TECHNISCHE UNIVERSITÄT WIEN
VIENNA UNIVERSITY OF TECHNOLOGY

DISSERTATION

Dynamic Virtualization and Service Provision in Multi-Provider GMPLS Networks

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Prof. Dipl. Ing. Dr.techn Admela Jukan
Institut für Datentechnik und Kommunikationsnetze
Technischen Universität Carolo-Wilhelmina zu Braunschweig, und
E 388 – Institut für Breitbandkommunikation
Technischen Universität Wien

und Begutachtung von

Prof. Dr. Dimitra Simeonidou
The Photonic Networks Laboratory
Department of Electronic Systems Engineering
University of Essex

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Dipl.-Ing. Slobodanka Tomic
Matr.-Nr. 9826910

Wien, im August 2007

Zusammenfassung

In den letzten Jahren wurde zunehmend erkannt, dass die weitere Evolution des Internets eines neuen Kontrollparadigmas bedarf, das den Netzbetreibern ermöglicht, eigene Netze auf der physikalischen Netzwerkschicht (L1) zu virtualisieren und virtuelle Infrastrukturen als Netzdienste anzubieten. Der Technologie der optischen Netze wurde aufgrund ihrer inhärenten Flexibilität eine Schlüsselrolle in diesem neuen Szenario zuteil Des Weiteren erkannte man, dass sich das volle Potenzial der L1 Virtualisierung nur mit den dynamischen und automatischen Mechanismen und Protokollen für die Netzsteuerung entfalten kann. Die bekanntesten und gebräuchlichsten Mechanismen und Protokollen werden vom IETF innerhalb des GMPLS Rahmenwerks vorgeschlagen.

Motiviert durch den Bedarf eines effizienten Ressourcenmanagements in virtualisierten optischen Netzen wird im Rahmen dieser Dissertation ein "Infrastructure Service Model" für die dynamische L1 Virtualisierung entwickelt, das auf zwei in diesem Kontext innovativen Konzepten basiert: der "resource visibility" und dem "GMPLS Exchange Point (GXP)".

Mit der "resource visibility" wird die dienstspezifische Verfügbarkeit der Ressourcen definiert, die den gemeinsamen Zugriff vieler Dienste steuert, während der "GMPLS Exchange Point (GXP)" eine dynamisce Verbindung von GMPLS Domänen verschiedener Netzanbietern ermöglicht. Basierend auf diesen beiden Konzepten wurden in dieser Dissertation die Methoden für "Traffic Routing and Topology Engineering (RToE)" für Infrastrukturdienste entwickelt und in einer Simulationsstudie evaluiert. Die erzielten Resultate demonstrieren signifikante Leistungsverbesserungen eines Ressourcenmanagements auf Basis der eingeführten "resource visibility" und der neuