subproblems, first a description of the network model and a short summary of the related works are given, then the solution of the subproblem is presented, illustrated with numerical results and shortly discussed.

Chapter 3 includes the achievements in the field of the theoretical analysis of automatically switched optical networks with dynamic optical channel requests.

In Chapter 4 the studies on the flow level analysis of elastic IP traffic in different network architectures are presented. New routing algorithms are introduced and analysed with simulation showing their benefits and drawbacks.

Chapter 5 deals with the analysis of dynamic grooming of lower speed traffic on optical channels. A theoretical solution is given for the case of guaranteed traffic and some elementary studies are presented for the case of a more complex model that considers elastic traffic.

Finally, Chapter 6 summarises and concludes the thesis, mentioning some possible directions of further studies.

Chapter 2

Problems and methods

In the previous chapter we presented the motivations that lead us to examine the general routing issues in networks from the performance point of view. Now we define more precisely the problems we wanted to solve. We also list and classify the methods that can be applied in the analysis.

2.1 General model of the network structure

Among the fix, cable based, non-local telecommunication architectures presently used for networking, the one with the brightest perspectives is the TCP/IP based internetworking over static or dynamic wavelength division multiplexed optical networks. We study IP over WDM in the core segment of the network, assuming a WAN or MAN environment.

According to the network model presented in [6] and without specifying the service model we define two layers, that compose the network:

- the *optical layer* provides high capacity connectivity by establishing optical connections that may span large physical distances,
- the *data layer* provides resources and networking functions to the user applications that can use several transport protocols.

The *optical layer* can be interpreted as a network that consists of optical links and switching nodes. The optical links contain several fibers, their number can go up to hundreds in one link. Each fiber can transport data on several wavelengths. A wavelength realises a high capacity optical channel on the link.

The nodes model optical cross-connects, OXCs with optional traffic adding and dropping functions as in OADMs and ROADMs. Some switching devices allow subchannel bundling based on timeslots. If this capability is available in the network we can define *subwave channels* on the links. The capacity of these optical channels is a fraction of the wavelength capacity. The nodes can have different capabilities of wavelength conversion and timeslot reordering. The latter realises the conversion of the *subwave channel*. There are two extreme architectures from this point of view: in the first full conversion is possible at each node and in the other neither wavelength nor *subwave channel* conversion is enabled.

A connection through a contiguous series of optical channels with equivalent capacity is called *lightpath*. It is established according to the RWA algorithm and the switching capabilities in the nodes, and it provides high bandwidth connectivity between its endpoints. There can be established more parallel *lightpaths* between any source and destination node pairs.

The main resources in the *data layer* are the routers in the nodes and the links providing bandwidth capacity for both guaranteed or best effort type user traffic. The most important functions of the routers are routing and queue management, including buffering capabilities. In order to take always the right decisions, these devices observe the traffic on the links and advertise among them the data collected on the network status.

Since the capacity of a *lightpath* is of several Gbps and users require bandwidths that are orders of magnitude smaller, the multiplexing of IP traffic on the optical channels is mandatory. This is the basic motivation of grooming.

In some of the network nodes there are special switching equipments that are necessary to harmonise the tasks of both layers and to perform the data transfer among them. The compound architecture of these nodes consist of one or more cooperating physical devices. User traffic reaches the *optical layer* through *interlayer channels* established via grooming ports. The number of these ports limits the number of parallel *interlayer channels*. In our studies we have assumed infinite number of grooming ports in each node.

The logical connection between the layers of our model is the mapping of the *light-*

paths to the links of the *data layer* topology. Since these links exist only during the lifetime of optical connections, they are virtual links from the network point of view, but the entities of the *data layer* see them as normal links. The virtual links form the virtual topology that can change dynamically if the *optical layer* is dynamically reconfigurable. This indicates, that the topology of the *data layer* differs from the topology of the *optical layer*.

Let us follow the whole route of an IP data unit in the network shown on Figure 2.1. The data is sent from router C_D to router E_D. We assume that the routing of the *data layer* assigned to this communication the route $C_D - A_D - E_D$, i.e., a two-hop path in the virtual topology. In the reality the data will be groomed into *lightpaths* and passes through the *optical layer*. First it takes *interlayer channel* $C_D - C_O$, *lightpath* $C_O - A_O$ and *interlayer channel* $A_O - A_D$ which series realises the virtual link $C_D - A_D$ and then an other *interlayer channel* $A_D - A_O$, *lightpath* $A_O - E_O$ and *interlayer channel* $E_O - E_D$ that realises the virtual link $A_D - E_D$. The real route passed even the optical equipments of node B but this remains hidden from the *data layer*.

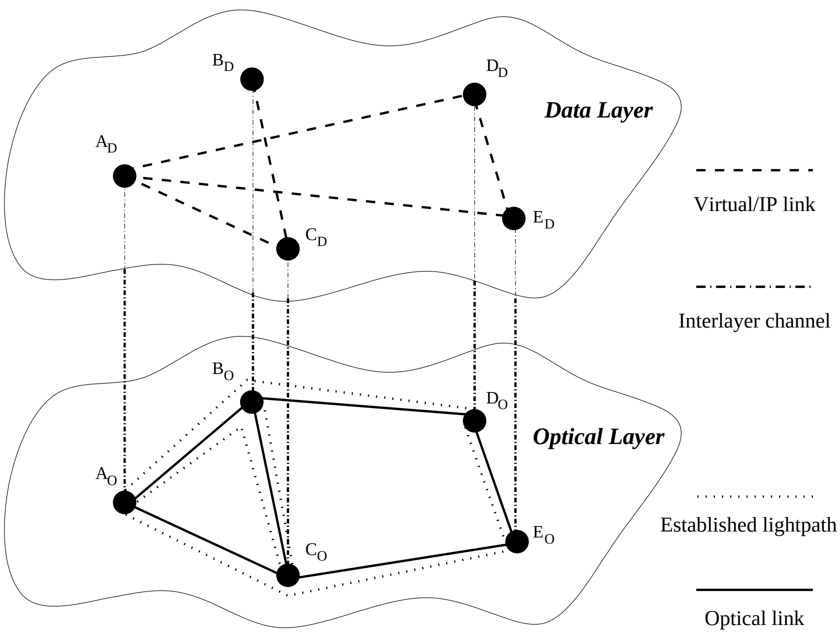


Figure 2.1: General multilayer network model

We discuss later the technologies and traffic types that characterise this multilayered model. Figure 2.1 presents the data planes of the two layers and the connection between them. Control plane details are neglected in this abstraction of IP over WDM networks[1] and we assume a single-domain environment in both layers.

2.2 Decomposition of the analysis task

The main objectives of the studies presented in the dissertation is to find the right performance measures that characterise the modelled networks and to evaluate them considering *general routing problem* solutions. The optimal solution would be the compound analysis of the multilayer network architecture, but beside the very complex modelling task, this meets with other difficulties too. On the one hand the different layers imply different characteristics to observe. The different issues may require also different methods to apply in the analysis.

On the other hand the *general routing problem* may include multiple functions, thus enlarges also the solution space to study. Using a compound model the separation of the effects caused by each cooperating function becomes rather complex. For instance, the effects of the current IP network routing may blend with that of the applied wavelength assignment and lead to a confusion in the analysis.

We focus the analysis on the special issues in a more effective way by decomposing the network and observing the layers separately. The performance of not compound routing functions can be analysed easier this way, since it remains tractable how the algorithm settings affect the behaviour of the network. However, there are questions that refer the issues of the cooperation of the layers and the compound analysis is surely indispensable, e.g. the performance of a given grooming solution.

Figure 2.2 illustrates the decomposition of the IP over WDM architecture. In the compound model data transfer requests arrive from the IP users to the network. The changes of the traffic generated by the service of the user requests implies requests to opening or closing optical connections, i.e., *lightpaths*, according to the decisions of the grooming policy.

[1]GMPLS is a possible choice for signaling.

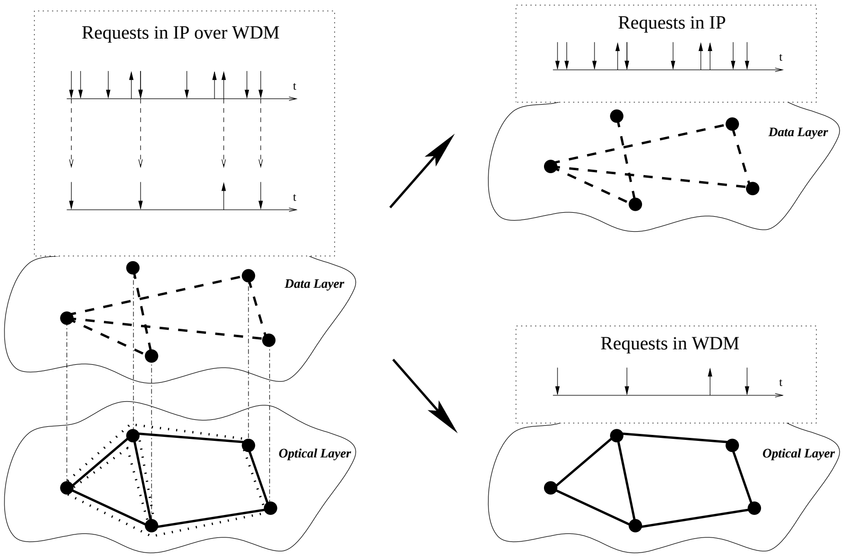


Figure 2.2: Separation of network layers

If we separate the layers, only the relating requests and functions need to be considered in their models. This simplifies the analysis.The decomposition of our general analysis task results in three subproblems:

- analysis of the *optical layer* as a WDM network with fix topology,
- analysis of the *data layer* as an IP network with fix topology,
- analysis of the IPoWDM network with fix topology in the *optical layer* but with variable topology in the *data layer*.

The following three sections deal with these three cases, introducing also the joint issues and presenting more details on the suitable models of the offered traffic.

2.3 Analysis of the *optical layer*

The bottom-up fashion exploring of the IP over WDM network implies first the observation of the optical network layer.

2.3.1 WDM technology

Wavelength division multiplexing network architecture is one of the most effective transport networking solution of our days [7]. In these optical switching networks the capacity granularity is the optical channel that is physically realised as a lightbeam of a given wavelength in a fiber. This means a connection of several Gbps. Since WDM technology allows the use of several beams of different wavelengths in the fibers, the capacity of a fiber is greater than 1 optical channel. Optical links can contain several fibers realising the multifiber environment.

A recent research and development area in the telecommunication field is the investigation of WDM networks that support dynamic reconfiguration, e.g., ASON [8, 9]. These networks treat dynamically arriving optical connection requests. On the one hand, the motivation of dynamics is to provide services with higher utilisation of optical network resources. On the other hand, this solution provides higher performance to the customers, since their resource needs can be satisfied dynamically and only the real usage of resources has to be payed. Optical channel provisioning allows end to end lightpath composition.

As it can be expected, the efficiency of such services depends strongly on the wavelength conversion capability [3], since without converters only the identical wavelengths of links can be connected in a *lightpath*. Not considering the technology differences, having full wavelength conversion in all nodes of the network reduces the problem. The issue is similar to that of a simple circuit switched network with very large capacity resources and high bandwidth requests.

The general case, i.e., the multifiber optical networks with limited wavelength conversion capabilities must be modelled in a rather complex way because of the link and wavelength utilisation dependencies [10].

2.3.2 Optical connection requests

In the case of optical connection requests we cannot make the assumptions typically used for the traffic of connection oriented networks. The provision of connections with the bandwidth of a whole optical channel for the traffic of a single IP user is not realistic. However, users with large traffic, e.g. Internet Service Providers may request large bandwidth connections and they can be considered as users of the *optical layer*. The data that the optical users want to transport comes from the aggregation of the traffic of users in the *data layer*. According to the decomposition of the IP over WDM architecture, also the entities that perform grooming decisions are modelled as optical users.

The characteristics of the requests are not obvious to model since the decision to set up a new optical channel depends strongly on the traffic of the upper layer and on the applied multiplexing-grooming policy. However, we can recognise two important types of traffic in the network that can be modelled in a tractable way.

One is the traffic of a static WDM network in the first phase of the progress towards on demand provisioning: the permanently provided channels can be torn down if no traffic is transported on them. A possible mathematical model of this traffic is the *Binomial* arrival process with a finite number n of generators that are able to generate optical connection requests. Each generator can have zero or one open connection, thus modelling an arrival process of an $M/M/n/n/n$ system when the ON and OFF periods are with exponentially distributed length.

The other traffic is present in scenarios with bursty traffic, streaming from overflows on resources in the *data layer* realised as permanently provided connections in the *optical layer*. Imagine an ISP that has overloaded resources, needs further high capacity links and thus requests an optical connection. This traffic type can be modelled by the *Pascal* or *negative binomial* arrival process that generates requests in a more bursty fashion [11, 12].

Obviously we can not assume a very large population of independently acting optical users and thus the classic *Poisson* arrival model fails in this scenario. However, it can be used as a good reference point, also because of its popularity. It assumes exponential distribution for the interarrival time of the connection requests and the connection duration is a random variable with exponential distribution. In the case of a *Poisson* arrival process the number of connection requests arrived in a given period of Δ has the peakedness $Z = \sigma^2[N]/E[N] = 1$. Z is less than 1 for the *Binomial* model and greater than 1 for *Pascal* process.

These traffic models were compared in [13]. In our studies we have not considered more complex, e.g. PH-based, models for the interarrival process of optical channel requests. The service time was assumed to have exponential distribution that models a memoryless service process.

The connection requests in the *optical layer* are very different from the traditional requests that come from traditional users. In the considered model of IP over WDM network these requests come from the IP control plane when the grooming policy demands to set up a new virtual link. However, in many cases the communication in the *data layer* can be performed also on the current virtual topology. Thus, the refusion of an optical connection request does not imply necessarily the blocking of the traffic of IP users and does not affect critically the data transport service. Considering grooming we can allow for the optical connection requests a higher blocking probability value than that usual in PSTN networks.

2.3.3 Performance of dynamic WDM networks

Though there are some relevant differences, the obvious similarities with classic circuit switched networks suggest us to study similar performance measures as in that research field. Such measures are the utilisation of the total network transfer capacity, that of individual links and blocking probability, i.e., the ratio of refused optical connection requests. These quantities represent the cost-efficiency and availability of services provided by the dynamically switched optical network, thus our studies were focused on the analysis of these measures.

We dealt mainly with the blocking probability and the impact of the following special properties:

- wavelength conversion constraints in nodes,
- links consisting of several fibers,
- special traffic models for the requests.

2.4 Studies of the *data layer*

Assuming the separation of layers, the analysis of the IP data transport issues becomes less complex and more tractable. Let us present now the considered model of the *data layer* and the related problems.

2.4.1 Model of the IP network and its traffic

Our main objective in this area is to study the performance of routing algorithms and for this sake we consider a rather simplified model of the IP network architecture.

The network topology consists of switching nodes and links that connect them. At this point we do not deal with topology changes. The nodes represent entities realising routing functionalities, they are entry and exit points of traffic and connect adjacent links. Links represent transporting elements with arbitrary finite capacity that carry data. Through the constraining effect coming from the finite capacity, they influence strongly the results of decision mechanisms.

We can consider the network management functions realised by a system of distributed decisions using information that are available about the whole network. Information on current utilisation of resources, however, is prone to error measurements, and, most of all, it quickly becomes outdated. Since the *data layer* does not include the model of the control plane of the IP network, we do not consider its technology details and the control traffic.

IP networks are packet based and originally without providing quality of service guaranties. However, architectures that provide QoS for IP users, like MPLS-based services IntServ [14] or DiffServ [15] emerge more and more with the increase of traffic coming from application with real time transport demands. Though classic IP is a strictly datagram packet switched architecture, applying these techniques we can identify data pieces belonging to the same session [16]. We can model the traffic generated by a user session with a *data flow*, allowing a more efficient analysis of routing effects on the quality. A flow can also refer to a set of connections sharing the same source and destination nodes and having the same QoS requirements. The way a flow is identified within the network, or whether a flow of packets is a single logical connection or an aggregation of connections, do not influence basic study results concerning routing performance.

Requests with a data amount to transport arrive to the network from the users according to predefined arrival processes. During the processing of a request an indispensable step is the routing: a path has to be selected that can carry the data of the flow.

According to the flow-based concept routers in our model do not operate on traffic units smaller than data flows, i.e., the implementation of any packet level function is not required. Thus, the model does not consider packet losses and we assume no switching capacity restrictions in routers, the only limiting factor is the capacity of the IP links.

Users may require guarantees on the service quality or – as more typical in the IP networks – they may generate *best effort* traffic. In this latter case, the interaction of flows concurrently transporting on a link leads to elastic rate behaviour, i.e., each flow achieves a transport rate that depends on the actual network state and on the coexisting flows. The properties of the elastic traffic and the mutual effects of the flows on each other were analysed in many works, e.g., in [16, 17, 18, 19].

In addition, the model of the *data layer* considers the concept of *flow starvation*. Caused by the lack of admission control, the number of flows using the network at the same time is virtually unlimited and thus the achieved bandwidth of them tends to zero, i.e., they 'starve' due to the lack of resources. As a result, the application that generates the traffic or the user itself may suspend the flow before it ends with transfer completion.

2.4.2 Performance of QoS routing solutions

Routing has traditionally been an active research field in both circuit- and packet-switched telecommunication networks. Routing strategies that make use of information relative to the network status, as well as information relative to QoS requirements of the traffic being routed, are generally known as *QoS routing* or *constraint–based routing*. Many algorithms of this type were introduced and analysed in previous works, e.g., in [20, 21, 22, 23].

Since these algorithms aim to adapt their choices dynamically to best suit to the current network state they are referred as *dynamic* or *adaptive* algorithms too. It is considered a very promising method for enhancing the performance of integrated services and possibly one of the enabling techniques for the deployment of the Internet, where heterogeneous multimedia traffic flows should coexist. Beside the improvement of the quality of service that the flows receive, the goal of QoS routing solutions is also the improvement of network resource utilisation.

In packet–switched networks QoS routing can be applied only if a sequence of correlated packets that belong to a single connection can be recognised and handled as a flow. Allowing the identification of flows ensures that routing decisions carried out by the router at the ingress of the network are coherently accepted by every other router: thus the path selected by the first router is consistently followed by all packets belonging to the same flow.

Assuming a QoS providing architecture the transport quality of data generated by the users becomes a principal performance measure of the network. QoS routing algorithms were introduced in order to find routes that maximise the average per flow throughput that best effort flows experience. We focused on three related subproblems as follows.

2.4.2.1 Models for elastic traffic

Most of the studies on routing assume virtual circuit switched connections to realise data flows. To have a correct insight into the properties of these algorithms an analysis method is required that considers also the elasticity of the network traffic that consist of requests of transferring finite data. Our results concern the following issues:

- find a suitable method to model the elastic traffic at the *data layer*,
- identify the performance measures that are pertinent to the special behaviour of such flows,
- formalise the routing problem in the IP-QoS environment,
- perform comparative analysis of routing algorithms applying the above model and measures.

2.4.2.2 State information inaccuracy

In the IP networks we have to assume that the distributed values describing the state of resources may be out of date. Indeed, stale load information can even lead to wrong routing decisions that can cause an avalanche effect forcing other route selections to choose the wrong paths [24, 25, 26]. New routing algorithms are often proposed without considering the key issue of robustness to non-optimal working conditions.

If the algorithms are candidates to work in real networks, these issues can not be neglected. Thus, we examined the following problems:

- observe the resistance of QoS routing algorithms to the link state information inaccuracy effects,
- find less sensible solutions.

2.4.2.3 Dependence of network load

A major drawback, however, affects all QoS-based routing algorithms. The cost function at the core of the algorithms tries to find portions of the network where resources are under-utilised and exploits them to the benefit of connections that would otherwise cross a congested portion of the network. Doing so, as shown in [27] for the case of simple alternate routing, when the network load is high the algorithm starts consuming more resources than shortest path routing does. Hence, in case of heavy congestion, QoS-based routing wastes resources and performs poorly compared with shortest path algorithm. The critical drawback of QoS routing in the Internet is clear: whatever is gained at low or medium network loads, it is paid for at high network loads.

A resilient algorithm that allows the migration of a QoS-based routing algorithm to shortest path routing as the network load grows would solve this issue. However, in IP networks the load is typically not known to the routing algorithm, not even in the case of centralised solutions. In addition, in any load dependent algorithm a key issue is that the load level where the migration has to start can depend on the network topology and traffic pattern.

To capture these problems we have achieved novel results on following topics:

- develop an algorithm that can identify the congestion level of the network,
- consider this information in the routing process,
- study the elementary behaviour of the new solution,
- compare the performance of the algorithm with that of previously published ones.

2.5 Analysis of the multilayer network

The layer separated investigation of IP over WDM networks lead to simplified models that are easier to analyse. However, it does not allow to study the issues concerning the interaction of the *data layer* and the *optical layer*.

2.5.1 Modelling IP over WDM

The big gap between optical channel capacity and the achievable data rate of user traffic, that is constrained by access link or internal application limits, implies sharing optical channels among users. This mechanism is called *grooming* and works similar as the multiplexing in circuit switched networks. It is widely applied in statically configured optical networks. In recent years, using the on-demand switched WDM networks some dynamic grooming algorithms were proposed and analysed as for instance in [28, 29].

To consider the dynamic grooming techniques in the model, we need to integrate the *optical layer* and the *data layer*. In this multilayer network we assume a compound architecture of switching elements in nodes. They realise the functions of both layers, but with their cooperation constrained only to the dynamic grooming functions that, however, may include also the routing in one or both layers.

In the compound model the optical connection requests are not directly generated by optical users according to a random arrival process as presented in Section 2.3 but driven by the grooming decisions. Neither the holding time can be determined at the arrival of the request, it will rather depend on the behaviour of the *data layer* traffic and the grooming policy. The other entities of the *optical layer* model do not change.

In the *optical layer* the *lightpaths* will be set up and torn down in a dynamic manner, and they realise virtual links, which transport the data traffic of the *data layer*. The IP network composed of virtual links works accordingly to the concept used in the fix IP network topology case, presented in Section 2.4. Routers work the same way as in the model of a separated *data layer*, but consider always only the actual virtual topology.

We assume that the optical layer does not consider what kind of traffic is transported on the optical channels and the interlayer connections required to resolve data transport between the layers are realised inside the nodes.

2.5.2 Dynamic grooming techniques

As mentioned before, the goal of grooming is to accommodate the user traffic of rather low bandwidth to high-capacity optical channels and to manage the cooperation of the *optical* and *data layers*. Dynamic grooming is strongly connected to three main issues of the *general routing problem*:

1. Routing and wavelength assignment in the *optical layer* which strongly influences the blocking of optical connection requests.

2. Routing in the *data layer* that selects the virtual link for the transmission of user data.

3. Suiting well the virtual topology to the traffic, which enhances the performance of the routing.

The last issue includes decisions whether, when and between which nodes has to be opened a new *lightpath*. If the optical connection request is not blocked these nodes will to be connected by direct, high bandwidth channels, i.e., virtual links.

It is a debated point here whether in their operation the *data layer* functions should consider any information coming from the underlying *optical layer* and vice-versa. Three basic architectures were introduced in [6]: the *peer*, the *augmented* and the *overlay* architecture. They differ by the amount of the information exchange and thus, by the level of cooperation of the layers. The peer model assumes the full cooperation of the layers while augmented allows only the exchange of summarised information between the control planes of the layers. The overlay architecture assumes completely separated routing solutions in the two layers and require only a very simple interface to connect them.

2.5.3 Performance of dynamic grooming

We can analyse an IP over WDM network with the same objectives as in the case of separated layers. In the *optical layer* we are interested in the resource utilisation and the connection request blocking probability while in the *data layer* we study the efficiency of data transport. For the analysis we can apply the same measures and techniques as before, but extended with some new metric that concern the interaction of the layers.

We studied two subproblems related to the analysis that considers both the *optical layer* and the *data layer*.

2.5.3.1 Grooming of guaranteed traffic

As first we analysed scenarios where the connection requests require guaranties on the transport bandwidth. Since the traffic with predictable bandwidth requirements can be carried on constant bitrate channels, a *lightpath* created of suited *subwave channels* can be assigned to the data-flows. To provide the guaranty, a simple admission control is applied in the *data layer* that considers whether in the *optical layer* there are enough available resources to set up a *lightpath* between the source and destination node. If the direct virtual link could be established the request will be accommodated on it. This procedure requires a grooming policy with peer architecture.

This grooming model was introduced in [30] and it leads to very similar problems as defined in Section 2.3. Our study was focused on how to estimate the performance of the network represented by the blocking probability. This characteristic was analysed in the light of the granularity of *subwave channels*, i.e., the difference between the capacity of a wavelength and the bandwidth required by the user traffic.

2.5.3.2 Grooming of elastic traffic

A more complex case has to be studied if we assume more realistic models for IP traffic in the *data layer*. We need to couple the optical network issues with the problems of the elastic nature of data traffic and integrate it with the layer-interaction issues. Such analysis were performed only very recently and few proposals are available [31, 32]. However, the use of more realistic models can lead us to study with more insight the existing solutions and to develop new, more effective ones thanks to the analysis of the results.

Considering the technology, the control plane integration possibilities and the service provision structure of the existing networks, overlay seems to be the most realisable architecture. In our studies the model of IP over WDM network is based on this architecture. Within this area of research, we dealt with the following problems:

- performance analysis and comparison of the elastic traffic models and different grooming algorithms using typical measures related to the *optical layer* and *data layer*,
- definition and analysis of special measures that characterise the interaction of the layers.

2.6 Approaches

The flavours of the different problems presented above can imply the use of different approaches in the analysis. Let us give a short summary of the available methods and compare their advantages and drawbacks also from the specific point of view: how they can be applied to solve our problems.

2.6.1 Performance measures in networks

To observe networking solutions in the most practical and veritable way, one should measure real networks. After assembling a test network or realising measurement points in a working one, we can collect different statistics and process them. However, there are several issues that obstruct us to apply such methods in the studies on the *general routing problem*:

- In the phase of the design or in the comparison of the possible solutions, it is far too expensive to build up a network with full functionalities as measurement environment.
- The new networking solutions have to be implemented in each network equipment.
- Special performance measures have to be defined that are not always directly available in the statistics set of the equipment and often they cannot be derived from the available statistics.

- The real-time observation takes long time.
- The identification of the stationary phase is not obvious.

2.6.2 Performance modelling with theoretical methods

In the last decades many works were performed and published on the theoretical approach in the analysis of networks supporting dynamic demands. The research started with the study of the simplest queuing systems and arrived to results on the analysis of very complex cases. Here we just mention some important works as [33, 34, 35].

Applying queuing models in the field of telecommunication networks is almost obvious. A large population of independent users is assumed. The users want to transfer data using network resources and the population offers dynamically changing traffic. Stochastic theory allows the evaluation of the dynamic situations and the calculation of statistical behaviour of performance values.

The results in this domain are mostly based on the following modelling techniques derived from stochastic theory:

- Markov chains,
- renewal processes,
- fluid models,
- combinatorics,
- Petri nets,
- queuing networks.

The list is not exhaustive and also other techniques were exploited in the last decades. The methods can be applied in the analysis of different issues of networking and in the modelling of different network component[2] characteristics.

[2] switching and transferring elements, schemes, algorithms, policies, etc.

The performance analysis of optical, virtual circuit switched or packet switched telecommunication networks that considers the routing implies problems that can be formalised within the framework of these disciplines. As generally in modelling, the most important and most complex problem is to find a suitable interpretation of network components and to identify the outputs we want to analyse.

Theoretical formulae provide us exact results whose calculation can be performed generally – but not always – quickly. However, the models are not always robust and work only under strictly defined conditions. Unfortunately, in the performance analysis we have to consider this drawback. It can then cause the need of significant changes of the model for even very small changes of an element, e.g., the routing algorithm or the traffic characteristics.

In our work we developed theoretical models using Markov chains and combinatoric approaches.

2.6.3 Simulation tools

In our studies that concern different networking solutions and algorithms, the robustness of development is a very important point. In simulators one can realise the new routing algorithms and other schemes easily. The functionality and specific effects of these solutions can be then observed and controlled directly. Simulation tools provide a rather easy way to analyse the network performance, and they can be used to verify the results obtained with theoretical models.

On the other hand, simulations demand long time to run to collect enough samples. Mostly, a huge amount of events have to be simulated to obtain confident and accurate values as results. Another issue of this analysis method is related to rare events. Some decisions of the general routing scheme determine the network behaviour for long time and in these cases more simulations have to be started with different seeds to get meaningful results considering each important network traffic situations.

Similarly to the case of theoretical models, the efficiency and accuracy of the simulation results depend on the interpretation of the network components. The selection and identification of the statistical variables that we use for modelling the performance and their representation in the model are critical tasks.

Several commercial and free network simulator tools can be found on the Internet that are developed for performance studies. They model the network entities and functions with different modelling depth. The more general tools capture the behaviour of nearly all possible network components, protocols and technologies. Thus they can simulate fine tuned, strictly specified scenarios providing detailed and accurate results. However, using these tools the researcher may be lost in the details and it is difficult to make general conclusions. When comparative studies has to be performed, such as those in our tasks, rather the use of simple but transparent simulators is suggested.

For the simulation results presented in this work we applied the tool Ancles and its special versions [31, 36, 37, 38, 39]. Originally it was developed at the *Politecnico di Torino* for ATM and IP network simulation and we extended it to cover the analysed issues. The newer versions of Ancles – ASONcles and Gancles – support both the *optical layer* and *data layer*, dealing even with their interaction, i.e., with dynamic grooming. Our choice to use this simulator was strongly influenced by the cooperation of the *Telecommunication Network Group* at the Politecnico, the *Department of Information and Communication Technology (DIT)* at the University of Trento and the *Networking Research Group* at our department. This cooperation resulted in many common publications as for instance [23, 32, 40, 41].

2.6.4 Application of the methods

Regarding the efficiency of the theoretical analysis and the simulation of telecommunication networks we can come to the general experience that the former can be faster and giving deeper insight in the functionality of the system. However, as we mentioned before, the simulation provides a framework where it is easier to extend the models and implement the network functions, while with real measurements very accurate results can be obtained. We used the following approaches in the analysis of the certain problems:

- Studies at the *optical layer*: we created a new theoretical model and we used simulation only for its validation.

- Studies at the *data layer*: we inserted new traffic models and networking functions in a simulation tool and simulated the IP network.

- Studies of the IP over WDM network: a theoretical model is developed for the case when guaranteed traffic is assumed in the *data layer*, while in the case of elastic traffic we used simulation.

Due to the difficulties listed in Section 2.6.1 we did not deal with real network measurements during our work. In the next sections we summarise the main reasons that we considered at the choice of the methods.

2.6.4.1 General constraints of theoretical models

The theoretical analysis can be performed effectively only when the model supports fast calculations. To develop such a model sometimes significant simplifications and approximations are required. However, even not very accurate results can be accepted, when the motivating issue of the analysis allows it.

Another general problem is the scalability of the theoretical methods. The computation time can grow very fast with the number of network components when the evaluation method is not well structured, e.g., it contains recursive calculation.

The accuracy depends strongly on how the used model fits well the network architecture under the scope. It often happens that a small change in any component of the modelled system implies the need of developing of a brand new model. As a simple example we just have to think on the differences that we can have analysing the service of different arrival processes or that of different routing algorithms.

Obviously, a rather significant difference can be observed when modelling the different layers of the IP over WDM network since they differ even in their basic concepts. Undoubtedly, the cooperation and interaction of the layers is a much more complex problem.

Using random variables in traffic modelling implies the complexity problem of having a large or even infinite number of network states to be evaluated. In many cases the theoretical analysis dissolves this problem and reduces the evaluation complexity with the help of simple but well suited models. However, in some scenarios simple models cannot catch complex network functions. This can lead to situations where multiplied evaluation has to be performed according to different values of parameters that describe the state of the network. For example, let us consider that the arrival order of connection requests affects significantly the current and mean performance measures. It is not enough to evaluate only the case of one – randomly chosen – arrival order, but at least the statistically most important or most probable cases have to be analysed one by one.

Calculation with theoretical models fails often due to problems with their implementation. Results of intermediate steps of calculation can be of different magnitude and thus their combination in a further step can lead to very inaccurate operations.

To capture the above problems often the combination of simulation and theory can be useful as done in [42].

2.6.4.2 Modelling the traffic

In the *optical layer* we assume connection based traffic and the obvious similarities imply to model this layer with similar methods as in the case of analysis of PSTN networks. The link behaviour can be considered as a Markov process on the number of the idle channels on it and this basic idea can be extended to networks in several ways. In our case, i.e., having the *optical layer* and links with optical channels, we need to consider the main constraints coming from the technology, e.g. wavelength continuity, and the specific, i.e., non-Poisson, characteristics of traffic. Though these difficulties, we can find theoretical solutions for some problems relating the *optical layer*.

The packet-based traffic in IP networks can be modelled with the help of queueing networks [34]. The main problem with these solutions is that they study the performance from the network element point of view, instead of that of the network user. By this we mean, that with these methods one can evaluate the characteristics of the switching node or link, e.g. packet loss, waiting time and load. Although these values are important in the analysis of the network, they are not representative of the network performance provided to the end users. Obviously, this latter is the more important point in the studies regarding QoS provisioning.

The characteristics of the traffic with a particular entry and exit point, and touching maybe more switching and transporting network components, can not be always derived easily from the characteristics of the involved components. Their relation can not be described using simple additive methods.

On the other hand, these models do not consider the deterministic routing algorithms of IP. In the classic queueing network analysis methods the incoming traffic is mixed and the next node of a packet is chosen stochastically. This way we neglect the correlation between the traffic coming from one input line and going to one predefined output line of the node. It results in inaccuracies even if at the output lines we use the probabilities suited to the pattern of the offered traffic load.

To resolve these problems analysis at the flow level might be used. The basic approach of such studies uses the principles of fluid models and ideal resource sharing in the network nodes. In the required model the non-persistent flow requests arrive in the network according to a given arrival process and a suitable route is assigned to them. The instantaneous bandwidth of a flow depends on the network state and the number of co-existing flows, due to the sharing of common resources. Depending on the assumptions about the queueing policy of the routers the analysis of the elastic traffic can be based either on the max-min-fair-sharing [35] or on the proportional sharing [43, 44] approach. Since we studied scenarios where the flows were with rather similar parameters, we used the first approach.

The problem is a simplified, formal approach of the problem that regards the throughput calculation of IP traffic with implicit or explicit feedback, e.g. with protocols TCP or UDP. Though the high importance of this issue in the analysis and preparation of planning, only a few, more or less suitable models were presented on it, e.g. [19, 45]. The accuracy and scalability of these models are mostly poor, that proves the complexity of the problem, and thus the constraints of the theoretical modelling of such traffic.

Analysing the compound network model, i.e., the IP over WDM architecture presented in Section 2.1 we have to follow two different ways depending on the traffic of the upper layer. If we assume elastic traffic in IP layer, then still hold the constraints and difficulties with the theoretical approach that were presented in the previous paragraphs. If the *data layer* is assumed to transport guaranteed traffic and to use admission control at the network entry points the analysis task of the two layers can be reduced. In this case a connection based theoretical model is feasible.

2.6.4.3 Analysis of routing and grooming solutions

The issue of performance analysis with theoretical methods becomes much more complex if dynamic decisions are allowed in the network functions. The methods trace the behaviour of the network with stochastic approximations, i.e., modelling the events in a probabilistic manner. In dynamic solutions the network state is fed back and that can lead to an explosion of the state-space in the analysis.

Many routing algorithms were proposed in both layers with the aim to utilise more effectively the low loaded resources of the network. They use decisions that consider the instantaneous state of the network and thus provide adaptive routing methods. Although this dynamics is not easy to introduce in the analysis, some authors proposed theoretic models of adaptive routing algorithms, as for example in [46]. These works mostly assume the choice among the available path alternatives to be a random variable according to the stochastic model of the network. This way even impossible events can have non-zero probability that leads in general to wide inaccuracy of such models.

We find the same problems in the case of dynamic grooming solutions. On the one hand, the decisions on the changes of the *data layer* topology should be modelled. This leads to a double feedback problem since the instantaneous state of both layers has to be considered to decide if a new virtual link is needed and if it is realisable. The required information can be obtained from the *data layer* and *optical layer* respectively.

On the other hand, to analyse accurately the performance of the *data layer* we should evaluate this layer of the network for each possible topology configuration. It is easy to see that the number of the possible cases is huge even for a small network.

A very important decision is how to choose the network topology for the studies. Specific characteristics of the analysed solutions can be emphasised better when we use an appropriate topology. The guidelines for the choice depends on the studied problem and thus different topologies may be required in the different tasks. Beside the topology characteristics also the size of the network plays a role at this point. On the one hand, evaluating scenarios with small networks can help to understand easier the behaviour of a solution. On the other hand, large networks are rather realistic, though their analysis is sometimes very complex and time consuming. Thus, in our studies we used different sizes and types of network topologies.

Chapter 3

Studies at the *optical layer*

3.1 Introduction

Recently the performance analysis of routing and wavelength assignment in dynamic WDM, i.e., automatic switched optical networks received a lot of interest. Many algorithms on this topic were presented and analysed, summaries on this research area can be found for instance in [46, 47, 48]. Some solutions consider also the use of protection schemes [8, 9], in order to provide reliable high bandwidth connections similarly to usual services of static WDM architectures.

In Section 2.3 we defined the main issues related to the analysis of the *optical layer* with dynamic reconfiguration capabilities and support of dynamically arriving connection requests. In the modelling and analysis tasks the constraints originating in the optical technology and specific arrival processes have to be considered.

Several previous work were presented in recent years on parts of the theoretical models of dynamic WDM networks. However, most of these models can be applied only among very strict conditions. In nearly all related models we find the constraint, that the requests arrive according to a Poisson process and with exponential holding times. As it was discussed in Section 2.3.2 this is not a very realistic scenario.

Clear models were presented in [12, 46, 49, 50, 51] and [52], but the authors ignore the dependency of link loads. Paper [1] introduces a complex derivation of performance bounds that considers any routing and wavelength assignment algorithm based on the solution of an ILP formalised problem.

A combinatoric approach is the basis of modelling the wavelength sets and the *lightpath* setup in [12, 46, 50, 51, 52, 53] and [54]. This approach is applicable to the examination of *random* wavelength assignment. [55] gives a solution based on the analysis of Markov chains without considering the load dependence of the wavelengths on links.

A very different method, the overflow analysis is used in [3] and [10], which present models for *first-fit* and other assignment of wavelengths. The papers [3, 46, 51, 56] consider also adaptive routing algorithms beside the *shortest path* algorithm.

The multifiber environment is introduced only in few references. A very clear model is that of [52] and its generalised version in [57], which considers different switching trunk sizes on the links. A very important drawback of these models is the assumption that optical links are with the same number of fibers and there is a uniform capacity on all of the network links. On the other hand, they have large computation time caused by recursive steps in the calculation.

The authors in [58] present a model that solves the problem effectively and provides accurate results due to consider most of the issues of the *optical layer*. This model proceeds only in the analysis of single fiber networks with Poisson traffic, but its extension in [59] leads to a rather general method.

3.2 Theoretical analysis of dynamic WDM networks

Approaching the study of dynamic WDM networks, we devised **WDMM**: *Wavelength Dependent Multifiber Model*. It is a theoretical model that considers also the possible constraints derived from the optical technology, e.g. need of wavelength continuity, and its special traffic characteristics in order to analyse efficiently such networks. Let us first present the basic assumptions that accord to our general network vision.

We assume that the all-optical communication network consists of a set of switching nodes modelling OXCs or OADMs and optical links connecting them. A continuous series of adjoint links is called path or route. According to the multifiber option a link consists of one or more fibers. An identical wavelength set is assumed on each fiber. A wavelength realises one optical channel with a bandwidth that depends on the technology.

A fiber in the model represents a pair of unidirectional fibers in the modelled optical network. Thus, also the optical channels are bidirectional and adopted to connection requests with bidirectional communication needs. Analogously to the PSTN model, where telephone calls occupy resources for both forth and reverse direction traffic, a set of bidirectional optical channels will be assigned to each connection. However, this very natural assumption does not constrain the interest matrix to be symmetrical.

We assume that the switching nodes are incapable to convert wavelengths, resulting in a wavelength continuity constraint for *lightpaths*. A *lightpath* connects two nodes with a high bandwidth connection. It is realised by connecting optical channels on the links of a route between these nodes.

Let us define the main terms of the model. A wavelength is free on a link when there is at least one fiber of the link where the wavelength is currently not occupied. This term can be extended for routes, a wavelength is free on a route if it is free on each link of the route.

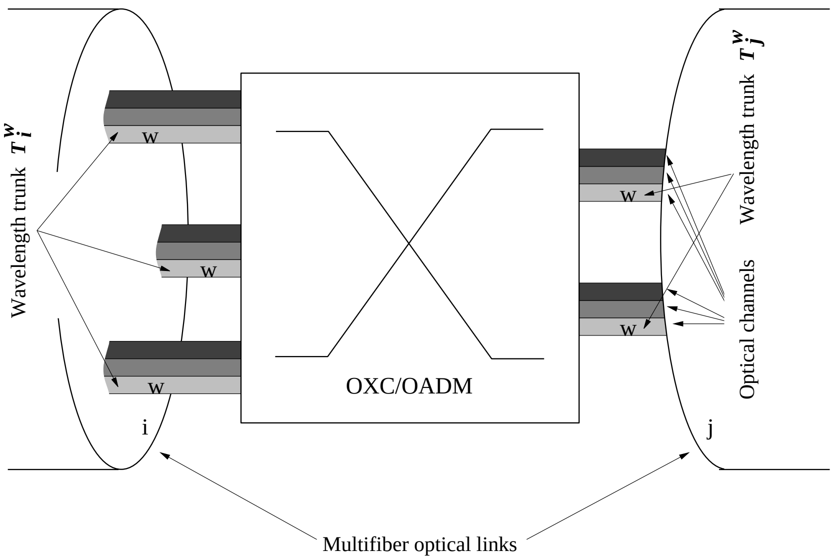


Figure 3.1: Wavelength trunks on joint links

We illustrate the concept of *wavelength trunks* on Figure 3.1. The wavelength trunk T_j^w is the set of optical channels in link j that are assigned to wavelength w. The possibly different number of fibers on the links imply that T_i^w and T_j^w on different links can have different sizes. On the other hand, since fibers are identical, the trunk sizes T_j^w and T_j^v are equal for each w and v. The channels in trunks with the same wavelength can be connected in the nodes without restrictions since no wavelength conversion is required among them. A trunk is free on a link when at least one of its channels is free, i.e., the corresponding wavelength is free on the link according to the above defined term.

Connection requests arrive in dynamic fashion according to a stochastic process and their duration is stochastically distributed. The routing and wavelength assignment task is performed every time a request arrives. The pairs of nodes can be referred by their assigned route and vice versa, since a single, predefined path is selected for each connection between a given source-destination pair. Thus our model considers fix routes that can be determined with any algorithm. This restriction originates in the general constraints of theoretical models mentioned in Section 2.6.4.

If there are one or more free wavelengths on the selected route, a *lightpath* can be established for the communication. The selection of the wavelength is realised by a weighted random choice. To achieve a uniform distribution of their usage, the current weight for wavelength w is set to the current minimum of the available optical channels of w on the links of the route. This is equal to the number of *lightpaths* that could be set up concurrently using w at the instant of the arrival.

The connection occupies one optical channel of the chosen wavelength w on each link of the route and it is called a connection of *colour* w. If there are no free wavelengths on the route assigned to the source-destination pair, the connection request will be refused and a blocking event has to be registered. Blocked requests are not repeated.

We used the following notation:

J : number of network links,

R : route, that consists of $|R|$ links,

M_j : number of fibers on link j,

M : maximum value of M_j,

C : number of different wavelengths on one fiber,

C_j : capacity of link j, it can be calculated as $M_j \cdot C$ and it is given in optical channel units,

H : maximum number of hops in the predefined routes of the network.

According to the definitions the number of trunks is C on every link and the size of T_j^w is M_j.

3.2.1 Description of the computation model

The analysis is performed by an iterative algorithm equipped with a feedback on the offered load level. The main steps are as follows:

1. Initialise the input values,
2. compute link loads considering the blocking effects originating from other links,
3. calculate the probability that a set of wavelengths is free on a single link,
4. extend the analysis to whole routes using an iterative method considering the mutual impact of adjacent links,
5. calculate total network blocking probability considering the offered traffic pattern,
6. if the required precision is reached then stop, else start again from step 2.

3.2.1.1 Traffic model and single link analysis

We model the WDM network considering all types of traffic that can be described by a memoryless arrival process with a possibly varying intensity and exponential holding times. The *Binomial*, *Pascal* and *Poisson* arrival models that were listed in Section 2.3.2 are of this type. From these traffic models we can derive the process that describes the number of occupied optical channels on a link as a birth-death process.

Let $\alpha_j(m)$ be the intensity of connection request arrival on link j, given exactly m free optical channels on it. At the calculation of this intensity we have to consider the characteristics of all the traffic that meets link j. According to the arrival model $\alpha_j(m)$

can depend on the current state of link j and it is also affected by the traffic arriving to other network links. We do not loose generality assuming normalised intensities by setting the mean connection holding time to 1.

From the steady state analysis of the birth-death process we can easily get the probability of being exactly m free channels on link j:

$$q_j(m) = \frac{C_j(C_j-1)\cdots(C_j-m+1)}{\alpha_j(1)\alpha_j(2)\cdots\alpha_j(m)} q_j(0) \tag{3.1}$$

where $q_j(0)$ has to be calculated via normalisation, i.e., according to the fact, that $q_j(m)$ is a distribution on m. All the following calculations use this distribution regardless of how the $q_j(m)$ values were obtained. Thus, **WDMM** works in the case of any traffic model, for which these values can be calculated.

A wavelength-set is available by definition if each wavelength in the set is free. The probability that a set I with cardinality i is available on link j can be derived according to the application of the random wavelength assignment algorithm:

$$\beta_{I,j}^{mul} = \sum_{m=i}^{C_j} q_j(m) \frac{\sum_{k=0}^{\min\left(i, \left\lfloor \frac{C_j-m}{M_j} \right\rfloor\right)} (-1)^k \binom{i}{k} \binom{C_j-M_jk}{m}}{\binom{C_j}{m}} \tag{3.2}$$

The sum in the nominator of Equation 3.2 describes the number of cases when the wavelengths of set I are free, given that there are m free channels on link j. We can get it as the number of all cases $\binom{C_j}{m}$ less the number of cases when at least one wavelength of I is not free. The latter is calculated using the inclusion-exclusion rule for the members of set I.

3.2.1.2 Analysis of multihop routes

Let us observe now the mutual effect of links that can be derived from the traffic correlation and from the lack of wavelength conversion assumed in our all optical network model. As it is mentioned in [58] too, this effect is not negligible if there are several routes that contain some common multihop sections[1].

[1]Sparse networks with few links have this characteristic, e.g. rings.

To simplify the problem we introduce some assumptions applied to each route R:

1. on adjacent links j and $j+1$ of route R we consider only the dependencies of the trunks with the same wavelength, i.e., T_j^w and T_{j+1}^w,

2. the dependencies of trunks with the same wavelength is considered only on the adjacent links of route R,

3. we do not consider the dependencies between the traffic on link j and any other traffic relation in the network that uses a route R not containing j.

Now we can estimate the probability that a set I is available on the two-hop route consisting of links A and B:

$$g_I^{A,B} \approx \beta_{I,B}^{mul} \prod_{k=1}^{i} \frac{\beta_{I_k,A}^{mul} - \beta_{I_{k-1},A}^{mul} \gamma_{k,AB}^0}{\beta_{I_{k-1},A}^{mul} \left(1 - \gamma_{k,AB}^0 - \gamma_{k,AB}^1\right)} \tag{3.3}$$

where $\gamma_{k,AB}^0$ is the probability that wavelength k is free on link A, but not free on link B, while the $\gamma_{k,AB}^1$ is the probability that k is not free on both link A and B. Wavelength-set I_l contains the first l members of set I and its cardinality is equal to l. After l steps the product in Equation 3.3 results in the conditional probability, that set I_l is available on link A, given that it is available on link B. Thus, after i steps we get the conditional probability for set I.

Let us consider the mean intensity of traffic on link o as λ_o and that of continuing traffic on two adjacent links p, q as $\lambda_{p,q}$. A continuing connection means, that the assigned route contains both the p and q links. Using a combinatoric approach we can derive the following probability values that hold for each wavelength w:

P_l^j is the distribution of the number of connections of colour w on link j:

$$P_l^j(k) = \sum_{m=M_j-k}^{C_j-k} q_j(m) \frac{\binom{M_j}{k}\binom{C_j-M_j}{m-(M_j-k)}}{\binom{C_j}{m}} \tag{3.4}$$

P_c^j is the distribution of the number of continuing and non-continuing connections of colour w on the adjacent links j and $j+1$:

$$P_c^j(l,k) = P_l^j(k+l)\left(\frac{\lambda_{j,j+1}}{\lambda_j}\right)^l \left(1-\frac{\lambda_{j,j+1}}{\lambda_j}\right)^k \binom{k+l}{l} \tag{3.5}$$

P_n^j is the conditional distribution of the number of continuing connections of colour w on the adjacent links $j-1$ and j, given the number of non-continuing connections:

$$P_n^j(k|l) = \frac{P_l^j(k+l)\left(\frac{\lambda_{j-1,j}}{\lambda_j}\right)^l \left(1-\frac{\lambda_{j-1,j}}{\lambda_j}\right)^k \binom{k+l}{l}}{\sum_{n=0}^{M_j-l} P_l^j(n+l)\left(\frac{\lambda_{j-1,j}}{\lambda_j}\right)^l \left(1-\frac{\lambda_{j-1,j}}{\lambda_j}\right)^n \binom{n+l}{l}} \tag{3.6}$$

Note that these values are independent from w due to the random assignment of wavelengths. To determine the values $\gamma^0_{k,AB}$ and $\gamma^1_{k,AB}$ we can use these distributions:

$$\gamma^0_{k,AB} = \sum_{i=0}^{M_A-1} \sum_{l=0}^{\min(M_B,i)} P_n^{(B)}(M_B-l|l) P_c^{(A)}(l,i-l) \tag{3.7}$$

$$\gamma^1_{k,AB} = \sum_{l=0}^{\min(M_B,M_A)} P_n^{(B)}(M_B-l|l) P_c^{(A)}(l,M_A-l) \tag{3.8}$$

Now we introduce the conditional probability that the set I of wavelengths on the j^{th} link of route R is available, given that it is available on the subsequent link $j+1$:

$$\beta'_{I,j} = \frac{g_I^{j,j+1}}{\beta^{mul}_{I,j+1}} \tag{3.9}$$

Starting from the above values we can calculate the probability that the wavelength-set I is available on the whole route route R of $|R|$ hops:

$$g_I^R = \beta^{mul}_{I,H} \prod_{j=1}^{|R|-1} \beta'_{I,j} \tag{3.10}$$

The blocking probability on the route R, i.e., the blocking of the traffic that uses this route, can be computed easily, using the inclusion-exclusion rule of sets:

$$B_R = 1 - \left(\sum_{k=1}^{C} (-1)^{k-1} \binom{C}{k} g_{I_k}^R \right) \tag{3.11}$$

To get the total blocking probability of the network, we only need to take the weighted sum of the B_R values. The weights are the normalised offered load values between each pair of nodes, i.e., the values in the interest-matrix.

3.2.1.3 Feedback on the arrival characteristics

Let us show how load correlation effects can be considered in **WDMM**. Blocking on route R affects the offered load on each of its links. Thus, we can derive the intensity of connection request arrival on link j, given m free optical channels on it:

$$\alpha_j(m) = \sum_{R:j\in R} \left(\lambda_R(j,m) \sum_{k=1}^{\min(m,C)} (-k)^{k-1} \binom{C}{k} g_{I_k}^{R,j}(m)) \right) \tag{3.12}$$

where $\lambda_R(j,m)$ is the intensity of the traffic offered to route R when there are m free channels on link j. For the *Poisson* traffic model $\lambda_R(j,m)$ does not depend on j and m. For the *Binomial* and *Pascal* traffic types a more complex calculation is required.

The conditional probability $g_I^{R,j}(m)$ means that the wavelength-set I is available on route R, given that there are m free channels on link j. Using the same idea as in Equation 3.2 this probability can be calculated as:

$$g_I^{R,j}(m) = \frac{g_I^R}{\beta_{I,j}^{mul}} \frac{\sum_{k=0}^{\min\left(i,\left\lfloor \frac{C_j-m}{M_j} \right\rfloor\right)} (-1)^k \binom{i}{k} \binom{C_j-M_jk}{C_j-M_jk-m}}{\binom{C_j}{C_j-m}} \tag{3.13}$$

3.2.2 Complexity of the algorithm

Let us present now the complexity of the **WDMM** algorithm in terms of the characterising values of the network:

- the applied binomial coefficients can be obtained in $O(C^2M^2)$ steps,
- the calculation of the $\beta_{I,j}^{mul}$ values is in order of $O(JC^3M)$,
- the auxiliary variables P_l, P_c and P_n are computable in order of $O(JC^2M^2) + O(HC) + O(HC)$,

- we need $O(J^2M^2) + O(J^2M) + O(JC^2)$ steps to obtain $\gamma^{(0)}$, $\gamma^{(1)}$ and β'.

The sum of these complexities gives the complexity of the whole calculation, that is $O(JC^3M) + O(JC^2M^2) + O(J^2M^2)$. This complexity has to be considered in each step of the iteration. The number of iteration steps depends strongly on the network load and can not be easily estimated.

The computation is obviously more complex for the multifiber case than for the case where only one fiber is allowed in each link. The model for this latter case is presented in [58] and has only the complexity in order of $O(JC^2)$.

3.2.3 Numerical results

In order to observe the accuracy of the **WDMM** algorithm, we compared it with the *Multifiber Link-Load Correlation* model **MLLC** introduced in [52] and with simulations. We present numerical results on the blocking probability of optical channel requests in function of the network load in different scenarios. The scenarios differ in topologies, fiber capacities and traffic patterns. Another basis for model comparison is the computation time, which refers the complexity of the algorithm.

Although our model accepts also traffic models that are better suited to WDM networks, in the studied scenarios the connection requests were assumed to arrive following a Poisson process and with exponential holding time. Considering, that as nearly all published solution, **MLLC** supports only the *Poisson* traffic model this is the only choice that allows the model comparison.

In these illustrative studies the load is set to get blocking probabilities in the range of 10^{-6} to 10^{-1}. This range is too high for a traditional connection based service but can be accepted in the IP over WDM scenario as we mentioned in Section 2.3.2. On the figures the network load is indicated in Erlangs. To accommodate the accepted requests the fixed shortest path routing is applied with the random wavelength assignment described in Section 3.2.

We used the ASONCLES [38] tool for the simulation task. This simulator considers Automatic Switched Optical Networks and allows the investigation of different RWA solutions. The simulations ran until the blocking probability as significant statistical variable reached the confidence level 0.99 with 0.01 accuracy around the point of estimate.

In the computations of **WDMM** the iteration was stopped when the difference between the blocking probability values got in two successive steps decreased under 10^{-6}.

3.2.3.1 Uniform ring topology

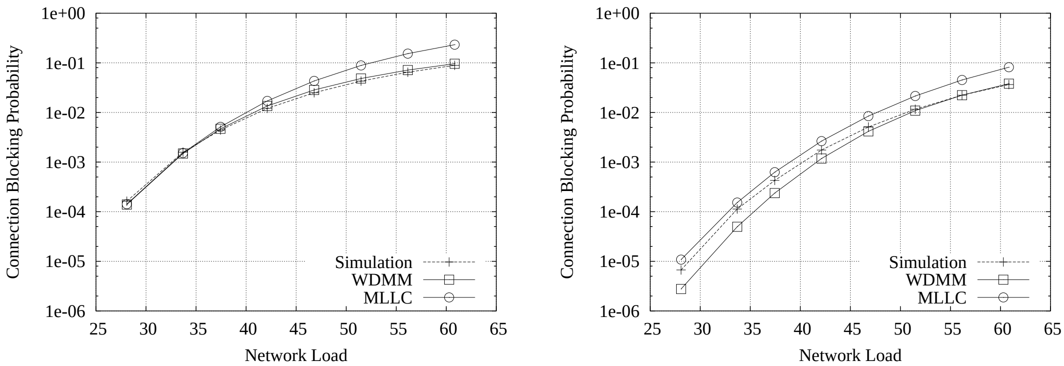


Figure 3.2: Blocking probability in the uniform 13-node ring: for uniform traffic (left) and non-uniform traffic (right)

The first results were obtained using a 13 node ring with $C_j = 24$ optical channels on each link. The number of wavelengths C was set to 24. This network is a single fiber uniform ring that is a very appropriate topology for the model validation because of the high dependence of paths. The numerous common links in the paths imply also the high dependence of the traffic coming from and going to different network nodes. We used a uniform traffic pattern. Due to the uniformity this scenario simplifies the study of basic model properties.

The comparison of the theoretic models **WDMM** and **MLLC** with simulation results can be seen on the left plot of Figure 3.2. One can see that **WDMM** estimates accurately the blocking probability, while the accuracy of the **MLLC** model is very good in the light-load cases, but it overestimates the blocking rate in cases of higher network load. This behaviour may come from the fact, that **MLLC** does not include any fixpoint search mechanism.

In the next study case we set C to 6 and, concurrently, M_j to 4 for each link using the same uniform ring topology. Here we applied a random generated non-uniform traffic pattern. For this multifiber network we obtained the results shown on Figure 3.2 in the right plot.

The large number of fibers causes that in the low-load area our model underestimates the simulated results of the blocking probability, while it is still accurate at high loads. **MLLC** works similarly as in the above case.

Let us introduce now the comparison of the computation times measured for the models that were used in the above study cases. All computations were done on an Intel P4 system running at 2 GHz. The results given in seconds are presented in Table 3.1.

Table 3.1: Computation times for the uniform 13 node ring

Load	*WDMM*	*MLLC*	*WDMM*	*MLLC*
28.08	1	263	< 1	78
33.70	3	262	1	77
37.44	5	261	< 1	78
42.12	7	262	1	77
46.80	10	262	1	78
51.48	15	262	1	78
56.15	21	262	2	78
60.84	31	263	4	78
	Uniform traffic, $M_j = 1$		*Non-uniform traffic*, $M_j = 4$	

The computation times of the model **WDMM** grows with the growth of the load. This comes from the fact that using the same accuracy limit in the fixpoint search, the iteration takes more steps to stop. It is easy to observe that the algorithm **MLLC** consumes much more time than our method. The complexity of this method is not less than $O(HC^5M^3)$ due to the use of recursive steps. This is considerably larger than the complexity of **WDMM** presented in Section 3.2.2.

3.2.3.2 Non-uniform ring topology

The next set of results were obtained using a ring of 13 nodes with different C_j link capacities. The total sum of the capacities in this network is the same as in the case of the previously studied uniform ring, i.e., $24 * 13 = 312$ optical channels, the link capacities

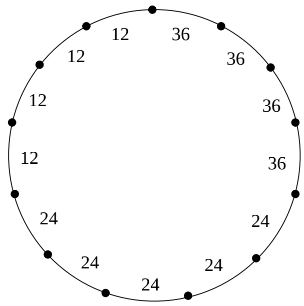


Figure 3.3: 13-node non-uniform ring

can be seen on Figure 3.3. For this topology the **MLLC** model can not be used due to its constraints and **WDMM** is compared only to the simulation results.

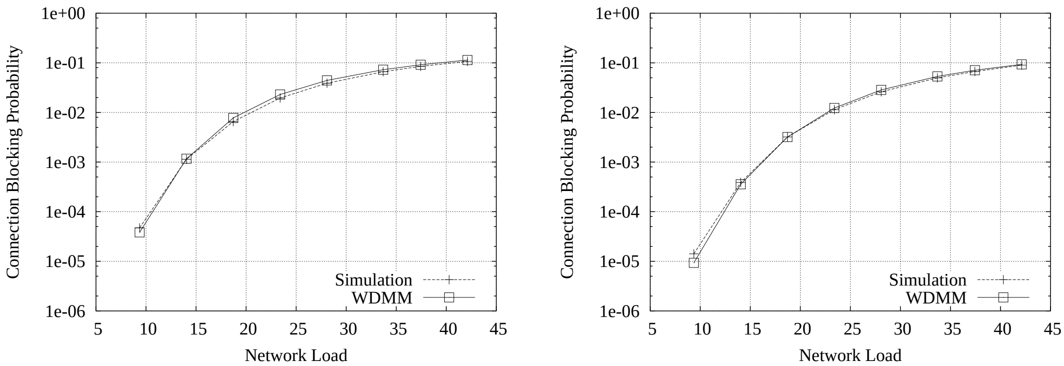


Figure 3.4: Blocking probability in the non-uniform 13-node ring: for uniform traffic (left) and non-uniform traffic (right)

First we applied uniform traffic pattern and the value C was set to 12, hence the number of fibers on link j was $M_j = C_j/C$ accordingly. Left plot of Figure 3.4 shows the results. For the right plot of Figure 3.4 the results were obtained by setting C to 4 and applying a random generated non-uniform traffic pattern. In both cases we can observe that the values computed by the **WDMM** model fit very well the simulation results.

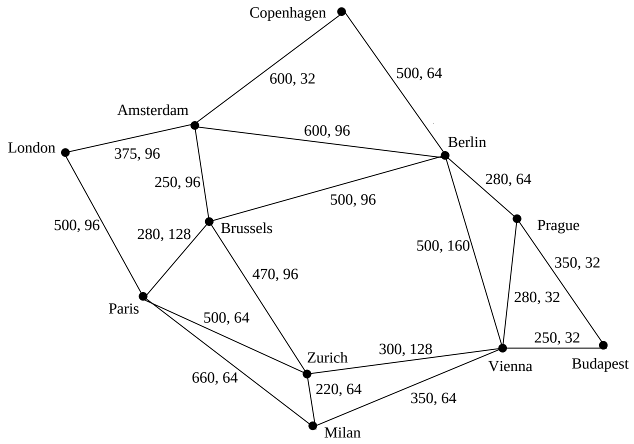


Figure 3.5: Central-West European Network

3.2.3.3 CWEN topology

After the evaluation of the proposed model on regular topologies let us present results using a meshed network. The hypothetic Central-West European network shown on Figure 3.5 consists of 11 nodes and 19 optical links. The numbers on the links are the length derived from the distance in kilometres and the capacity in wavelengths. Its design was based on previous publications [60] and [61] that studied pan-European optical network opportunities. The link capacities and the applied traffic pattern were determined considering a population-distance-based traffic relation matrix.

The number of wavelengths C was set to 32 in the first case and to 16 in the second case. The results are presented on Figure 3.6 left and right plot respectively. We can observe the robust accuracy of the **WDMM** model in both cases.

3.3 Discussion

In this chapter we presented the results of the studies performed considering only the *optical layer*. A theoretical model was introduced to capture the issue of the optical request blocking probability computation.

Summarising our experiences, on the one hand, we have that the **WDMM** model estimates well the blocking probability of optical connection request. This is due to its

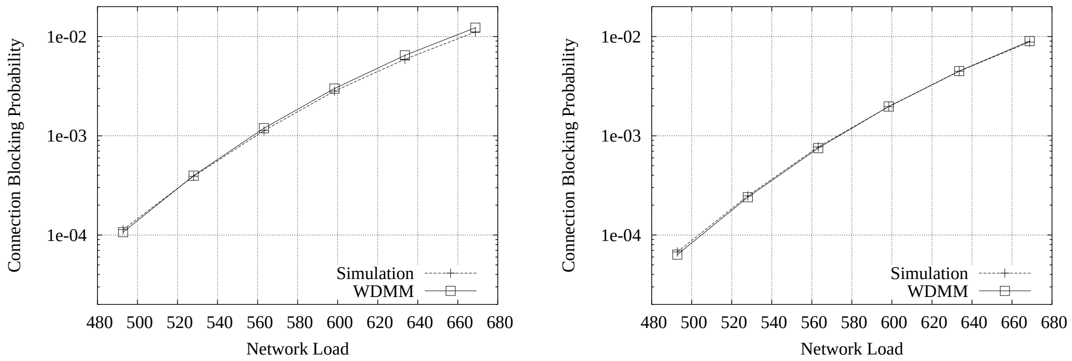


Figure 3.6: Blocking probability in the Central-West European network with 32 (left) and 16 (right) wavelengths per fiber

structure, which considers wavelength and load dependency among network links. On the other hand, the time needed for the computation is acceptable.

The results obtained in the comparative studies in Section 3.2.3 show that **WDMM** might provide a light underestimation in some scenarios. This range of the inaccuracy depends on load, traffic pattern and topology of the network. We found significant inaccuracy only in the cases where at least one of the following properties held:

- The network topology design did not suit well the offered traffic pattern.
- The number of wavelengths was moderate and correspondingly the number of fibers on the links was elevated.
- The load, and accordingly the blocking probability, was not considerable, e.g. the connection blocking probability was less than 0.001.

Chapter 4

Performance of the *data layer*

4.1 Introduction

This chapter treats performance analysis in the *data layer*, with a specific interest on the IP routing solution. The aim of this work was on the one hand, to find efficient but correct ways to assess network capabilities. On the other hand, the study of the existing algorithms can help us to develop new ones providing higher performance.

Many works were presented in the last decade on IP networks performance analysis and about the viable solutions. The emerging of the IntServ [14] and DiffServ [15] architectures for the future Internet, together with the possibility of building backbone Autonomous Systems based on Multiprotocol Label Switching (MPLS) [62] and the need for provisioning Quality of Service (QoS) in the Internet spawned a burst of work concerning new, QoS-based, dynamic routing algorithms suitable for implementation in IP networks.

Some examples obtained from the research effort on QoS routing can be found in [2, 20, 63, 64]. Besides the examination of the service quality aspect that is mostly represented in our studies as the effective transport rate of flows and the prevalence of starvation situation, also other points of view can be found for comparison of routing performance. Researchers focused their attention upon other aspects of QoS routing too, namely protocol overhead [2, 21] and implementation issues [63, 64].

Nearly all of the previous works assume flow based models representing the IP traffic via dynamically arriving requests, but with constant and fixed requirements, like circuit switched connections. They do not consider the elasticity issue in any grade.

Among others, the studies [24, 25, 26] addressed the problem of stale link state information, analysing different policies for triggering information exchange among nodes. The authors of these works considered network models where connection requests are generated with arrival processes and holding times that suit well to the IP traffic. However, they assumed constant bandwidth requirements and evaluated the connection blocking probability. Their results clearly show that different policies for information exchange lead to different performance of routing algorithms.

The benefits of the shortest path algorithm in heavy-loaded networks was analysed in [35]. The author even proved that for guaranteed bitrate flows this solution is optimal when the load offered to the network tends to infinity. The extension of these results to elastic flows is not straightforward, nevertheless, many simulation results confirm this intuition [20, 22]. On the other hand, dynamic routing algorithms provide significant gain in light-loaded networks. Some papers as for instance [22], present routing solutions working in a load-dependent manner to catch the gains of both static and dynamic algorithms.

In the following sections we summarise the results of [23, 40, 41, 65, 66] by presenting models introduced for the observation of elastic traffic and by proposing new routing algorithms. The results are obtained by simulation, dealing with both previously proposed and novel routing techniques.

4.2 Evaluation of routing algorithms in the Internet

This section addresses the problem of evaluating routing algorithms via simulation in packet-switched networks when elastic traffic is involved.

4.2.1 Motivation

Our aim is to highlight some deficiencies of classic approaches that fail to capture both the complex interactions of connections traversing multiple bottlenecks and common user behaviours. Our approach is devised to overcome these limitations and it is particularly suited for the evaluation of routing algorithms in presence of best-effort traffic. The simulation results presented offer a deeper insight into well-known routing algorithms. Through this analysis it is clear that quantitative and also qualitative behaviours of dynamic routing algorithms based on traffic measurements may be fairly different depending on the nature of the traffic loading the network, as well as depending on its interactions with the network parameters and behaviour.

As one could see in Section 2.6.4 purely theoretical tools can have hard limits if more realistic traffic and more complex algorithms need to be modelled. Most of the network functions can be efficiently implemented in a simulator, although not always in a very easy way. A big advantage is the robustness in realising new models.

The drawback of simulation can be that no general and straightforward statements can be done based on the obtained results. While theory can provide clear and proved answers on several questions regarding the performance of network solutions, simulation results can always depend on the investigated scenario. For instance, analysing the performance of QoS routing algorithms to get general conclusions several different topologies and traffic loads need to be simulated. Since networks need a routing strategy that is first of all robust to changing traffic patterns and loads, achieving high performance in some selected scenarios, while performing poorly in others may be worse than using a routing strategy whose performance is possibly not the best for any traffic pattern, but is predictable and ensures reasonable performance under any circumstance.

4.2.2 Basic assumptions

Performance evaluation in IP networks can be performed with analysis on packet-level. Note that this solution results in a very complex model, whose simulation is generally too expensive to obtain reasonable results for any realistic network. Instead, we use flow-level traffic models in the simulations. Going more in details on the *data layer* in the general network vision presented in Section 2.1 we use the following general assumptions in the rest of this section.

- The network nodes have unlimited capacity in both switching and throughput. A node of the *data layer* is able to switch any flow from any incoming link to any outgoing one if the links can accept the traffic.

- Buffering is not modelled, assuming that the packet loss and delay problem is implicitly solved with the elastic model of the flows.
- We do not consider the details of the switching mechanism and its timing.
- Network users are attached to the nodes.
- Generators connected to network users pump the traffic in a flow oriented manner according to the total network load ρ and the traffic pattern.
- In any instant, for any link, the sum of instantaneous data rate of flows carried by the link can not be greater then the link capacity.
- Flow management functions are present in the network and we assume a control plane that fully connects them and carries the needed OAM-type information. This allows us to model network traffic as flows, even if from network management point of view the endpoints are not connection oriented.
- Once a path is assigned to a flow, the data transport is performed using this path until the user finishes to carry traffic.
- The instantaneous data transport rate of the flow f is identical on each link of the assigned route and is stated as B^f.
- Flows may have a maximum data rate B_M derived from the access link or application output constraints.

Connections originating from users with service guarantees receive the resources contracted in the SLA. However, analysing the *data layer* that models the IP network, we are rather interested in the performance of the best-effort traffic. For this type of traffic the network does not guarantee any value of the data rate and these flows share link resources adaptively. We assume max-min-fair-sharing described in [35] and analysed in [67], since no priority among the flows is considered. This leads to a behaviour where elasticity is a dominant phenomenon.

The elastic approach can be easily validated if we consider that the Internet traffic is mostly carried over TCP. Adaptivity is embedded in the congestion control algorithm of the TCP protocol which is a closed-loop protocol with implicit feedback from the network. The modelling technique used in our studies does not address the details of TCP protocol, but tries to partially capture its self-controlling behaviour from the consumed bandwidth point of view. For instance, the reaction to network congestion spread the connection over time and result in congestion periods that last longer than what can be estimated with a model devoid of any feedback feature. The aim of TCP congestion control is sharing network resources following max-min fairness, and this is exactly the *adaptivity* criterion we use. Indeed, some authors in [68] have suggested that feedback phenomena, making the load offered to the network dependent from the network status and parameters, might be responsible for the LRD (long range dependent) behaviour of Internet traffic[1].

In addition, the model considers the concept of *flow starvation*, another typical feature of the Internet, specially related to Web browsing. It is well known that caused by the lack of admission control the number of connections using the network at the same time is virtually unlimited. As a result, TCP packets might be lost due to congestion, leading to frequent retransmissions and consequently longer transfer completion time; similarly, UDP packets belonging to a streaming multimedia connection could be dropped, leading to unacceptably poor playback quality. In either case, users will abort the connection and may retry to download the data later.

This behaviour can seriously affect the network performance, since the network does some effort to transfer information which might turn out to be useless. Furthermore, resources devoted to aborted connections appear to be unnecessarily taken away from other connections. To emulate the sudden closure of starved connections, we define a connection as *starved* when the amount of bandwidth resources it receives falls below a given threshold.

4.2.3 Modelling elastic traffic connections

We defined two modelling paradigms of data transfer in [41], they are used in the flow-level analysis of different user traffic. After their description we give a more detailed insight into the issue of modelling the starvation effect.

[1]The study of this issue is out of our scope.

4.2.3.1 Time-Based model

The most common traffic model used in high-level communication networks simulation is Time-Based (*TB*). In this model, connections are described by their duration (holding time), and by their bandwidth requirements. For example, in telephone networks, connections require a CBR (Constant Bit Rate) service; this piece of information, integrated by the connection interarrival time and the required bandwidth, is sufficient both to determine the traffic intensity produced by the connection generator and to completely characterise the connection from the network point of view. The network can allocate enough resources to guarantee the required QoS. A similar approach is valid for variable rate connections. Any given CAC (Connection Admission Control) scheme can be implemented in simulation to control the network load and satisfy the QoS requirements of connections.

The effectiveness of this model is questionable if best-effort, data-based connections are considered. Defining a priori a *duration* for such a traffic is virtually impossible, since these connections are data-centric, considering the information to be exchanged the most important part of the communication. Typically using a best effort service, these data-centric connections adapt their sending rates to the current network congestion, which cannot be known a priori. Thus a new model must be used to describe this kind of traffic.

In a Time-Based model, we can still derive some kind of connection duration, but it can usually be obtained making hypothesis on the bandwidth the connection could obtain in the communication. Since *TB* is not able to model veritably the whole process of data transfer, indeed, a distortion of the network throughput can affect the analysis. While in the case of CBR traffic the total amount of traffic offered to the network could be precalculated using the predetermined holding time, in the case of elastic traffic no obvious derivation can be done. If the actually achieved per flow throughput, i.e, the bandwidth that is currently assigned to the flow, decreases, the instantaneous offered traffic decreases too. The flow request arrival rate remains invariable but the data amount transferred by the connection gets smaller.

4.2.3.2 Data-Based model

The (*TB*) model fails when we try to apply it to the typical data exchange in Internet, such as data downloads from the Web or FTP transfers. In all these situations the objects of communication are files and the communication ends when the last bit of the file has been acknowledged. Even if we declare a maximum communication rate B_M that emulates the limitations introduced by the application, the transport protocol, or the access network, it is of little help, because it only sets the lower bound for the connection duration. However, considering the *starvation effect* an upper bound can be also calculated for this value.

Indeed, the actual time required to successfully end the data transfer depends on many factors, and mainly from the varying available bandwidth while the connection is active. Thus we introduce the Data-Based (*DB*) traffic model, where the connection lasts until all the data amount S_D associated to the flow is transmitted. This way no distortion affects the offered traffic. Figure 4.1 illustrates the main difference to *TB*.

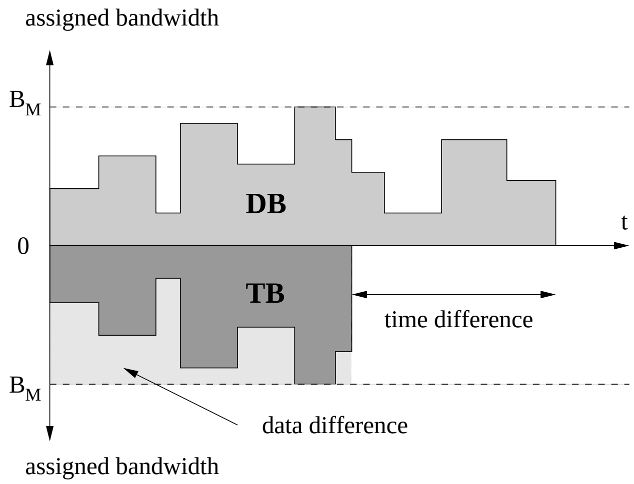


Figure 4.1: Traffic models *TB* and *DB*

Since in the flow-level simulation approach the packet transmission is not simulated, we need to estimate dynamically the time the transmission ends. Since connections with elastic bandwidth requirements generally use all their share of available bandwidth (up to their maximum transmission rate) duration can be determined only by monitoring *semi-*

continuously the instantaneous bandwidth allocated to the connection. In a simplified model the values of assigned bandwidth can change only at time-points of the arrival or departure of connections with any characteristic. We can obtain them by running a max-min fair-share algorithm that requires a full recalculation of the bit rate of all connections currently routed in the network every time a flow is opened or closed. Actually this solution provides an upper bound to the performance, since the max-min fair-share represents an ideal working situation for any congestion control protocol that aims at equally dividing resources among the elastic flows, as the TCP protocol tries to achieve.

4.2.3.3 Modelling the starvation

To model the starvation effect, i.e., users that abort the data transfer due to poor performance, we introduce a *starvation threshold* b_m that will be used to identify starved connections and realises an implicit definition of the minimum acceptable bandwidth for the flow. If for any flow f the current per-flow bit rate estimate on a bottleneck link B^f drops below b_m, then the connection with the longest remaining time of data-transfer is picked and terminated. The choice of the connection that would hold on for the longest time models the behaviour of a general user that aborts the most delaying data transfer. The only exception for the choice is a currently started connection whose termination, emulating this way an implicit *a priori* CAC, is not allowed.

A starvation situation can evolve not only after the opening of a new flow. In certain scenarios the departure of a flow causes the starvation of other flows.

This mechanism allows us to define the *starvation probability* p_s as the ratio between connections that are prematurely terminated and the total number of connections that entered the network. The information about service quality that this probability value gives us, is similar to the blocking probability value in networks with connection admission control function and has to be considered in the performance evaluation.

4.2.4 Formulation of the routing algorithms

The flow-based approach of the introduced data and traffic models reflects the concept of flow identification possibility in the transport services that support QoS, e.g. IntServ, DiffServ or MPLS. According to this idea we make an assumption about the routes chosen to carry best effort traffic. A transport path is assigned to a flow request arriving in node s, choosing it with the routing algorithm once and for all, i.e. the same path is used for all the flow-data. This way we exclude the option of using multiple paths, as for instance in-node flow control techniques of packet-based network models that allow to assign different paths to the fragments of the transported data.

To understand the workout of the used routing algorithms we introduce a general notation and formulation: a generic network is referred as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices (accord to network nodes), and $\mathcal{E}$ is the set of edges (accord to network links) that can be weighted by different metrics ω.

A path $\pi(v_s, v_d)$ of length $n = ||\pi(v_s, v_d)||$ hops is defined as a sequence of n distinct edges e_i joining v_s and v_d, where $v_s, v_d \in \mathcal{V}$, $e_i \in \mathcal{E}$, $\pi(v_s, v_d) = \{e_1, e_2, ..., e_n\}$.

A set $\mathcal{P}_\mathcal{G}$ is defined as the set of all paths existing between any two distinct vertices of $\mathcal{G}$, i.e., $\mathcal{P}_\mathcal{G} = \{\pi(v_s, v_d) | v_s, v_d \in \mathcal{V},\ v_s \neq v_d\}$. This set can be divided into disjoint subsets corresponding to the starting and ending node of network route: $\mathcal{P}_\mathcal{G}(s, d) = \{\pi(v_s, v_d)\}$.

A weight w_i^ω can be assigned to each link e_i using any metric ω that combines topological, physical or traffic-related characteristics of that link. A cost $c^\gamma(\pi)$ is assigned to each path π using a combination of the weights of its links where γ identifies the combination method. Following this, an order relation $\prec$ in each subset $\mathcal{P}_\mathcal{G}(s, d)$ is established among the paths according to their costs c^γ, and we can refer to a path π_i as "lighter than" π_j if $\pi_i \prec \pi_j$ holds. If neither $\pi_i \prec \pi_j$ nor $\pi_j \prec \pi_i$ hold, π_i and π_j are *equicost* paths by the given cost function γ.

If we do not indicate it differently, path costs are real values and the applied ordering relation is the ascending ordering among real values, i.e., π_i is lighter than π_j according to the cost metric γ if and only if $c^\gamma(\pi_i) < c^\gamma(\pi_j)$.

In routing decisions the combined use of different costs are allowed. The simplest combination is the concatenation, formulated as: $c^{\gamma_1} \oplus c^{\gamma_2}$. This combined solution implies that after the ordering by c^{γ_1} the *equicost* paths will be reordered by c^{γ_2}. Conditional use of different cost functions and, as one can see in latter sections, reduction of the subsets $\mathcal{P}_\mathcal{G}(s, d)$ can be applied too.

Commonly used weight metrics of edges are the following.

H assigns the same constant weight to each link, $w_i^H = a|e_i \in \mathcal{E}$. As most common value, a can be set to 1 representing one *hop*, in the following we will use this value.

L assigns constant weight for link e_i, representing its *length*. $w_i^L = l_i|e_i \in \mathcal{E}$.

RES assigns as weight the value of *residual* bandwidth measured on the link, i.e., the bandwidth available for the best effort traffic. $w_i^{RES} = B^{RES}(e_i) = B^{TOT}(e_i) - B^{GAR}(e_i)|e_i \in \mathcal{E}$, where $B^{TOT}(e_i)$ is the total bandwidth capacity of link e_i and $B^{GAR}(e_i)$ is the bandwidth currently assigned to traffic with bandwidth guarantees.

ABW weight metric is based on the bandwidth value that is *available* for a new best effort connection on the link, assuming fair sharing of the residual bandwidth among the best effort connections. $w_i^{ABW} = \frac{B^{RES}(e_i)}{N_i^{BE}+1}|e_i \in \mathcal{E}$, where N_i^{BE} is the current number of best effort connections, i.e., identified elastic flows on link e_i.

The weight assigning functions H and L are time independent, while RES and ABW depend on the current network state. Routing algorithms that consider information on the network state are known in the literature as *dynamic* or *adaptive* schemes.

Let us define a special cost function that is not derived from edge weights: $c^{RND}(\pi)$ is a randomly chosen value between 0 and 1. RND costs can be assigned either a priori or at the moment of cost evaluation, this two versions are referred as $RNDA$ and $RNDC$ respectively. We impose that there are no *equicost* paths in the subsets $\mathcal{P}_{\mathcal{G}}(s,d)$ through the costs $RNDA$ and $RNDC$.

Now we give a brief description of some known routing algorithms. A set of them will be used in the analysis assessing the difference between the traditional traffic modelling technique based on holding time (*Time-Based*), and the novel one based on the amount of generated data (*Data-Based*). Later on these algorithms will be used in the analysis of new routing algorithms as comparison basis. Unless otherwise stated, the choice is always the “lightest” path with the given ordering system.

- *Fixed-Shortest-Path* (*FSP*): for each source-destination pair, the algorithm determines the path with the minimum hop count and routes flows along that path. If two or more *shortest* paths exist, the algorithm chooses one at random[2], that is used during the whole data transmission period of the flow. This is the routing algorithm commonly used in the Internet (e.g., OSPF). Its formal definition is a concatenation of orderings by the cost functions $c^H(\pi) = \sum_{e_i \in \pi} w_i^H$ and $c^{RNDA}(\pi)$, or in shorter form: $c^{FSP} = c^H \oplus c^{RNDA}$.

- *Widest-Shortest* (*WS*): for each source-destination pair, the algorithm determines the path with the minimum hop count. If more than one such path exists, it breaks the tie by choosing the one with the largest available bandwidth for the new connection. The formal definition of the ordering: $c^{WS} = c^H \oplus c^{ABW} \oplus c^{RNDA}$, with

$$c^{ABW}(\pi) = \frac{1}{\min_{e_i \in \pi} w_i^{ABW}} \tag{4.1}$$

  The implemented scheme operates as in the original proposal in [69], except in the case of multi-class scenarios. We use the share of bandwidth available to the new flow taking into account that resources used by guaranteed traffic are no longer available to the elastic best-effort traffic. The original proposal considers in the link weight calculation the whole physical link capacity $B^{TOT}(e_i)$.

- *Minimum-Distance* (*MD*): as proposed in [20, 67], for each source-destination pair the path π is chosen which minimises a quantity that consider both the number of hops in path and the available bandwidth on its links, and can be described as a special kind of distance between the end nodes. Formally we order by: $c^{MD} = c^{RS} \oplus c^H \oplus c^{RNDA}$, where

$$c^{RS}(\pi) = \sum_{e_i \in \pi} \frac{1}{w_i^{ABW}} \tag{4.2}$$

  Note that the authors in [67] state that the algorithm is to be implemented over networks employing max-min fair-share-based congestion control. In our case, we assume that the network can determine what the current max-min fair-share situation is at the time when the routing algorithm is executed, though this hypothesis may be optimistic.

[2]It is possible to optimise the $RNDA$ order, for example using load-balancing criteria.

- *Shortest-Widest* (*SW*): first, it identifies the maximal-bandwidth paths, then it breaks tie by choosing, among the paths with the maximum available bandwidth for a new best effort connection, the one with the minimum hop count. This scheme is similar to the one proposed in [64], adapted to the presence of different QoS classes. Its formal definition is given by the ordering cost: $c^{SW} = c^{ABW} \oplus c^{H} \oplus c^{RNDA}$.

- *Load-Dependent* (*LD*): this algorithm was proposed in [22] noticing that when the network is heavily overloaded the best possible algorithm is *FSP* while an adaptive algorithm may enhance network performance when the load is light. The algorithm chooses the path π minimizing the cost $c^{MD}(\pi)$ (see the *MD* algorithm) if

$$\frac{c^{H}(\pi)}{c^{MD}(\pi)} > kB_M \tag{4.3}$$

  otherwise the choice is based on *FSP*. The predefined value k is a fine-tuning constant for the algorithm, its impact was studied in [22].

Advertising updated QoS parameters, such as the currently available bandwidth among network routers is assumed to occur every s seconds. However, in the simulations presented now we chose an *ideal* instantaneous update, in order to avoid complicating the interpretation of results with the problem of stale routing information [2]. Later, in Section 4.3 we introduce the model of the update latency and analyse how it affects network performance.

4.2.5 Comparison of *TB* and *DB* approaches

In this section we analyse the traffic models and present numerical results. The simulation results were obtained by running ANCLES [37]. This tool offers a set of routing algorithms and performance indices to study. The simulator supports the above introduced traffic models but does not model the overhead resulting from signaling traffic.

Each simulation ran until the 95% confidence level with 0.05 accuracy of the estimate. We applied the same tool and settings in all the further studies presented in this chapter.

4.2.5.1 Simulated network and scenario

In order to decouple performance comparisons from a particular topology, a randomly generated network topology was selected. The generating tool was the software GT-ITM, introduced in [70]. The resulting topology comprises 32 nodes with an average connectivity degree of 4 and uniform link capacity $B^{TOT}(e_i) = 10$ Mbps for each link e_i. The offered traffic is generated by best-effort source generators (one per node), trying to set up connections with a maximum bandwidth B_M of 1 Mbps. Each connection requires a bulk data transfer whose size S_D is randomly chosen from an exponential distribution with average 500 kbytes. A uniform traffic pattern is simulated, i.e., when a new flow request is generated, the source-destination pair is randomly chosen according to a uniform distribution.

This is hardly a scenario favouring QoS routing, and we chose it exactly for this reason: non-uniform, "hot-spot" traffic scenarios may favour specific QoS routing algorithms, altering the perception of the relative merits of each algorithm. As a consequence of this choice, the gains we can expect from QoS routing are limited compared to *FSP*.

The load offered to the network is expressed in Mbps. The mean number of flows that arrive in the network each second can be easily computed knowing the mean bulk data size S_D. Dividing it by 32 we get the requests per second value per generator. For instance, an offered load of 320 Mbps corresponds to a nominal network flow request arrival rate of $(320\,Mbps)/(8 \cdot 500\,kbit) = 80$ flow per seconds, hence each single generator offers $80/32 = 2.5$ request per second load, i.e., averagely 10 flow requests in every 4 seconds. The actual throughput carried by the network depends on the efficiency of the applied routing algorithm.

The holding time H_t of Time-Based connections is computed based on the information amount and the maximum required bandwidth:

$$H_t = S_D/B_M \tag{4.4}$$

If instead we use the DB model, this calculation gives a lower bound of the real holding time. The elastic manner of their assigned bandwidth can affect the time needed to transfer the currently remaining traffic in each moment.

As performance indices, we report the starvation probability p_s and the average bandwidth or throughput per connection T, defined as the mean value of bandwidth that connections obtain during their lifetime, averaged among all the source/destination pairs. Only the connections that successfully complete the transfer are taken into account. Notice that in a resource sharing environment this is not the average resource occupation divided by the number of flows, since flows have all the same weight, regardless of the amount of transferred data. In the Data-Based scenario, we report results for the *dilatation factor* D_f, i.e., the ratio between the average completion time of a connection and its minimum completion time (equation 4.4). All results are reported versus the average network offered load.

4.2.5.2 Shortest-Path performance evaluation

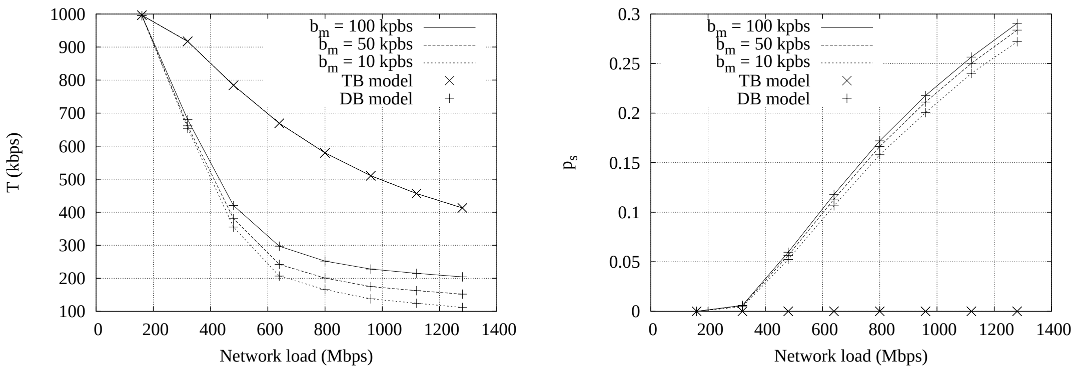


Figure 4.2: Average bandwidth per connection (left) and starvation probability (right) for *DB* and *TB* models using *FSP* routing with different starvation threshold b_m

The left plot of Figure 4.2 presents a comparison of the average bandwidth T obtained by modelling best-effort flows with the Time-Based approach (x points) and the results obtained with the novel Data-Based connection model (+ points), when the *FSP* algorithm is used. For both *TB* and *DB* model, we present simulations with different starvation threshold b_m set to 100 kbps, 50 kbps and 10 kbps.

The difference in the performance results of the two approaches is striking. Both approaches show T starting from 1 Mbps when the offered load is low, i.e., there is no congestion in the network. In a non congested network the two models work the same way, since data is transferred with the maximum bandwidth and there is no holding time dilatation in the case of the *DB* model.

On the other hand, the two models drastically differ when the offered load starts increasing. Indeed, the Data-Based model shows a decrease in T and the performance tends to the threshold b_m that acts as a lower bound on this performance index. On the contrary, the Time-Based approach shows a smoother decrease of the average bandwidth, and independency on the threshold b_m.

When the available bandwidth goes below the starvation threshold, connections begin to be aborted. We presented simulation results on the starvation probability in *DB* and *TB* models on the right plot in Figure 4.2. In the case of the Time-Based approach p_s is always zero, explaining the independence of T on b_m. Although simulation were performed with different b_m thresholds, in the case of the Data-Based approach we do not measure a significant variation in the starvation probability. This behaviour can be interpreted as a much lower impact of the threshold on the starvation of connections than on the assigned bandwidth T. In other words, major part of the connections that suffer the starvation effect with threshold b_m, would be interrupted also using a lower threshold b'_m. Since flows with threshold b'_m are less sensitive on congested situations, the mean lifetime of the starved connections is not smaller than of those using threshold b_m. Thus, using lower starvation thresholds implies that a not completed flow wastes the network resources for longer time, while the ratio of the starved flows does not change too much.

To explain the not negligible differences in the performance measures, on the one hand the experimented characteristics of T and p_s are due to two factors derived from the model structure:

- In the Data-Based approach, when congestion arises in the network, connections are throttled. This causes a stretching-out of the duration of the connections that require more time to be successfully ended, thus spreading congestion over time. A sort of avalanche effect occurs, as the number of connections increases on a bottleneck link, reducing the available bandwidth for each of them. In other words, there are always more connections that share network resources when using the *DB* model, with respect to the *TB* model.
- On the contrary, in the Time-Based model, the congestion does not spread over time, since the holding time of the flows is determined *a priori* and it is not affected

by the congestion. Thus, no avalanche effect on congestion is possible, since the average number of connections is independent of congestion.

On the other hand, we experienced that the data amount transmitted with finalised flows, is only a part of the total offered load. When b_m was 100 kbps the ratio of finalised and offered traffic decreased in a roughly linear way down to 40% with the load increasing up to 1260 Mbps. Although this degradation effect of network throughput at high loads is similar for both the *TB* and *DB* models, the causes are different:

- Using the *DB* approach we get not finalised traffic due to the starved flows.
- In the case of *TB* model there is no starvation. Here we have to refer the throughput distortion mentioned in Section 4.2.3.1: the same arrival rate as in *DB* can reflect in lower amount of data to be transferred in *TB*.

4.2.5.3 Comparison of routing algorithms

In this subsection, we present a set of results that aims at showing the difference in performance obtained with the two traffic models when QoS routing algorithms are adopted in the network. Generally, when comparing routing algorithm performance, we are interested in the relative merit with respect to well-established algorithms, such as the Fixed-Shortest-Path (*FSP*). Thus, as performance index, we can select the relative gain of the average throughput η obtained using a QoS-aware routing algorithm with respect to *FSP* i.e., $\eta(r) = T(r)/T(\mathit{FSP})$, where $T(r)$ is the average throughput T measured when using the r routing algorithm.

The left plot of Figure 4.3 presents the relative throughput gain obtained with the Time-Based model, while the right one shows that with the Data-Based model. Results are for a scenario where the threshold b_m is 100 kbps.

Throughput results are comparable to previous studies (e.g. [20, 22]). The adaptive *MD* and *WS* algorithms outperform *FSP* routing on networks with relatively low load, as they manage to exploit the spare bandwidth that is present on lightly loaded links. Instead, they provide a worse performance when the network becomes overloaded, because of the waste of bandwidth that occurs when longer paths are selected. This behaviour of adaptive routing algorithms can be observed in nearly all cases that we examined.

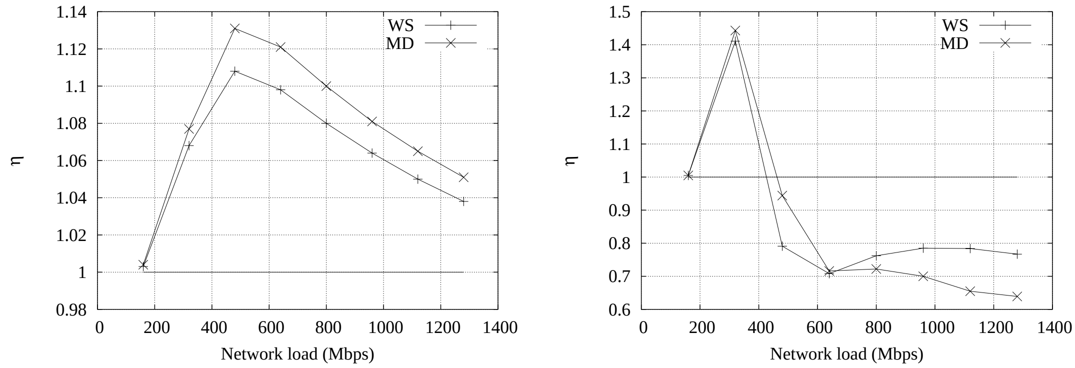


Figure 4.3: Relative gain η of *MD* and *WS* algorithms for *TB* (left) and *DB* (right) models

Let us analyse the differences of performance using the different traffic models. On the one hand, the *TB* approach shows a much wider range of offered load where the dynamic routing algorithms outperform the *FSP* algorithm, although the gain is never larger than 15%. On the other hand, the more realistic *DB* approach shows that the load range where the dynamic algorithms perform better than the *FSP* is much smaller. Moreover, in this range, the maximum obtained gain can be much larger (up to 45%), but the transition to the overloaded region where the *FSP* performs better is much sharper.

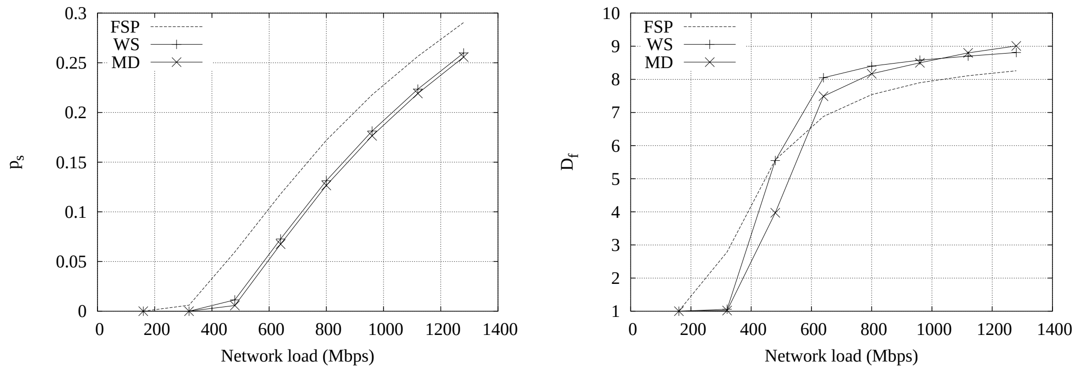


Figure 4.4: Starvation probability of *MD* and *WS* algorithms (left) and dilatation factor with *DB* models (right)

To provide a deeper insight, the left plot of Figure 4.4 reports the starvation probability. Only the values achieved with the *DB* model are presented, since for *TB* all the values were zeros as in case of *FSP*. They suggest that the waste of bandwidth caused by the starved connections can be rather high, as the starvation probability grows to large values as soon as the offered load is larger than 450 Mbps.

Correlating Figures 4.3 and 4.4, it seems clear that the peak throughput gain of *MD* and *WS* with respect to *FSP* is coincident with the network load where connections begin to starve if *FSP* is used, but still receive a satisfactory service if *MD* or *WS* are used. Even for higher load regions we find that the starvation probability is smaller when the *WS* or *MD* algorithms is used. This suggests that the network bears a larger number of simultaneous connections, each one obtaining a smaller throughput, but still higher than its starvation limit.

The average dilatation factor for the different routing algorithms is plotted on the right plot of Figure 4.4, obviously, only for the *DB* model. It can be noted that the *WS* algorithm degrades its performance very quickly as soon as the offered load is higher than about 400 Mbps, while the *MD* algorithm performs better than the *FSP* algorithm up to 600 Mbps. As one could expect, the curves of this plot are correlated with the the values of average assigned bandwidth that are not plotted explicitly. Even if the relation of averages can be complicated, it is easy to see that the less is the assigned bandwidth, the larger is the stretch-out of connection holding time. According to the starvation threshold of the scenario, the maximal dilatation can be calculated as $B_M/b_m = 10$, which is an asymptote in the right hand plot of Figure 4.4.

The elastic traffic model, together with the run-time starvation detection, shows that the evaluation of routing strategies based on inadequate (or too simplistic) models, such as the Time-Based one, is prone to gross approximations and even mistakes. Thus in all further studies we used the more realistic *DB* model.

4.3 Networks with stale link state information

As referred before, adaptive, dynamic routing schemes outperform *FSP* in situations where network load is not too heavy. On the other hand, benefits of QoS routing are often questioned, mainly on the basis of the additional costs[3] and the risk that stale information about the network status may lead to wrong routing decisions.

In this section we introduce a model that includes a timeout-triggered protocol function of link state information exchange. We analyse the impact of out-of-date information with different timeout values on several QoS routing algorithms.

4.3.1 Information distribution

Any distributed routing algorithm requires the exchange of information among nodes in order to compute the available routes and their current costs. The advantage of QoS routing is maximal when the routing algorithm exploits information on current utilisation of resources. This information, however, is prone to error measurements, and, most of all, it quickly becomes outdated. Indeed, out-of-date load information can even lead to wrong routing decisions that can cause an avalanche effect forcing other route selections to choose the wrong paths. New routing algorithms are often proposed without considering the key issue of robustness to non optimal working conditions.

The amount of the exchanged information, together with the impossibility of distributing load information in real-time, may eventually counterbalance the advantages of QoS routing in favour of the simple, traditional Fixed Shortest Path routing. *FSP* was shown to guarantee a good network utilisation in heavily–loaded networks with uniform traffic pattern [67].

Several studies [2, 25, 40] addressed the problem of unreliable link state information, analysing different policies for triggering information exchange among nodes. These works clearly show that different policies for information exchange lead to different performance of routing algorithms.

We used the above introduced traffic modelling technique and evaluation methods to compare different QoS routing algorithms with stale information. The main performance measures for elastic traffic are the average throughput T obtained during the flow life and the probability of starvation p_s. The aim of a QoS routing algorithm in this case is finding routes that maximise the average throughput that best effort flows experience while minimising the starvation.

[3]both in term of computing complexity and increased signaling

This section contains studies on the sensitivity of different routing algorithms to inaccurate link state information. This is mainly a consequence of the impossibility of distributing information in real-time as mentioned above. Other phenomena, like inaccurate link measures, may also influence the results, but are generally less critical. The inaccuracy comes from two separate effects. On the one hand, it is difficult to measure the current bandwidth *available* for elastic traffic. On the other hand, the number of best effort flows N^{BE} currently active on a link can be only estimated, since this type of traffic is packet based in reality. The measurements made by the router can consider only a window of finite number of packets and there is no guarantee that this window contains packets from each flow which currently uses the link. Since the number of connections is an indispensable factor during the calculation of the *estimated* available bandwidth for a new connection, the inaccuracy influences the decision of dynamic QoS routing algorithms, e.g. those introduced in Section 4.2.4.

4.3.2 Model and problem formulation

We consider a Timer Based Trigger information update protocol [2], disregarding implementation details about the protocol and focusing on the degradation of performance induced by large update times t_u. While it has been shown that other update policies can outperform the time-driven policy, in stationary traffic conditions (as in the simulations) all threshold-based policies can be described in terms of time-driven ones, since no sudden change in the traffic triggers an information update before the timeout expiration. In our model the triggers of network links are not synchronised among themselves and in a more detailed model different t_u values could be considered for the links.

The inaccuracy in the number of connections on a link is modelled by a multiplication with a random variable with uniform distribution over the $(1 - f_{ina}, 1 + f_{ina})$ interval, where f_{ina} is a tuneable *inaccuracy factor*. Since this inaccuracy is strictly connected to the burstiness of the best effort traffic entering the network as well as to the measuring interval, the factor has to be set to a reasonable value that will be than valid for each link of the network.

We need a rather simple extension to the notation introduced in Section 4.2.4 to include the update protocol in the model. Two of the weighting metrics are touched by the non–continuous update of link state information:

$RES(t_u)$ is the version of metric RES, but the $B_{RES}(e_i)$ value of the residual bandwidth on link e_i is not updated at each traffic change. For each link a series of samples is available, captured by asynchronous evaluation in each t_u timeunits.

$ABW(t_u, f_{ina})$ weights vary in two main points from the simple ABW weights. First, the $B_{RES}(e_i)$ values of the bandwidth on link e_i that remains available for the best effort connections is updated only each t_u timeunits asynchronously for the different links. Second, also the N_i^{BE} number of active best effort flows on link e_i is updated this way and, moreover, it contains the inaccuracy of f_{ina} factor as described above.

The QoS routing algorithms that consider the stale link state information work the same way as described above, but using these varied weighting metrics. The evaluation of these changes can be performed only if the information error can be expressed in stochastic terms as a function of t_u (see for instance [2]). Indeed, the only viable approach seems to be the *a-posteriori* verification that the heuristics selected for the QoS routing are resilient to incorrect or outdated link state information.

4.3.3 Simulation Results

In the following we analyse how the use of outdated information affects the performance of the QoS routing functions. Here we present simulation results to show the indispensable influence of the update period scale.

4.3.3.1 Topology and scenario

We used the network introduced in [22]. The topology, called Switched Cluster Topology (SCT), is shown in Figure 4.5. This is a mesh topology, with 16 node/routers and 26 links. The capacity of the links is either 150 Mbps or 600 Mbps, as shown in Figure 4.5. The overall network capacity is 6150 Mbps and can be used as comparison for the total traffic offered by best-effort flows.

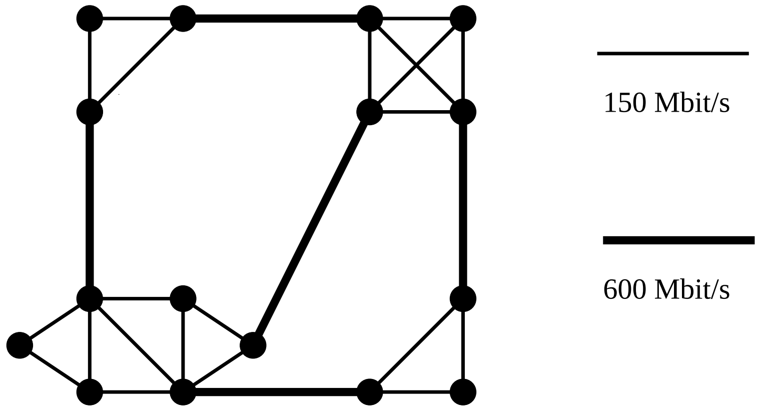


Figure 4.5: The SCT network topology

To highlight the effect of outdated link information we use a scenario where the bandwidth that can be used by the best-effort traffic varies in time. Two types of traffic generators are present in the network: guaranteed CBR and best-effort. One generator of each type is connected to any network node. Traffic from both generator types is directed towards sinks that are located in every node of the network. We consider a uniform traffic scenario.

CBR flows request 5 Mbps and their durations are set through i.i.d exponential random variables, with mean value equal to 120 s. CBR flows are always routed using a fixed shortest-path algorithm with load balancing as described in [71], i.e., balancing the load between equivalent paths instead of choosing among them at random. CBR flows are subject to a simple peak-rate admission control: they can reserve bandwidth and are served with priority.

Best-effort generators activate flows whose traffic amount to transfer is exponential with mean value 50 Mbyte. Each flow is bounded by B_M of 10 Mbps and by b_m of 1 Mbps. This part of the traffic realises the elastic behaviour presented in Section 4.2.3.2.

In the traffic scenarios 10% of each link capacity is reserved to best-effort traffic; this means that CAC procedures on CBR flows consider 90% of the link capacity as the available bandwidth. The average CBR load is roughly 150 Mbps, leading to a high variability of available bandwidth for best-effort flows.

We consider periods ranging from a supposedly instantaneous distribution of metrics (i.e., $t_u = 0\,\mathrm{s}$) to periods that are of the same order of magnitude as the average of minimum duration of flows in the network (i.e., $t_u = 100\,\mathrm{s}$). The rationale beyond this choice is that, if update periods and flow durations are comparable, nodes are less and less dependable upon to have an accurate representation of the state of the network. Conceivably, this induces less-than-optimum choices when the routing path is selected. We set f_{ina} to 5% in order to model the inaccurate estimation of the current flow bandwidth B^f.

4.3.3.2 Performance evaluation

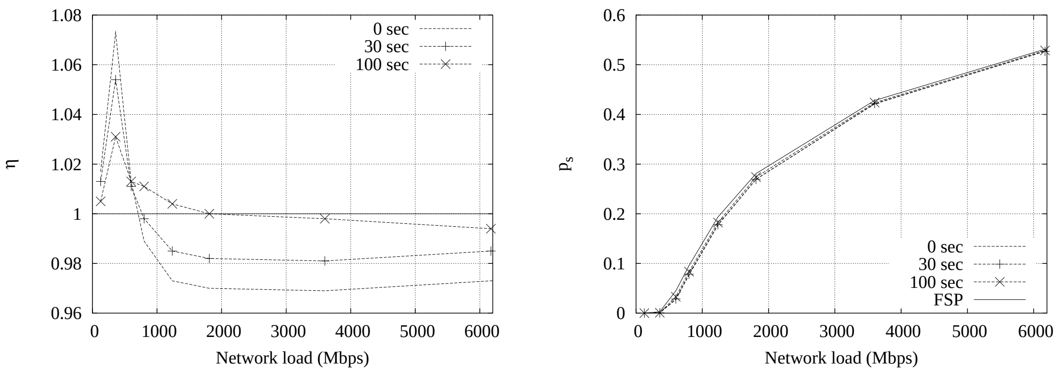


Figure 4.6: Relative gain η (left) and starvation probability (right) of the Widest-Shortest algorithm for different update periods

The plots of Figures 4.6, 4.7 and 4.8 depict results for the $\eta(WS)$, $\eta(MD)$, $\eta(LD)$ gains on the left hand side and the corresponding starvation probability values on the right, with different t_u update periods. Indeed, the results with *FSP* algorithm using a weighting metrics based on time-invariant topological information are not affected by the outdated link state information, but the adaptive algorithms suffer from this effect. As it could be expected, the lower is the frequency, the lower becomes the mean throughput of the connections that terminate ordinary. The gain of QoS algorithms in the low-load region vanishes almost completely if the update period is comparable to connection duration.

Indeed, when updates are fewer and farther between, adaptive algorithms start mis-

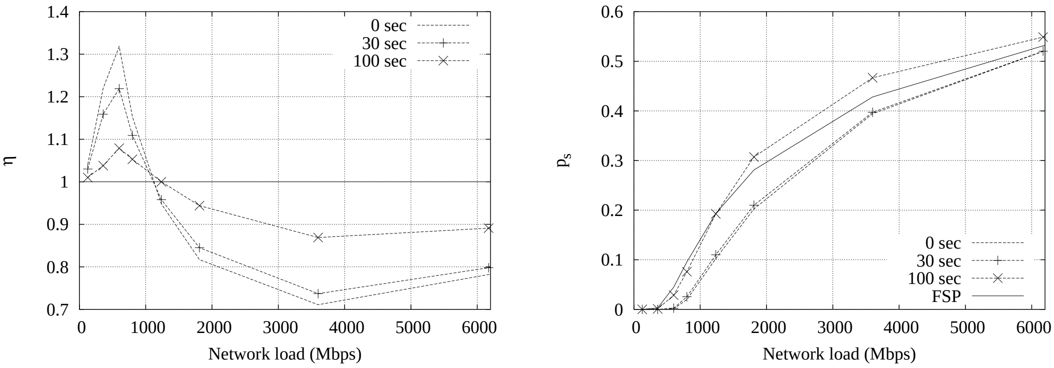


Figure 4.7: Relative gain η (left) and starvation probability (right) of the Minimum-Distance algorithm for different update periods

judging the available bandwidth. Therefore, flows may be routed neglecting the less-congested paths. In the low-load region this leads to the above mentioned loss of assigned bandwidth gain. However, when the load is high every route are quite congested and the misjudging can lead to use rather the shorter paths, thus tending to the *FSP* choice. This effect can be observed for the *WS* and *MD* algorithm when the load is higher than 800 and 1100 Mbps respectively: the impact of growing t_u turns to the opposite and we achieve even better performance from the throughput point of view.

On the other hand, we have to take in consideration the values presented on the right plots. The update period does not affect strongly the starvation probability except for the *MD* algorithm with large t_u. In the other cases the p_s value of the QoS algorithms grows with the update period but still remains under the starvation probability of *FSP*.

We analysed as well the impact on the performance if we vary the f_{ina} inaccuracy factor, but we found that its effect is marginal with respect to that of the update period. Thus, the value of 5% is used in each further simulation.

4.4 Novel QoS routing strategies

In the previous sections some results on routing performance were presented and in many cases the advantages of QoS routing algorithms were obvious. In general, one can ob-

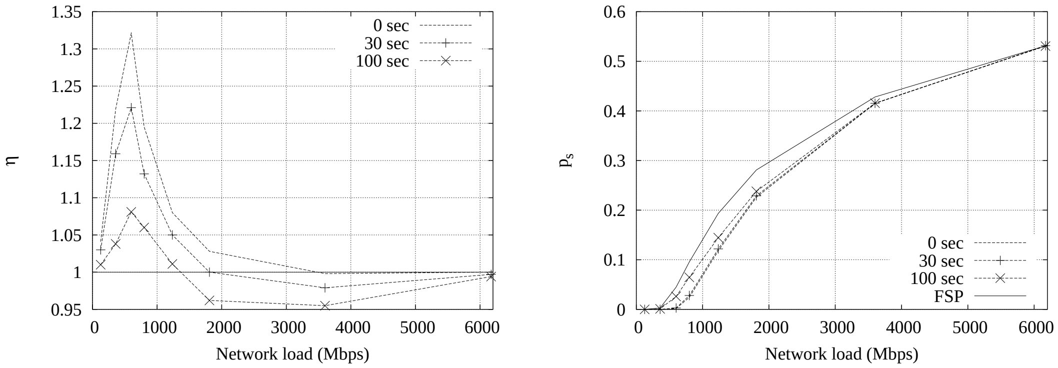


Figure 4.8: Relative gain η (left) and starvation probability (right) of the Load Dependent algorithm for different update periods

serve a gain in light load regions that becomes a loss when the network load grows significantly. Additional drawbacks are the time-consuming calculations based on the network state and possible wrong decisions when the information is out of date.

Our aim was to provide QoS algorithms that are robust in two ways, i.e., they tolerate high network loads and link-state information instability, while working with moderate computing complexity. We developed and analysed two groups of novel algorithms: *MSF* and *NGR* presented in [40] and [23, 66] respectively.

4.4.1 Multimetric Sequential Filtering algorithms

The routing algorithms introduced in this section are based on the notion of *Multimetric Sequential Filtering* (*MSF*).

Any *MSF* algorithm operates through a number of steps in the route selection task. At each step, the set of feasible paths from the source to the destination is ordered according to a metric, and the paths in the bottom half of the ordered set are dropped. The remaining paths are then re-ordered according to a secondary metric, again dropping the bottom half. The filtering process is repeated a number of times until the path where flows will be routed is picked out of a small set.

The filtering rules, and the order in which they are applied, are the key issues of the algorithm. Since one of the goals of the algorithm is to reduce the processing load of a router, even if considering multiple metrics, larger path sets should be filtered less frequently, while smaller sets could be recomputed in a shorter time frame. Consequently, the filtering rule applied to the *whole* set of feasible paths from a source to a destination should be a ‘topological’ rule, such as the number of hops, or the physical link capacity. Path ordering and filtering according to the hop count is bound to hold unless the topology changes and so does a physical capacity ordering as well. When other ‘volatile’ metrics are taken into account, such as the above defined RES or ABW, a more frequent reordering, reflecting changes in the state of the network, is necessary. For this reason, filtering based upon available bandwidth should be done *after* the path set has already been reduced by previous filtering.

Using topology-based filtering first and bandwidth-based at a later stage also reduces the risk of routing flows on too long paths because of a marginally larger available bandwidth.

4.4.2 Formal description

The proposed algorithms employ three ordering rules according to the weighting metrics and path costs introduced in Section 4.2.4:

H applies the cost c^H

RES applies the cost

$$c^{RES}(\pi) = \frac{1}{\min_{e_i \in \pi} w_i^{RES}} \tag{4.5}$$

ABW applies the cost c^{ABW}

Using these ordering rules let $\mathcal{P}_{\mathcal{G}}^{\gamma}(s,d)$ be equal to the set $\mathcal{P}_{\mathcal{G}}(s,d)$ ordered by the cost metric c^{γ}. We introduce the function $\chi(n)$, $n \in \mathbf{N}$. It returns an even number computed as the quotient of $n/2$, without remainder, and incremented by one. If $n \leq 2$, it returns 1. More precisely:

$$\chi(n) = \begin{cases} 2 * \lceil n/4 \rceil & \text{if } n > 2 \\ 1 & \text{otherwise} \end{cases} \tag{4.6}$$

The $\chi(\cdot)$ function will be used in determining the number of paths to drop from the ordered set before applying a different filtering. Let $\Xi(\mathcal{P}^\gamma)$, be the filtering function of the ordered set of paths $\mathcal{P}^\gamma$:

$$\Xi(\mathcal{P}^\gamma) = \{\pi_k | \pi_k \in \mathcal{P}^\gamma,\ k \leq \chi(||\mathcal{P}^\gamma||)\} \tag{4.7}$$

where π_k is the k^{th} element of ordered set $\mathcal{P}^\gamma$. The returned set $\mathcal{Q} = \Xi(\mathcal{P}^\gamma)$ is *unordered*. An *MSF* algorithm performs a sequence of combined steps that include both filtering and ordering by different cost metrics. Due to the included filtering, this method differs significantly from the simple concatenation of cost metrics introduced before.

To model the routing functions working in a network that considers the periodic distribution of link state information we need to use the link weights $RES(t_u)$ and $ABW(t_u, f_{ina})$ instead of RES and ABW respectively.

Using these formalisms, we give now the description of four different flavours of Multimetric Sequential Filtering algorithms.

4.4.2.1 Algorithm MSF1

Given a set of paths $\mathcal{P}_\mathcal{G}(s,d)$ with m_0 paths between a source s and a destination d in graph $\mathcal{G}$:

1. Considering the hop-ordered set $\mathcal{P}^H_\mathcal{G}(s,d)$, let us first restrict our scope to the first m_1 paths; thus, we will define:

$$\mathcal{Q}_\mathcal{G}(s,d) = \{\pi_k | \pi_k \in \mathcal{P}^H_\mathcal{G}(s,d),\ k \leq m_1\} \tag{4.8}$$

   (m_1 is just an upper bound to the initial number of paths in the set: in the simulations we used $m_1 = 32$).

2. $\mathcal{R}_\mathcal{G}(s,d) = \Xi(\mathcal{Q}^H_\mathcal{G}(s,d))$

   given the m_1 paths in $\mathcal{Q}^H_\mathcal{G}(s,d)$, let us select the first $m_2 = \chi(m_1)$ and assign them to the unordered set $\mathcal{R}_\mathcal{G}(s,d)$.

3. $\mathcal{S}_{\mathcal{G}}(s,d) = \Xi(\mathcal{R}_{\mathcal{G}}^{RES}(s,d))$

   given the m_2 paths in the residual-bandwidth ordered set $\mathcal{R}_{\mathcal{G}}^{RES}(s,d)$, let us select the first $m_3 = \chi(m_2)$ and assign them to the unordered set $\mathcal{S}_{\mathcal{G}}(s,d)$.

4. Best-effort flows between s and d will be routed over the path π in set $\mathcal{S}_{\mathcal{G}}(s,d)$ that is the "lightest" by the cost metric ABW. In other words, the choice is the first path of the ordered set $\mathcal{S}_{\mathcal{G}}^{ABW}(s,d)$.

As discussed above, steps 1, 2 and, if only one traffic class is present then even step 3, of the MSF1 algorithm can be executed off-line. In step 4 the selection of the path that has the largest amount of bandwidth available to a single flow must be performed on-line.

4.4.2.2 Algorithm MSF2

Version 2 of the *MSF* routing algorithm further filters paths according to the hop count. As a result, steps 1 through to 3 are the same as in MSF1. The further steps are:

4. $\mathcal{T}_{\mathcal{G}}(s,d) = \Xi(\mathcal{S}_{\mathcal{G}}^{ABW}(s,d))$

   given the m_3 paths in the available-bandwidth ordered set $\mathcal{S}_{\mathcal{G}}^{ABW}(s,d)$, let us select the first $m_4 = \chi(m_3)$ and assign them to the unordered set $\mathcal{T}_{\mathcal{G}}(s,d)$.

5. Best-effort flows between s and d will be routed over the path π in set $\mathcal{T}_{\mathcal{G}}(s,d)$ that is the "lightest" by the cost metric H, i.e., the choice is the first path of the ordered set $\mathcal{T}_{\mathcal{G}}^{H}(s,d)$.

4.4.2.3 Algorithm MSF3

Version 3 of the *MSF* routing essentially captures the spirit of the *Load-Dependent* (*LD*) algorithm that was introduced in [22] and outlined in Section 4.2.4. The algorithm chooses the same π path as MSF1 if the available bandwidth on it is large enough for a new connection, i.e., if $\frac{1}{c^{ABW}(\pi)} < kB_M$, and the *FSP* choice otherwise. However, note that this solution is not identical to the *LD* algorithm.

4.4.2.4 Algorithm MSF4

The last version is a Load-Dependent extension of the MSF2 algorithm: specifically, the comparison with the maximal bandwidth B_M is carried out after algorithm MSF2 has selected a path π, and the routing decision is taken following the guidelines given for algorithm MSF3 too.

4.4.2.5 Performance evaluation results

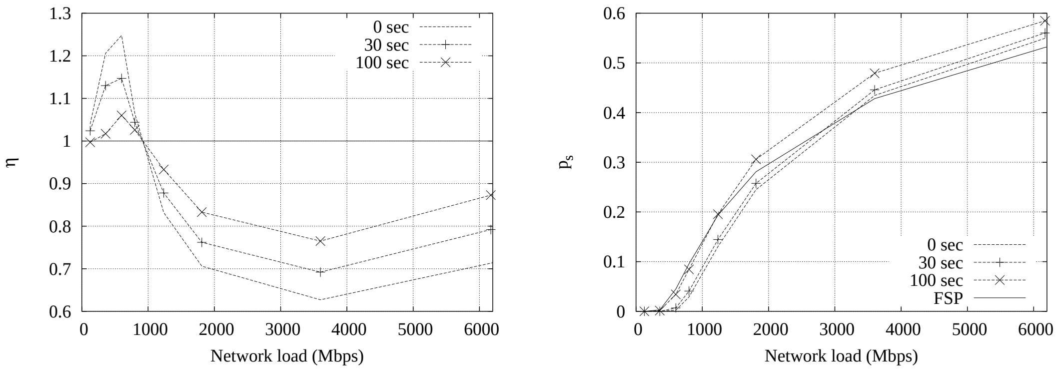


Figure 4.9: Relative gain η (left) and starvation probability (right) of the MSF1 algorithm for different update periods

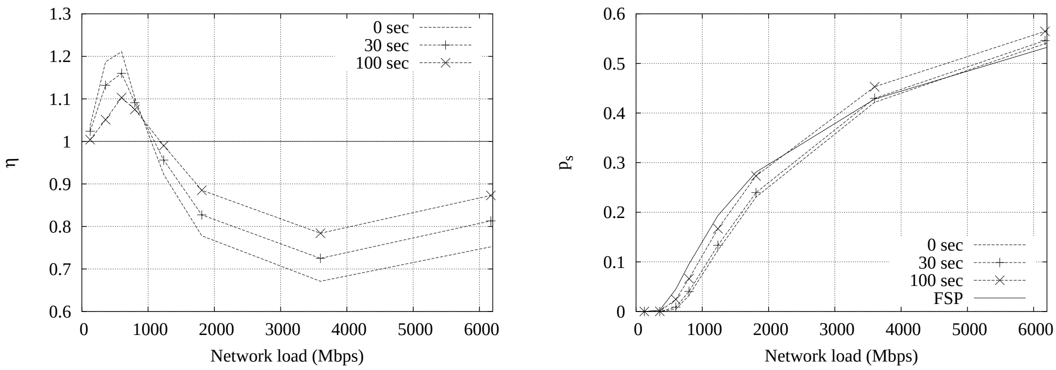


Figure 4.10: Relative gain η (left) and starvation probability (right) of the MSF2 algorithm for different update periods

To observe the robustness of the Multimetric Sequential Filtering algorithms from the link state update frequency point of view, we considered the same network topology and traffic scenario as in Section 4.3.3. Regarding Figures 4.9 and 4.10 we can state that MSF1 resists hardly to the "bad" effects of the large update period, while MSF2 suits better to our aim and looses less of its low-load gain, even if it does not stand to our expectations. When the offered load grows over about 1000 Mbps η starts to decrease strongly and in higher load regions even the starvation probability of MSF1 and MSF2 algorithms is larger than that of *FSP*.

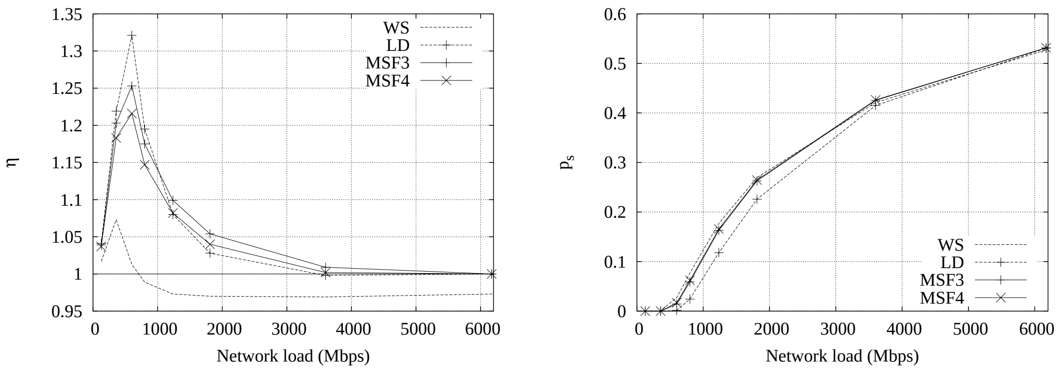


Figure 4.11: Relative gain η (left) and starvation probability (right) for continuously update

In the second set of results we present a comparison of the relative per connection throughput gain and the starvation probability achieved by algorithms MSF3 and MSF4 to those achieved by *WS* and *LD*. Figures 4.11 and 4.12 present an overview of the gains and probabilities for $t_u = 0$ s and $t_u = 100$ s respectively.

As we can see, the adaptive algorithms on the one hand outperform *FSP* and the *MSF* ones when the traffic load is low and the link state information is distributed continuously or with high frequency. On the other hand MSF3 and MSF4 do not loose to much of their performance even if the information are wide out of date. As we discussed earlier, *MSF* algorithms inherently lessen the impact of outdated information by restricting their scope to a limited set of paths, namely those with a smaller hop count and a larger capacity. It is very important to stress that algorithms like MSF3 and MSF4 do not perform worse

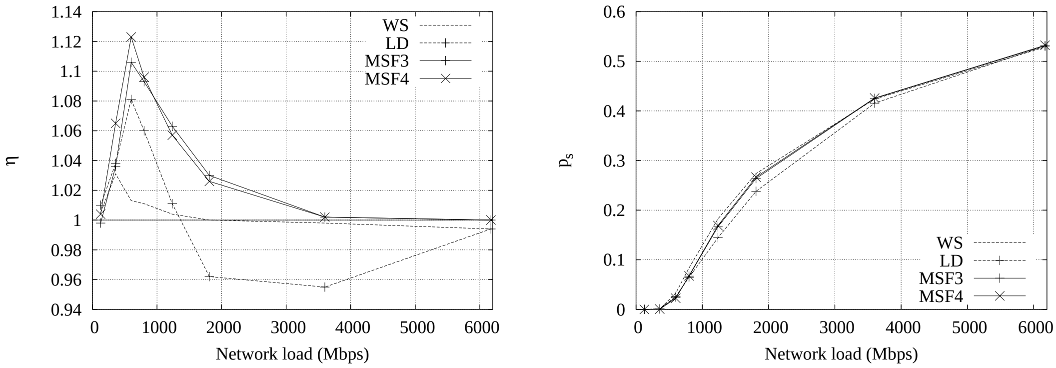


Figure 4.12: Relative gain η (left) and starvation probability (right) for 100 s update period

than *FSP*, even when t_u is comparable with the flow average minimal duration.

Simulation results show that some *MSF* routing algorithms can be resilient to out-of-date link state information, while offering a non marginal gain when the update is reasonably frequent compared to the average duration of flows and the network load is moderate.

4.4.3 Routing based on Network Graph Reduction

We introduce a new approach to QoS routings that can be summarised as "Network Graph Reduction". This method performs a modification of the graph describing the network before the routing path is computed, in order to exclude from the path selection the over-congested portions of the network. This solution leads to a class of two-step routing algorithms, where both steps are simple, hence allowing efficient implementation.

4.4.3.1 Motivation

The cost function at the core of a QoS routing algorithms tries to find portions of the network where resources are under-utilised and exploits them to the benefit of flows that would otherwise cross a congested portion of the network. As stated many times before, a major drawback of all such algorithms is that in case of heavy congestion, QoS routing ends up to waste resources and performs poorly even if compared with *FSP*.

We stress at this point that the poor performance of these routing algorithms at high loads is not due to the sub-optimality of the used algorithms, but is rooted in the locality of the routing decision. Whenever a flow is routed, the given cost function is minimised for the current state of the network, necessarily disregarding the future network evolution. Under heavy load, the overall network benefit does not coincide with the cost function of a single route, hence the minimisation of the one does not lead to the maximisation of the other.

As it appears from the results presented in Section 4.2.5.3 the adaptive algorithm *LD* performs quite well in both low and high load case. The conditional decision in this routing method tries to model a choice of route that is dependent on network load: when the network seems to be overloaded, the algorithm chooses the *FSP*. Since the network load ρ is typically not known to the routing algorithm, not even in the case of centralised schemes, the identification of the high load zone remains the main problem. Indeed, simulation results show that the performance of the *LD* algorithm depends strongly on the value of the factor k used during the route selection procedure. The most suitable value of k can be derived only with a wide analysis of the given topology and traffic.

To eliminate these problems, we may look at the routing problem from a slightly modified perspective. Until now a QoS routing algorithm was a choice of a suitable path considering a series of $c^{\gamma}(\cdot)$ cost function. Now, instead of trying to find new adaptive metrics, the same metric that works well for light loads can be applied at high loads to a *reduced graph* that contains only uncongested links, i.e., links whose load is under a given threshold. Since the reduced graph may not be connected, i.e., no path may exist from a source to a destination, the *FSP* should always be included as possible solution of the routing problem. Thus, as the offered load ρ grows, all links become congested and the algorithm necessarily chooses the minimum-hop path. While the direct measure (or evaluation) of ρ is a complex task, the identification of single congested links is extremely easy, so that, overall, the methodology is simple and its implementation straightforward. This key idea leads to the definition of the Network Graph Reduction (*NGR*) strategy.

4.4.3.2 Formal description

NGR can be applied to any existing QoS routing algorithms. It operates two subsequent alterations upon the path set $\mathcal{P}_{\mathcal{G}}$, and applies selection criteria on the resulting subset.

First, it transform the directed graph $\mathcal{G}$ into $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$, $\mathcal{E}' \subseteq \mathcal{E}$, selecting only those links whose weight satisfies a *cut-off criterion* $\mathcal{C}$: only links with non-zero free capacity are acceptable, i.e., edge e_i has to be cut off, if

$$\sum_{f=1}^{N_i^{BE}} B^f = B^{RES}(e_i) \tag{4.9}$$

As a result, the path set $\mathcal{P}_{\mathcal{G}}$ is reduced to $\mathcal{P}_{\mathcal{G}'}$ after removing those paths containing cut-off links.

Second, let $\mathcal{F}_{\mathcal{G}}$ be the set including the *FSP* path from each source to each destination. The second alteration transforms the path set $\mathcal{P}_{\mathcal{G}'}$ into $\mathcal{P}'_{\mathcal{G}'} = \mathcal{P}_{\mathcal{G}'} \bigcup \mathcal{F}_{\mathcal{G}}$. The new subset is built from i) the paths belonging to $\mathcal{P}_{\mathcal{G}'}$ and ii) the *FSP* paths between any two nodes in $\mathcal{V}$, *even if their links were discarded in the first step*.

Finally, the decision on how to route packet flow between source s and destination d is made by applying the order relation $\prec$ to the subset $\mathcal{P}'_{\mathcal{G}'}(s, d)$ and picking the path that turns out to be the *lightest* according to the cost metric defined by the original QoS routing algorithm.

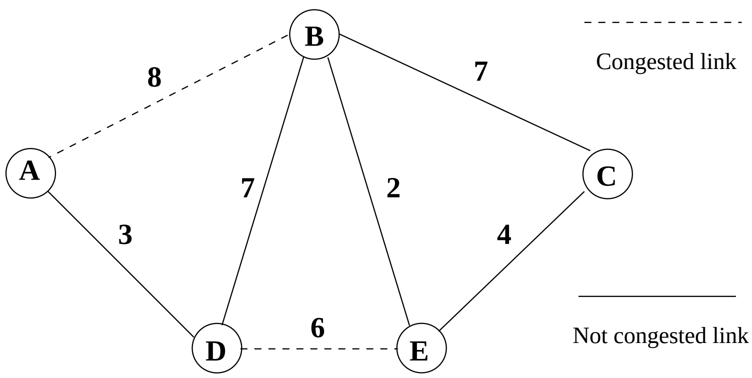


Figure 4.13: Network topology for the illustration of *NGR*

Let us show a small illustration considering the network depicted on Figure 4.13 and using the indicated instantaneous link costs and an additive route cost function. We consider link AB and DE to be congested. For source A and destination C the minimum cost path is selected from the set of available paths $\{ADEC$, $ABEC$, ABC, $ADBEC$,

$ADBC$, $ADEBC$, $ABDEC\}$ in the classical algorithm case, while in the *NGR* case this set is reduced to $\{ABC$, $ADBEC$, $ADBC\}$. Thus, the classical route selection takes the path $ADEC$ while introducing *NGR* we get ABC that is indeed the shortest path.

Combining *NGR* with either the *MD* or *WS* algorithms introduced in Section 4.2.4, we obtain the *NGR-MD* and *NGR-WS* algorithms.

4.4.3.3 Implementation issues

After defining the algorithms, we are interested in the viability of a *consistent* hop-by-hop implementation, in which routing decisions taken at an upstream node are not liable to be superseded by downstream nodes finding a locally-optimal alternative to the path chosen earlier by upstream nodes. Unfortunately, if we combine the *NGR* methodology with the *MD* and *WS* algorithms, the hop-by-hop implementation property is not preserved. The proof in case of the *NGR-MD* metric is given by counter-example in [66], while *NGR-WS* cannot be considered since the *WS* criterion itself is not implementable hop-by-hop as proven in [72].

In the following, we will refer to a *consistent* version of the *NGR* algorithms if routing decisions taken at upstream nodes are not overruled by downstream nodes, while we will refer to a *hop-by-hop* version of *NGR* (*NGR HbH*) if every node implements its own locally-optimal version of the algorithm.

As pointed out above, for consistent routing to be preserved, the forwarding procedures of the *NGR* algorithm must be integrated so as to signal the routing decision taken at upstream nodes. This solution can be implemented using the route-pinning property of the MPLS technology [62], or a slight modification of the IP forwarding procedure as it is described more detailed in [23].

The asymptotic computational complexity of the *NGR* algorithms is the same as the complexity of the QoS algorithms without graph reduction.

4.4.3.4 Algorithm performance on random topologies

The algorithm described in Section 4.4.3.2 necessarily behaves as a *MD* or *WS* algorithm for non congested networks and as a *FSP* algorithm as the offered load increases causing congestion; however, the behaviour at intermediate loads can only be investigated by simulation. For the sake of comparison, result reports include the classic *MD*, *WS* algorithms and the *FSP* algorithms.

To go in more details on the elementary mechanism of the *NGR* algorithms the reader is referred to the study of a trivial network published in [23]. Those results show how this method tries indeed to avoid the overloaded resources in the route selection process.

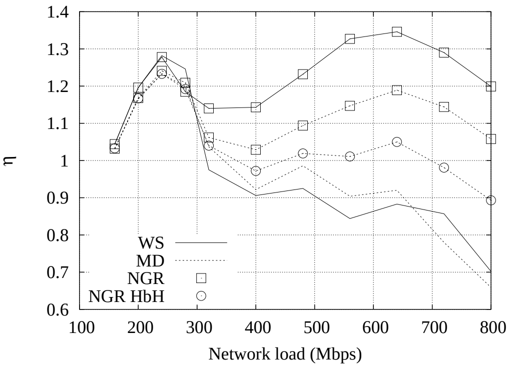
As done in Section 4.2.5.1 we present results obtained on randomly-generated network topologies using the same topology and link parameters. The *DB* traffic in this case is generated in each node with 1 Mbps as B_M and 50 kbps as b_m, while the mean size of bulk data to transfer is 2.5 Mbytes. Each node adopts a fixed update period for the QoS metric, set to 30 seconds, similarly to the update period used in OSPF.

We report results for two networks, named (A) and (B) for short. Results for network (A) reflect a simple stationary traffic scenario. A uniform traffic pattern is simulated, i.e., when a new connection request is generated, the source and the destination are randomly chosen. Each sources have the same intensity of generating a new connection. These intensities are stationary, i.e., they do not change during the simulation.

On the contrary, results reported for network (B) refer to a more complex traffic pattern, mimicking a "client-server" scenario. Besides, the traffic matrix changes over time. In more detail, among the 32 nodes, 5 nodes, picked among those with the highest number of outgoing links, are defined as "servers", while all the remaining one are defined as "clients". Each server node generates 10 times the traffic generated by a client, with a non-stationary behaviour: it randomly cycles over time among two states, high traffic and low traffic. When it is in the low-traffic state, it reduces its offered load by a factor of 10, i.e., it behaves as a client node. The average time spent in either the high- or low-traffic state is 3 hours, following an exponential distribution.

We evaluate the performance indices introduced before: the relative gain of average throughput η and the starvation probability p_s.

Let us first present the results achieved in the scenario with network (A). The left plot of Figure 4.14 reports the relative gain of average throughput. The advantage of QoS routing algorithms over *FSP* at low loads is clear, but while the classic *MD* and *WS* implementations fall below the *FSP* at high loads, *NGR* algorithms always remains above *FSP*. It is clear that the *MD* and *WS* algorithms are extremely sensitive to the network load, while the *NGR* versions perform better than *FSP*, converging to its performance as ρ



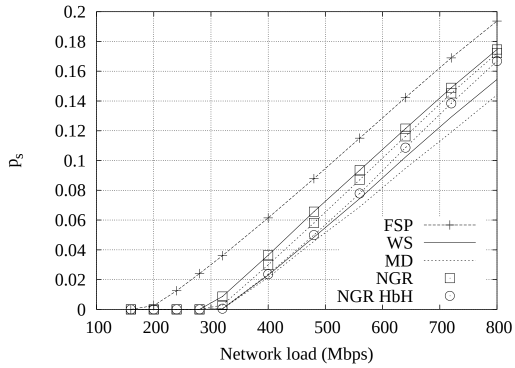


Figure 4.14: Relative gain η (left) and starvation probability (right) on network (A)

grows larger. Interestingly, the hop-by-hop implementation of *NGR* routing (*NGR HbH*) feature a performance midway between the *consistent* version and the classic algorithms, thus confirming that locally-optimal decisions are affected by performance degradation that limit the benefits of the *NGR* approach.

In addition, the *NGR-WS* algorithm performs better than the *NGR-MD*, since it chooses from the path set which includes only minimum-hop ones. The non-monotonic relative behaviour of the algorithms is due to the randomness of the topology that is neither regular nor a “well defined” hierarchical topology, which causes different portion of the network to become congested depending on the values of offered load, and on the specific QoS routing algorithm implemented.

One might wonder whether connections completing their information transfer obtain a larger throughput simply thanks to the larger number of starved connections shutting down thus leaving more for others. On the righthand plot of Figure 4.14 *NGR* algorithms show a starvation probability that is between *FSP* (the highest) and the classic *MD* and *WS* algorithms. In more detail, the starvation probability is negligible for all the QoS algorithms for values of the offered traffic smaller than about 300 Mbps, while the *FSP* algorithm shows an earlier rise in the starvation probability, at about 200 Mbps of offered traffic.

In the case of the client-server model with non-stationary traffic of network (B), it is more interesting to separately look at client’s and server’s behaviour, especially in view

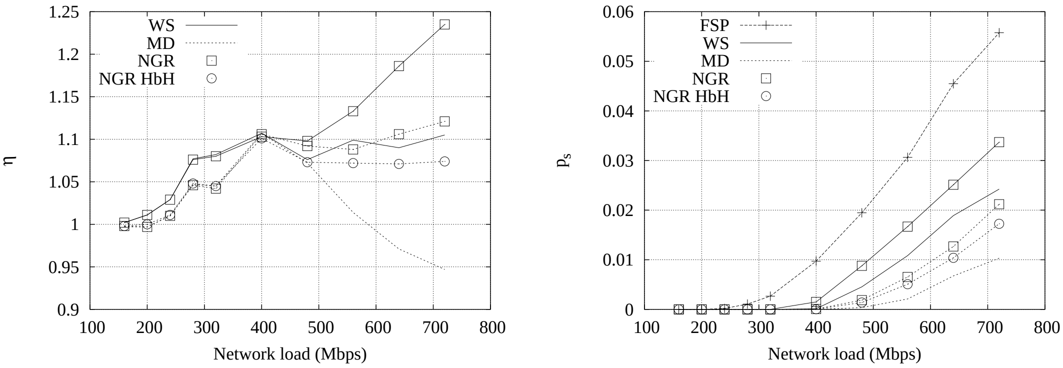


Figure 4.15: Relative gain η (left) and starvation probability (right) of the clients in network (B)

of the high level of traffic generated by the servers during their active cycles. Figures 4.15 and 4.16 show the performance of clients and servers. The left plots refer to the relative gain of the average throughput while the right plots refer to the starvation probability. Also in this case the *NGR* version of the algorithms outperforms the classic QoS routing. The more complex nature of the traffic pattern which keeps changing during time, reduces the gain observed for the client connections to less than 20%.

On the contrary, the performances obtained by server nodes are greatly improved by any QoS routing for relatively low load, with the classic *MD* showing a clear edge. But as soon as the total offered load of the network increases over 500 Mbps, their performance worsen, while the *NGR* versions still exhibit a better utilisation of the network capacity. The starvation probability confirms results discussed about network (A). It shows that the use of QoS routing greatly reduces the starvation probability, reflecting a better distribution of the traffic on the network. Also in this case, the *NGR-MD* and *NGR-WS* algorithms exhibit a slightly higher starvation probability.

4.5 Discussion

The first issue we addressed in this chapter was the impact of the applied traffic model on the performance in the *data layer*. Since no analytical tools are available for the

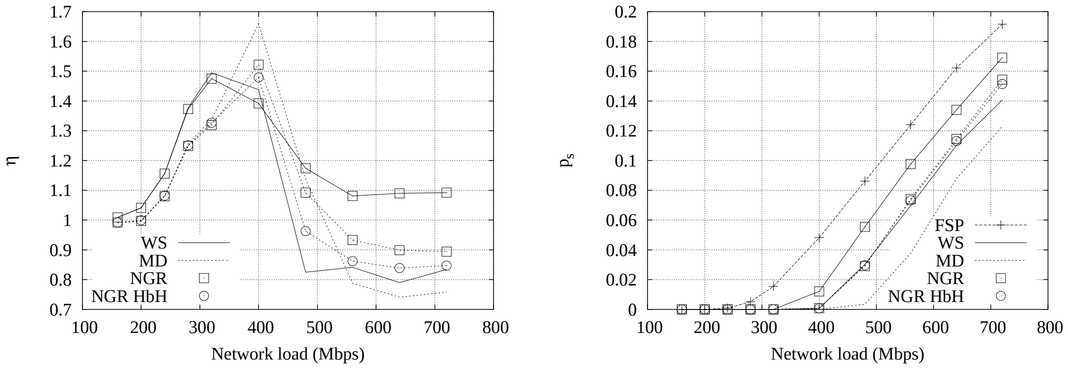


Figure 4.16: Relative gain η (left) and starvation probability (right) of the servers in network (B)

evaluation task and the simulation is very complex when using packet-based concept, we investigated a realistic flow-based traffic model. It is based on three main characteristics of IP traffic:

- the holding time of Internet connections is based on the amount of data to be transferred and hence depends on network conditions,
- the elastic nature of the Internet traffic implies a closed loop between network and sources: the connection completion time depends on the instantaneous load of the network and, hence, cannot be computed "a-priori" when the connection is generated,
- the Internet does not provide CAC functions, but connections are generally closed when the perceived quality is too poor.

We formalised the routing problem and compared some previously introduced solutions. The presented results show that the new performance evaluation methodology can help to understand the behaviour of the different QoS routing algorithms. Besides, comparison with a more traditional evaluation method, where connection are time-limited, though elastic, highlights significant differences, hinting that traditional models might fail to grab the relevant characteristics of the current data networks.

As second issue we analysed the sensitivity of QoS routing algorithms to stale link state information. A very natural assumption in IP networks is that the information on resources can not be updated continuously in each node of the network because, on the one hand, the control traffic grows. On the other hand, the continuous measurements of link state parameters can be complicated. Thus, the adaptive routing decisions may be based on inaccurate values. We investigated the problem of how the information update period influences the performance of adaptive routing algorithms.

Based on the experiences with the traditional ones, we proposed two new classes of QoS-based routing algorithms. Both of them are robust and combine good performance with a reduced computational complexity:

- Multimetric Sequential Filtering (MSF) realises a multi–step selection of the suitable route. In each step a different type of cost can be considered as ordering basis and the set of the available paths is halved.

- Network Graph Reduction (NGR) can be applied with any routing algorithms, but the selection is performed on a special path set. This set consist of paths in the reduced network graph that contains only uncongested links plus the *FSP* path itself. This concept implicitly realises a load dependent solution.

Chapter 5

Analysis of dynamic grooming

5.1 Introduction

In the previous chapters of the dissertation we analysed the *data layer* and the *optical layer* in a separated manner. Let us now turn to the complex problem of dynamic grooming, i.e., the interaction of these layers. The aim of this cooperation is the effective usage of network resources in both layers. While the customers of the *data layer* want a higher quality in the transmission of their data, the service provider that manages the *optical layer* has different aims. Operators want to provide services without wasting too much of network capacity and leave free as many resources for further customers as possible. Traffic grooming helps to resolve this conflict [73]. The goal of using dynamic algorithms is to adapt the virtual topology of the *data layer* to the changes of user traffic. Several solutions have been developed recently, but there are still many open question in this research area.

As outlined in Section 2.6.4 the application of a theoretical approach for getting insight in the behaviour of IP over WDM networks meets several difficulties. However, we can extend the results obtained for the optical layer when flow requests that arrive at the *data layer* are assumed to require guaranteed bandwidth. With respect to the networks that allow *subwave channels,* Srinivasan et al. introduce a model in [30], that can provide results considering even inhomogeneous converting capabilities in the nodes. However, some constraints are still present and the calculation is slow and complex. In [74] Xin et al. presented an accurate and robust model of traffic grooming analysis considering alternative paths and first fit wavelength assignment. The complexity is quite high also in these cases due to the recursive steps of the calculation. In [75] we analysed the problem by reducing it to a model that was introduced for the analysis of all-optical networks. This issue is addressed in Section 5.2.

On the other hand, the rather interesting problem is the analysis of the interaction in the IP scenario, i.e., when the *data layer* is modelled to catch the special characteristics of Internet traffic. Many grooming algorithms were proposed in recent years, see [4, 28, 76, 77] to cite just a few, and compared one another, or simply against a standard wavelength routed network where no grooming is performed. Some works assume static grooming [77, 78], and generally tackle the problem with same optimisation technique, while others assume that grooming is dynamic [5, 28, 79, 80]. All these works, however, simply disregard the elastic nature of TCP/IP traffic: IP over WDM is indeed modelled like a traditional circuit switched traffic with bandwidth guarantees.

Regarding the results of Section 4.2 we can state that the analysis with a too simple traffic model may lead to wrong conclusions, since the choice of the model affects seriously the performance measures. Another weak point of many analysing works is ignoring the dynamic interaction of the layers that implies ignoring also some extraordinary effects. Our studies presented in [32, 81, 82] instead, give deeper insight in how grooming schemes may work in realistic IP over WDM networks. Section 5.3 addresses the problem of considering elastic traffic in dynamic grooming issues.

5.2 Grooming guaranteed traffic

Here we show how the grooming analysis can be reduced to a problem that can be studied with a theoretical model based on **WDMM** introduced in Section 3.2. To perform the reduction we need some specific assumptions on the *data layer* traffic and the adaptation of the model to handle this traffic.

5.2.1 Considered traffic model and grooming scheme

Although it is not well suited to characteristics of IP traffic we consider at the *data layer* constant bitrate flow requests with bandwidth guarantees. We assume homogeneity in the required bandwidth, i.e., the bitrate is identical for all the requests. For the arrival process and for the holding time of flows we use the same assumptions as taken in Section 3.2 for the optical connection requests.

Allowing time division multiplexing inside the optical channels, we can define channels, called *subwave channels* that correspond to the timeslots. A data transmission channel consists of a series of *subwave channels*. The connection of *subwave channels* can be restricted by several rules derived from the wavelength conversion and timeslot reordering capability of the OXCs.

As a flow request arrives the generalised routing task is performed in both the *data layer* and the *optical layer*. In the latter we apply fixed routing and random wavelength assignment described in Section 3.2. Considering the concept of *subwave channels* the routing in the virtual layer is very simple: only one hop routes are allowed, i.e., a direct transmission channel transmits the flow data. This channel is dedicated to one flow request and it is set up and torn down at the arrival of the request and at the finishing of the data transmit of the flow respectively.

Note, that such a routing solution can be realised only when the dynamic grooming scheme supports the full cooperation of the two layers, i.e., it has a peer architecture. We refer the reader to [6, 28, 29] for more about the grooming architectures of different level of cooperation.

We assume that network nodes are homogenous from point of view of the wavelength conversion and the timeslot reordering capability. Moreover, they apply the same guidelines in connecting *subwave channels*. According to the assumptions in Section 3.2 each fiber supports the same number of wavelengths and each optical channel on each fiber is divided the same way into *subwave channels*. Thus, a set of trunks can be constructed out of these channels identified by their wavelength, or timeslot number, or both according to the node capabilities. Only *subwave channels* of the trunks with identical identifier can be connected in a transmission channel.

We apply a very simple connection admission control mechanism. Since users require bandwidth guarantees, if no suitable path is found, i.e., no direct channel can be set up, the request is blocked. As main performance measure we find the blocking probability defined as the number of blocked requests divided by the number of all requests.

5.2.2 Description of the model

Let us present the extension of the **WDMM** to a *Homogenous Guaranteed-traffic Grooming Model*, **HGGM**. The algorithm of the blocking probability computation does not change, but some parameters are interpreted in a different way, according to the concept of *subwave channels*.

Let us consider b_h the bitrate of the connections to be the bandwidth of a transmission channel. The number of *subwave channels* in an optical channel equals $U = \lfloor C_O/b_h \rfloor$, where C_O is the bandwidth of one whole wavelength. On each links the number of *subwave channel* trunks T_{SO} is equal to:

- C, when only timeslot reordering is allowed,
- U, when only wavelength conversion is allowed,
- 1, when both these functions are enabled,
- $C \cdot U$, when none of them are allowed.

Each step of the analysis can be performed using *subwave channel* trunks instead of wavelength trunks. In the computations we have to use the value T_{SO} instead of C and the values C_j/T_{SO} instead of the values M_j.

5.2.3 Numerical results

To evaluate the accuracy of the extended model we compared its results to simulation by reporting request blocking probability in function of network load. The simulation model had to be extended to consider the grooming scheme and the concept of *subwave channels*. The trunk interpretation was changed according to the extension of the theoretical model.

We analysed the impact of traffic granularity on blocking in regular and irregular topology scenarios similarly as in Section 3.2.3. Here we show the results network CWEN presented on Figure 3.5. The link capacity values has to be understood this time in 100 Mbps, e.g., 64 means 6.4 Gbps. Each fiber contains 4 wavelengths of 0.8 Gbps capacity. We consider neither wavelength conversion nor reordering of timeslots and we use the same traffic relation matrix as in Section 3.2.3.

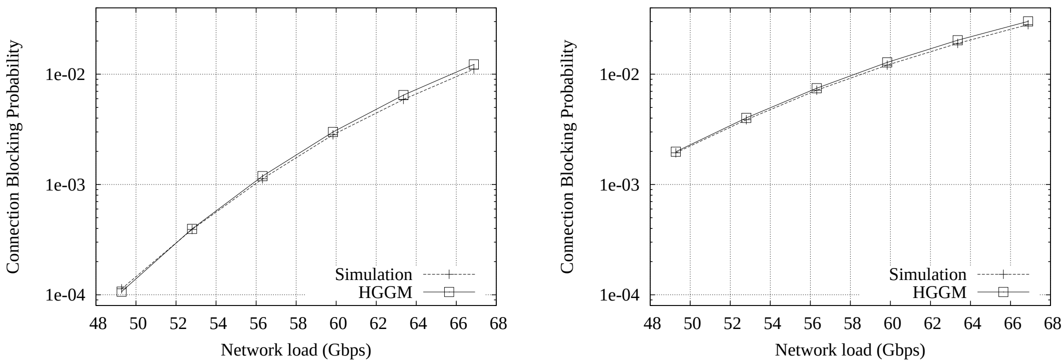


Figure 5.1: Blocking probability in the CWEN network, guaranteed bandwidth b_h is 100 Mbps (left) and 200 Mbps (right)

First, the guaranteed bandwidth of the flows was 100 Mbps and the left plot of Figure 5.1 reports the results obtained with **HGGM** and simulation. A different granularity of the traffic was analysed using b_h equal to 200 Mbps. The results can be seen on the right plot of Figure 5.1. On these plots the total network load is given in Gbps.

Thanks to the consistent extension of the theoretical and the simulation models we can observe high accuracy, similarly to the case of the *optical layer* analysis in Section 3.2.3.

5.3 Grooming elastic traffic

Our aim in this section is to investigate how traffic elasticity impacts on some basic grooming algorithms and assess their performance in dynamic networking scenarios where the *optical layer* and *data layer* interact. Beyond the resource sharing and traffic modelling problem analysed in Section 4.2 we meet also the issue of how the optical management plan behaves and assigns resources to traffic relations.

5.3.1 Considered models and schemes

As mentioned above, the adaptivity of traffic has a deep impact on the network performance and in particular on routing algorithms. The reason lies in the feedback nature of the interaction of elastic traffic with the network: the network status, e.g. congestion induces a negative or positive feedback reaction in the source. As usual in closed loop systems with delay, the nature of feedback, i.e., whether it is positive or negative, can change with changing conditions. For instance, a negative feedback at low loads can change to a positive feedback at high loads, leading to instability phenomena.

Considering the foreseen trend of the very dynamic and aggressive use of optical paths in metro-area optical network and the behaviour of the aggregated flows in more traditional wide-area optical networks it is easy to derive that elasticity in groomed traffic can arise in very different scenarios. To model the elasticity we used *TB* and *DB* model for the *data layer* traffic in the analysis of IP over WDM networks with dynamic traffic grooming.

Let us concretise more the general network vision of Section 2.1. The *data layer* assumes traditional routers with arbitrary routing, that is performed on the virtual topology. In this topology each *lightpath* of the *optical layer* is seen as a single hop regardless of the number of OXCs it crosses. Further, we consider two node architectures: a node can be a pure OXC which allows to switch entire *lightpaths* from an ingress port to an egress port, or it can be a Grooming OXC (G-OXC) which supports elastic traffic flows and groom them onto optical channels and realise also IP router functions. The *data layer* traffic can then be transmitted or received only in G-OXCs and transit traffic, which does not terminate in the IP router, can be groomed with incoming traffic. We are not concerned on the technology (optical or electronic) used for grooming.

In this architecture, a path connecting two routers in the IP layer is called a virtual or logical path, because it is created over some established *lightpath* in the optical layer, i.e., over virtual links. We assume the overlay grooming architecture [6] that does not allow the exchange of information between the layers. The consideration of this separated control is rather realistic in networks where different providers manage the services of the layers as usual in networks of our days. In this case the routing in the *data layer* and the RWA in the *optical layer* work without any direct interaction. The lack of knowledge of the actual resource use in the other layer may lead to taking decisions based on a local optimum that affects negatively later decisions.

Here, the only, but not trivial function of grooming is to decide whether a new traffic relation must be routed at the IP level, i.e., using the current virtual topology. The other option is to demand the setup of new virtual links realised by establishing *lightpaths* in the *optical layer*. We considered the most cited grooming algorithms introduced in [80] for CBR traffic, and adopted them to the elastic traffic scenario as follows:

- *Virtual-topology First* (*VirtFirst*). Each time a new IP request arrives in some router, the current virtual topology is considered first to route the request. If, once routed, the amount of bandwidth for some flow, not necessarily the one being routed, is less than a threshold th_o, then if possible, a new *lightpath* is set-up between source and destination. If the setup is successful, the IP request is routed over it and a new virtual topology is computed at the *data layer*. The new topology does not affect already routed requests, i.e., no re-routing is considered, but it will be used for routing all new requests. If the new *lightpath* cannot be set up, the request is routed based on the current virtual topology.

- *Optical-level First* (*OptFirst*). Each time a new IP request arrives in some router, the G-OXC always attempts first to set up a new *lightpath* in the optical layer, in order to route the request over it. If no free wavelengths are available here, the IP router routes the incoming request over the current virtual topology. Indeed, a virtual topology is addressed only when optical resources for the considered source-destination pair are exhausted.

Since elastic traffic is considered, the upper limit to the possible number of flows which exist in the *data layer* depends from the minimal acceptable bandwidth b_m. However, this parameter can not be applied directly in grooming, thus the need for the establishment of a new *lightpath* in *VirtFirst* must be introduced based on some suitable parameter. We define the above referred parameter, called *optical opening threshold* th_o, as a threshold on the instantaneous throughput obtained by connections.

Whenever a closing flow leaves a *lightpath* empty, the *lightpath* is closed too and the virtual topology is re-computed. A generalised formulation and more detailed analysis of grooming policies with overlay architecture is presented in [82].

5.3.2 Performance measures

To analyse the complex function of dynamic grooming we joined and extended the features of the previously used simulation tools ANCLES and ASONCLES. A more detailed description of the resulting tool GANCLES [39] can be found in [81].

In [32] we derived some results on a trivial topology to demonstrate that dynamic grooming analysis can lead to surprising results. Since the theoretical models are of rather limited generality, here we present only some simulation results to give an illustration of the performance of such networks. As we will see, the phenomena involved in routing and grooming elastic traffic are rather complex, and often far from intuitive. To highlight some basic effects caused by the interaction of grooming algorithms and elastic traffic we use only very simple routing algorithms both at the optical and at the IP level.

We compared the traffic models *TB* and *DB* introduced in Section 4.2.3 similarly as in Section 4.2.5, but in the environment of IP over WDM. Thus, beside the parameters used in case of a separated analysis of the *optical layer* and *data layer*, some new parameters and performance measures have to be considered. Three of them are at the IP level and three at the optical level:

T: the average throughput per flow as introduced in Section 4.2.5.1,

p_s: the starvation probability as introduced in Section 4.2.3.3,

N_l: average number of IP hops per flow,

N_{lo}: average number of optical hops per *lightpath*,

L_o: average of the open optical channels per optical link,

R_o: the ratio between the opening rate of optical paths and the arrival rate of flows at the IP level. It is a measure of the optical level routing effort.

Note that for an optical routed network without grooming $R_o = 1$, while for a network with static grooming $R_o = 0$.

Clearly, the goal of a grooming algorithm is maximising T while minimising p_s, R_o and L_o. For the other parameters there are no obvious guidelines on the optimal behaviour.

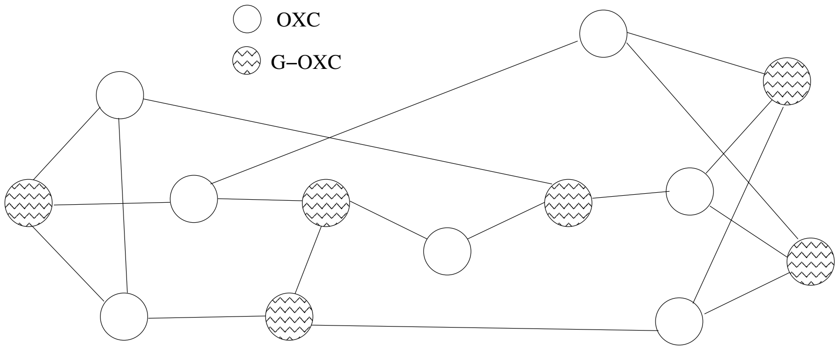


Figure 5.2: NSFNET network

The dynamic grooming policies were compared in scenarios with regular, e.g. full mesh and ring, and irregular topologies in [32, 81, 82]. Here we present results obtained on the NSFNET network [83] shown in Figure 5.2, which has 13 nodes and 18 fiber links. Each fiber carry up to 4 wavelengths, and only 6 nodes out of 13 have grooming capability, i.e., they are G-OXCs. Each wavelength has a capacity of 20 Gbps. A best-effort traffic source is connected to each G-OXC, opening flows with $B_M = 10$ Gbps. Each flow transfers data whose size is randomly chosen from an exponential distribution with average 12.5 GBytes. A uniform traffic pattern is simulated, th_s is set to 0.1 in all simulations and $th_o = th_s$ for the sake of simplicity.

The left plot of Figure 5.3 presents a comparison of the average throughput T obtained by modelling best-effort traffic relations using the *TB* and the *DB* approach. Simulations were performed for both *VirtFirst* and *OptFirst* grooming algorithms. The right plot reports the starvation probability.

The difference in performance results of the two approaches is striking. Let us consider first the *OptFirst* grooming policy. Both approaches show T starting from 10 Gbps when the offered load is low and they immediately diverge as the offered load increases.

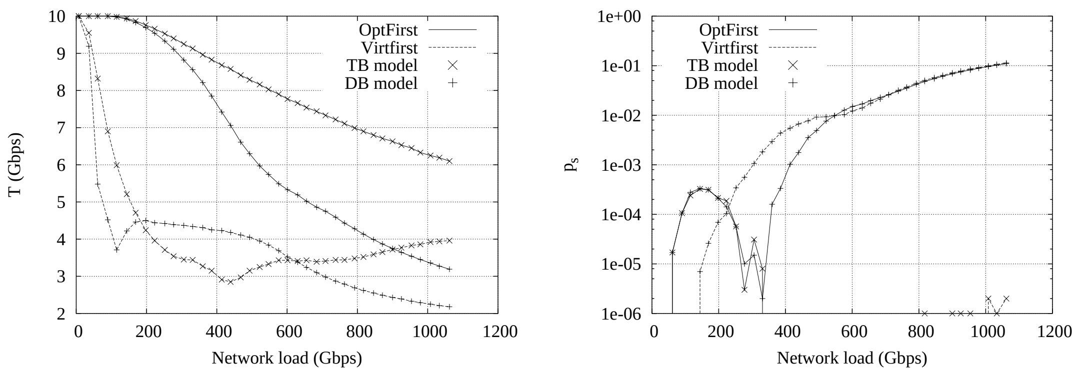


Figure 5.3: Average bandwidth per connection (left) and starvation probability (right) for *DB* and *TB* models

Indeed, the *DB* traffic model shows much faster decrease in T as soon as the offered load increases and this is due to the spreading of congestion over the time with a sort of avalanche effect mentioned in Section 4.2.5.2. On the contrary, the *TB* traffic model shows a smoother decrease of the average bandwidth.

Analysing the starvation probability on the righthand plot of Figure 5.3 adds more insight. When the traffic is very low, i.e., below 350 Gbps, both traffic models show the same, very strange behaviour for *OptFirst*: p_s increases and then decreases sharply. This form of starvation is independent of the traffic model and it is due to a very aggressive and dynamic use of optical resources of this grooming solution. This greedy policy leads sometimes to have no connectivity in the *data layer* between the source and the destination at flow request arrival. This means, that there is neither any available path in the virtual topology, nor the possibility to establish a new direct *lightpath* in the *optical layer*. Note that this effect of starvation can be interpreted rather as flow request blocking.

However, when the load increases *lightpaths* become more stable, because there is always traffic keeping them open. Thus the probability that the virtual topology is not completely connected becomes negligible. When the load increases further, the two traffic models behaviour diverges: the *TB* model shows no starvation at all, while it increases steadily in the *DB* model. This difference in the p_s behaviour enhances the differences in T, since aborting flows cause a waste of bandwidth.

When considering the *VirtFirst* grooming policy instead, the behaviour of both traffic models is different from the previous one: T decreases sharply even when the offered load is low, due to the conservative policy. In fact, *VirtFirst* sets up the minimum number of *lightpaths* in order to guarantee the minimum network connectivity, and keeps this configuration unchanged until some flow would cross the *optical opening threshold*. Only in this case *VirtFirst* increases the resources at IP level by setting up new *lightpaths*, if at all possible. In particular, the T for *DB* traffic model decreases very rapidly, causing an earlier set-up of new *lightpaths* compared to *TB* traffic. This leads the mean assigned bandwidth T to "bounce" taking advantage of the higher number of *lightpaths* in the network. Using the Data-based approach it happens at a load much smaller than for the Time-based model, that starts increasing again at higher loads. Obviously, both models show definitive decrease in T for higher loads, not presented here. It is interesting to notice that the starvation rate for the *TB* model is in this case always zero apart from some points at very high loads with a p_s around 10^{-6}. Using the *DB* model p_s increases steadily and shows a behaviour similar to the *DB* model with the *OptFirst* policy.

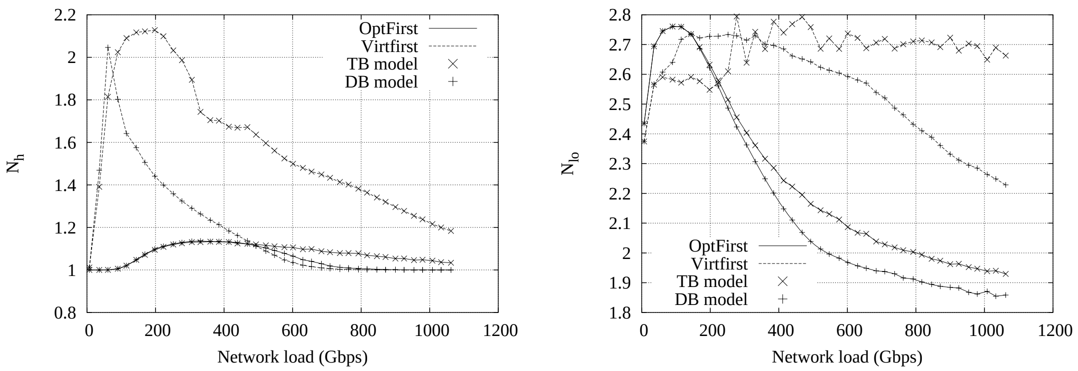


Figure 5.4: Average number of hops N_l in the *data layer* (left) and N_{lo} in the *optical layer* (right)

The left plot of Figure 5.4 reports the average number of IP hops per flow N_l. As expected the *OptFirst* policy for both *DB* and *TB* model keeps this parameter close to one since it attempts to open always direct *lightpaths* between the source and destination. The *VirtFirst* for *DB* traffic instead shows a sharp increase for low loads due to the high probability that a flow arrives and finds the virtual topology already connecting the source and destination. As the load increases, N_l decreases and converges to one because also the *VirtFirst* policy starts to open direct *lightpaths*. The same behaviour can be observed for the Time-based approach but with a smoother peak at a higher load value. This difference is due to the fact that in the *TB* case the T of flows decreases smoother and thus *VirtFirst* starts to open new *lightpaths* at higher loads.

Figure 5.4 plots the parameter N_{lo} on the right part. Once again the behaviour of the *OptFirst* policy is more predictable, with the number of links that decreases steadily with the load, and roughly converges to the weighted average distance in number of links between G-OXCs for both *TB* and *DB*. The *VirtFirst* policy shows instead longer optical paths even for high loads. This effect is due to the intrinsic behaviour of this grooming policy: most of the *lightpaths* set up by *VirtFirst* are in fact never torn-down since they are carrying traffic almost all the time. On the other hand this policy uses the *optical layer* resources in a conservative way, thus, when a new, longer *lightpath* must be established, it is more likely to find enough free optical channels to set it up. The difference of traffic models for *VirtFirst* comes again from the earlier decrease of T with *DB* that implies *lightpath* setups.

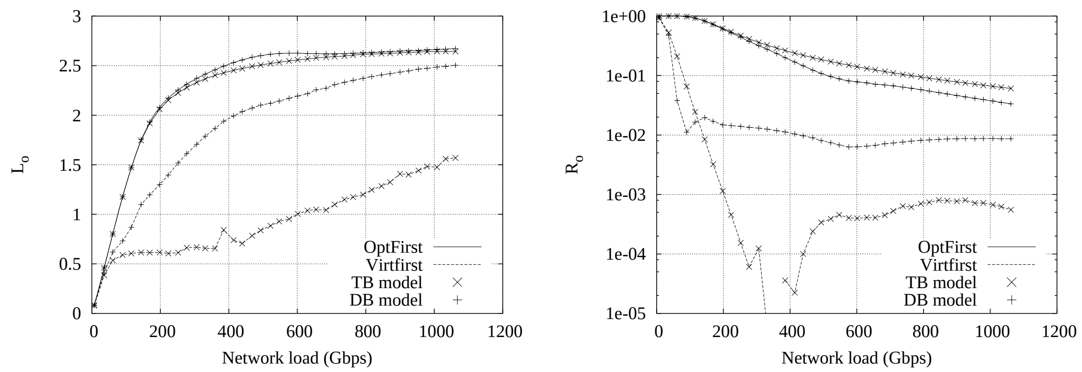


Figure 5.5: Average optical link load L_o (left) and *lightpath* opening rate R_o (right)

Considering the characteristics discussed above for the different policies and traffic models, the average number L_o of opened optical channels per link in the *optical layer* plotted on the lefthand side of Figure 5.5 does not present any surprise. The policy *VirtFirst* uses a lot of resources with both traffic approaches and converges to a saturated *optical layer* at high loads. However, due to the topology, at every time there are some channels that remain free and L_o does not reach the total number of channels in an optical link. Using *VirtFirst* the optical link load increases in a less aggressive manner. L_o converges to that of *OptFirst* for the model *DB*, while for *TB* it ends up in a much lower value.

The right plot of 5.5 shows the ratio R_o. As expected, when *VirtFirst* grooming policy is used, R_o decreases quickly with the load, indicating a burden for the optical level that does not increase with the traffic. It is interesting to notice the different behaviour of this parameter between the two traffic models. Using the *DB* model *VirtFirst* shows a lower and static rate of *lightpath* opening when the load increases, while for *TB* we find a decreasing and then re-increasing characteristics. When instead, the *OptFirst* grooming policy is used, R_o decreases slowly and smoothly, indicating a rather high burden for the optical level.

Results in other scenarios verified our comments by showing similar behaviour of the algorithms.

5.4 Discussion

In this chapter we dealt with the problem of dynamic traffic grooming in IP over WDM networks.

First we show how the **WDMM** model of Section 3.2 can be extended to study the grooming of homogenous and guaranteed traffic. The concept of *subwave channels* lead us to the derivation of model **HGGM**. The model provides accurate approximation of the connection request blocking probability. However, it is not a very robust solution and can be applied only for networks with homogenous conversion and multiplexing characteristics in each nodes.

Second, we presented some initial steps of the analysis of dynamic grooming algorithms with elastic traffic. The research of this area is still in an introductory phase and the solutions published in the relating works place little attention on how to suit the algorithms to realistic scenarios in IP over WDM networks.

The elasticity of traffic interacts with the grooming algorithms as well as with routing both at the IP and optical level, leading to unexpected results. Two grooming algorithms, both using an overlay architecture, have been considered and analysed comparing the Time- and Data-based approach in the elastic traffic modelling. In both the *OptFirst* and *VirtFirst* grooming algorithms the impact of elastic traffic is dramatic, showing clearly that approximating IP traffic with a not well suited traffic model can lead to wrong conclusions when routing and grooming are considered.

The results for both policies show that they are not appropriate to manage the two-layer network due to the lack of layer coordination in the overlay architecture. On the one hand, using *OptFirst*, it can lead to wasting resources in *optical layer*, that may cause even request blocking due to not completely connected virtual topology. On the other hand, using *VirtFirst*, the conservative guidelines applied to set up a new virtual link may lead to low quality of data transport at the *data layer*.

Chapter 6

Conclusion

In the PhD dissertation we addressed the problem of the performance analysis of the general routing solution in IP over WDM networks, including the routing algorithms in both layers and the grooming policies. This task was separated into three subtasks and analysed with different methods.

First, in Chapter 3 the *optical layer* was studied. We introduced the **WDMM** theoretical model that calculates the blocking probability of optical connection requests in multifiber optical networks without wavelength conversion. The results show the rather low complexity of the computation algorithm and the accuracy of the model.

However, in the case of low load and long routes assigned to the connections some inaccuracy can be observed, thus further refinement of the model is required. On the other hand, **WDMM** shall be extended to support more general traffic models, RWA solutions and network architectures.

Second, Chapter 4 dealt with the analysis of the *data layer*. Two flow-based traffic models representing elasticity and a model for flow starvation were introduced and tested with known routing algorithms. Simulation results show the clear difference of the *TB* and *DB* approaches proving that a simple traffic model fails in the performance analysis and may lead to wrong conclusions.

We studied also the impact of stale link state information on the performance of the dynamic routing solutions. We found a strong degradation when the link state update frequency decreases and proposed **MSF**, a new class of routing based on multiple filtering of available paths. Some of these algorithms resist to the degrading effects caused by the large update periods. Another group of QoS routing algorithms **NGR** was developed to realise a kind of built-in load dependence by avoiding heavily loaded segments of the network. We found that the routing solutions adopting the concept of network graph reduction during the path selection process allow good performance for high network load as well, even in complex scenarios.

Third, the compound IP over WDM network scenario was studied in Chapter 5 regarding the interaction of *data layer* and *optical layer* when assuming dynamic grooming. In the first part of the chapter we proposed **HGGM**, a theoretical model that considers requests with bandwidth guarantees. In the case of this model several constraints of applicability have to be loosed.

The second part presents the study of grooming policies of overlay architecture in elastic traffic scenarios. We found again the striking difference between the *TB* and *DB* traffic models in the *data layer* and highlighted some special behaviours due to the grooming policy characteristics. A further analysis of this issue can help us to determine the roots of all the observed effects and to design new algorithms that suit better to this environment. A possible direction of further research is the definition and analysis of new performance measures. To generalise our studies, the observation of dynamic grooming policies of more complex architectures and combined with different routing algorithms in the *optical layer* and *data layer* is required.

Bibliography

[1] R. Ramaswami and K. N. Sivarajan. Routing and wavelength assignment in all-optical networks. *IEEE/ACM Transactions on Networking*, 3(5):489–500, 1995.

[2] G. Apostolopoulos, R. Guérin, S. Kamat, and S. K. Tripathi. Quality of service based routing: A performance perspective. In *Proc. of ACM SIGCOMM'98*, 1998. Vancouver, Canada.

[3] E. Karasan and E. Ayanoglu. Effects of wavelength routing and selection algorithms on wavelength conversion gain in wdm optical networks. *IEEE/ACM Transactions on Networking*, 6(2):186–196, 1998.

[4] R. Dutta and G. N. Rouskas. Traffic grooming in wdm networks: past and future. *IEEE Network Magazine*, 16(6):46–56, 2002.

[5] R. Srinivasan and A. K. Somani. Dynamic routing in wdm grooming networks. *Photonic Network Communications*, 5(2):123–135, 2003.

[6] B. Rajagopalan, J. Luciani, and D. Awduche. RFC 3717: IP over optical networks: A framework, March 2004.

[7] A. Jajszczyk. Optical networks - the electro-optic reality. *Optical Switching and Networking*, 1(1):3–18, 2005.

[8] B. Craignou, R. Clemente, J. Robaday, L. Jereb, Z. Ioannidis, and J. Santos. Network operator perspectives on optical networks - evolution towards ason. In *Proc. of Networks 2002*, June 2002. Munich, Germany.

[9] Eurescom P1012, Flexible Automatic SwitcHed Optical Networks. http://www.eurescom.de/public/projects/P1000-series/p1012/.

[10] H. Harai, M. Murata, and H. Miyahara. Performance analysis of wavalength assignment policies in all-optical networks with limited-range wavelength conversion. *IEEE Journal on Selected Areas in Communications*, 16(7):1051–1060, 1998.

[11] et al. Kaminow, I. P. A wideband all-optical wdm network. *IEEE Journal on Selected Areas in Communications*, 14(5):780–799, 1996.

[12] S. Subramaniam, A. K. Somani, M. Azizoglu, and R.A. Barry. A performance model for wavelength conversion with non-poisson traffic. In *Proc. of IEEE INFOCOM'97*, pages 499–506, April 1997.

[13] Zoltán Zsóka. Ason hálózatok forgalmának modellezése. In *Proc. of HTE-BME 2002 diákkonferencia*, May 2002. Budapest, Hungary.

[14] S. Shenker and J. Wroclawski. RFC 2216: Network element service specification template, September 1997.

[15] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. RFC 2475: An architecture for Differentiated Services, December 1999.

[16] J. Roberts. Traffic theory and the internet. *IEEE Communications Magazine*, 39(1):94–99, 2001.

[17] S. Shakkottai, A. Kumar, A. Karnik, and A. Anvekar. Tcp performance over end-to-end rate control and stochastic available capacity. *IEEE/ACM Transactions on Networking*, 9(4):377–391, 2001.

[18] S. Rácz, B. P. Gerõ, and G. Fodor. Flow level performance analysis of a multi-service system supporting elastic and adaptive services. *Performance Evaluation*, 49(1-4):451–469, 2002.

[19] A. A. Kherani and A. Kumar. Stochastic models for throughput analysis of randomly arriving elastic flows in the internet. In *Proc. of IEEE INFOCOM'02*, 2002.

[20] Q. Ma and P. Steenkiste. Routing traffic with quality-of-service guarantees in integrated services networks. In *Proc. of 8th IEEE/ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 1998)*, July 1998. England.

[21] A. Shaikh, J. Rexford, and K.G. Shin. Load-sensitive routing of long-lived ip flows. In *Proc. of ACM SIGCOMM'99*, pages 215–226, August 1999. Cambridge, MA USA.

[22] C. Casetti, G. Favalessa, M. Mellia, and M. M. Munafò. An adaptive routing algorithm for best-effort traffic in integrated-services networks. In *Proc. of 16th International Teletraffic Congress, IEE ITC-16*, June 1999. Edinburgh, UK.

[23] C. Casetti, R. Lo Cigno, M. Mellia, M. M. Munafó, and Zoltán Zsóka. A new class of qos routing strategies based on network graph reduction. *Computer Networks*, 41(4):475–487, 2003.

[24] R. Gu'erin and A. Orda. Qos-based routing in networks with inaccurate information. In *Proc. of IEEE INFOCOM'97*, April 1997. Kobe, Japan.

[25] G. Apostolopoulos, R. Guérin, S. Kamat, and S. K. Tripathi. Improving qos routing performance under inaccurate link state information. In *Proc. of 16th International Teletraffic Congress, IEE ITC*, June 1999. Edinburgh, UK.

[26] A. Shaikh, J. Rexford, and K.G. Shin. Evaluating the impact of stale link state on quality-of-service routing. *IEEE/ACM Transactions on Networking*, 9(2):162–176, 2001.

[27] S. Sibal and A. De Simone. Controlling alternate routing in general-mesh packet flow networks. In *Proc. of ACM SIGCOMM'94*, August 1994. London, UK.

[28] H. Zhu, H. Zang, K.Zhu, and B. Mukherjee. A novel generic graph model for traffic grooming in heterogeneous wdm mesh networks. *IEEE/ACM Transactions on Networking*, 11(2):285–299, 2003.

[29] S. Koo, G. Sahin, and S. Subramaniam. Dynamic lsp provisioning in overlay, augmented, and peer architectures for ip/mpls over wdm networks. In *Proc. of IEEE INFOCOM'04*, Mar. 7-11 2004. Hong Kong, China.

[30] R. Srinivasan and A. K. Somani. Analysis of optical networks with heterogeneous grooming architectures. *IEEE/ACM Transactions on Networking*, 12(5):931–943, 2004.

[31] R. Lo Cigno, E. Salvadori, D. Severina, and Z. Zsóka. GANCLES: A network level simulator to study optical routing, wavelength assignment and grooming algorithms. Technical Report DIT-04-017, DIT - University of Trento, 2004.

[32] R. Lo Cigno, E. Salvadori, and Z. Zsóka. Elastic traffic effects on wdm dynamic grooming algorithms. In *Proc. of IEEE GLOBECOM'04*, Nov. 29 - Dec. 3 2004. Dallas, Texas USA.

[33] L. Kleinrock. *Queueing Systems*. John Wiley, 1976.

[34] T.G. Robertazzi. *Computer Networks and Systems: Queueing Theory and Performance Evaluation*. Springer, 2000.

[35] D. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, 1992.

[36] M. Ajmone Marsan, A. Bianco, C. Casetti, C. F. Chiasserini, A. Francini, R. Lo Cigno, and M. M. Munafò. An integrated simulation environment for the analysis of atm networks at multiple time scales. *Computer Networks and ISDN Systems, Special Issue on Modeling of Wired - and Wireless - ATM Networks*, 29(17-18):2165–2185, 1998.

[37] ANCLES - A Network Call-Level Simulator.
http://www.tlc-networks.polito.it/ancles.

[38] ASONCLES - ASOn Network Call-Level Simulator.
http://www.hit.bme.hu/ zsoka/asoncles.

[39] GANCLES - Gmpls cApable Network Call-Level Simulator.
http://ardent.unitn.it/gancles.

[40] C. Casetti, R. Lo Cigno, M. Mellia, M. M. Munafó, and Zoltán Zsóka. Qos routing strategies when link state information is out-of-date. In *Proc. of 8th IFIP Workshop on Performance Modelling and Evaluation of ATM and IP Networks*, July 17-19 2000. Ilkley, West Yorkshire U.K.

[41] C. Casetti, R. Lo Cigno, M. Mellia, M. M. Munafó, and Zoltán Zsóka. A realistic model to evaluate routing algorithms in the internet. In *Proc. of IEEE GLOBECOM'01*, Nov. 25-29 2001. San Antonio, Texas USA.

[42] T. Yung, J. Martin, M. Takai, and R. Bagrodia. Integration of fluidbased analytical model with packet-level simulation for analysis of computer networks. In *Proc. of Internet Performance and Control of Network Systems II*, volume 4523, pages 130–143, 2001. Denver, CO USA.

[43] F. P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37, 1997.

[44] Chiu. Some observations on fairness of bandwidth sharing. In *Proc. of ISCC 2000. Fifth IEEE Symposium on Computers and Communications*, July 2000.

[45] G. de Veciana, T. Konstantopoulos, and Tae-Jin Lee. Stability and performance analysis of networks supporting elastic services. *IEEE Journal on Selected Areas in Communications IEEE/ACM Transactions on Networking*, 9(1):2–14, 2001.

[46] L. Li and A. K. Somani. Dynamic wavelength routing using congestion and neighborhood information. *IEEE/ACM Transactions on Networking*, 7(5):779–786, 1999.

[47] H. Zang, J. P. Jue, and B. Mukherjee. A review of routing and wavelength assignment approaches for wavelength-routed optical wdm networks. *Optical Networks Magazine*, 1(1):47–60, 2000.

[48] Y. Yoo, S. Ahn, and C. S. Kim. Adaptive routing considering the number of available wavelengths in wdm networks. *IEEE Journal on Selected Areas in Communications*, 8(21):1263–1273, 2003.

[49] R. Barry and P. Humblet. Models of blocking probability in all-optical networks with and without wavelength changers. In *Proc. of IEEE INFOCOM'95*, April 1995. Boston, US.

[50] S. Subramaniam, M. Azizoglu, and A. Somani. All-optical networks with sparse wavelength conversion. *IEEE/ACM Transactions on Networking*, 4(4):544–557, 1996.

[51] L. Li and A. K. Somani. Blocking performance analysis of fixed-paths least-congestion routing in multifiber wdm networks. In *Proc. of SPIE Photonics East'99*, pages 56–67, September 1999. Boston, MA USA.

[52] L. Li and A. K. Somani. A new analytical model for multi-fiber wdm networks. *IEEE Journal on Selected Areas in Communications*, 18(10):2138–2145, 2000.

[53] A. Birman. Computing approximate blocking probabilities for a class of all-optical networks. *IEEE Journal on Selected Areas in Communications*, 14(5):852–857, 1996.

[54] S. Subramaniam, A. K. Somani, M. Azizoglu, and R. A. Barry. The benefits of wavelength conversion in wdm networks with non-poisson traffic. *IEEE Communications Letters*, 3(3):81–83, 1999.

[55] Y. Zhu, G. Rouskas, and H. Perros. A path decomposition approach for computing blocking probabilities in wavelength routing networks. *IEEE/ACM Transactions on Networking*, 8(6):747–762, 2000.

[56] A. Mokhtar and M. Azizoglu. Adaptive wavelength routing in all-optical networks. *IEEE/ACM Transactions on Networking*, 6(2):197–206, 1998.

[57] R. Srinivasan and A. K. Somani. A generalized framework for analyzing time-space switched optical networks. In *Proc. of IEEE INFOCOM'01*, April 2001. Anchorage, USA.

[58] A. Sridharan and K. N. Sivarajan. Blocking in all-optical networks. In *Proc. of IEEE INFOCOM'00*, March 2000. Tel-Aviv, Israel.

[59] Zoltán Zsóka. Blocking probability of all-optical multifiber networks. In *Proc. of 7th IFIP Working Conference on Optical Network Design and Modelling, ONDM 2003*, February 3-5 2003. Budapest, Hungary.

[60] COST239 homepage. http://www.ita.hsr.ch/cost239/.

[61] A. Fumagalli, I. Cerutti, M. Tacca, F. Masetti, R. Jagannathan, and S. Alagar. Survivable networks based on optimal routing and wdm self-healing rings. In *Proc. of IEEE INFOCOM'99*, March 1999. New York, NY USA.

[62] E. Rosen, A. Viswanathan, and R. Callon. RFC 3031: Multiprotocol Label Switching architecture, January 2001.

[63] D. Cavendish and M. Gerla. Internet qos routing using the bellman-ford algorithm. In *Proc. of IFIP Conference on High Performance Networking, HPN 1998*, 1998. Vienna, Austria.

[64] G. Apostolopoulos, D. Williams, S. Kamat, R. Guerin, A. Orda, and T. Przygienda. RFC 2676: QoS routing mechanisms and OSPF extensions, August 1999.

[65] C. Casetti, R. Lo Cigno, M. Mellia, M. M. Munafó, and Zoltán Zsóka. Routing algorithms evaluation for elastic traffic. In *Proc. of IEEE Workshop on High Performance Switching and Routing, HPSR 2001*, May 29-31 2001. Dallas, Texas USA.

[66] C. Casetti, R. Lo Cigno, M. Mellia, M. M. Munafó, and Zoltán Zsóka. A new class of qos routing strategies based on network graph reduction. In *Proc. of IEEE INFOCOM'02*, June 23-27 2002. New York, NY USA.

[67] Q. Ma, P. Steenkiste, and H. Zhang. Routing high-bandwidth traffic in max-min fair share networks. In *Proc. of ACM SIGCOMM'96*, August 1996. Stanford, Canada.

[68] A. Arvidsson and P. Karlsson. On the traffic models for tcp/ip. In *Proc. of 16th International Teletraffic Congress, IEE ITC-16*, June 1999. Edinbourgh, UK.

[69] Z. Wang and J. Crowcroft. Qos routing for supporting resource reservation. *IEEE Journal on Selected Areas in Communications*, 14(7):1288–1294, 1996.

[70] GT-ITM: Georgia Teach-Internetwork Topology Models. http://www.cc.gatech.edu/project/gtitm.

[71] C. Casetti, R. Lo Cigno, and M. Mellia. Qos-aware routing schemes based on hierarchical load-balancing for integrated services packet networks. In *Proc. of International Communication Conference, IEEE ICC'99*, June 1999. Vancouver, Canada.

[72] J. L. Sobrinho. Algebra and algorithms for qos path computation and hop-by-hop routing in the internet. *IEEE/ACM Transactions on Networking*, 10(4):541–550, 2002.

[73] T. Cinkler. Traffic and lambda grooming. *Network*, 17(2):16–21, 2003.

[74] C. Xin. Computing blocking probability of dynamic traffic grooming in mesh wdm optical networks. In *Proc. of First International Conference on Broadband Networks, 2004. BroadNets 2004*, October 2004.

[75] Zoltán Zsóka. Dinamikus forgalomkötegelés (grooming) analízise optikai hálózatokban. *Magyar Távközlés*, 15(2):33–37, 2004.

[76] X. Zhang and C. Qiao. An effective and comprehensive approach to traffic grooming and wavelength assignment in sonet/wdm rings. *IEEE/ACM Transactions on Networking*, 8(5):608–617, 2000.

[77] K. Zhu and B. Mukherjee. Traffic grooming in an optical wdm mesh network. *IEEE Journal on Selected Areas in Communications*, 20(1):122–133, 2003.

[78] M. Brunato and R. Battiti. A multistart randomized greedy algorithm for traffic grooming on mesh logical topologies. In *Proc. of 6th IFIP Working Conference on Optical Network Design and Modelling, ONDM 2002*, February 2002. Torino, Italy.

[79] M. Kodialam and T. V. Lakshman. Integrated dynamic ip and wavelength routing in ip over wdm networks. In *Proc. of IEEE INFOCOM'01*, Apr. 22-26 2001. Anchorage, USA.

[80] X. Niu, W. D. Zhong, G. Shen, and T. H. Cheng. Connection establishment of label switched paths in ip/mpls over optical networks. *Photonic Network Communications*, 6(1):33–41, 2003.

[81] R. Lo Cigno, E. Salvadori, and Z. Zsóka. Dynamic grooming in overlay networks: A study with realistic traffic based on gancles simulation package. In *Proc. of 9th IFIP TC6 Conference on Optical Network Design and Modelling, ONDM 2005*, February 7-9 2005. Milano, Italy.

[82] R. Lo Cigno, E. Salvadori, and Z. Zsóka. Dynamic grooming in ip over optical networks based on the overlay architecture. *submitted to Optical Switching and Networking*.

[83] O. Crochat and J. Boudec. Design protection for wdm optical networks. *IEEE Journal on Selected Areas in Communications*, 16(7):1158–1165, 1998.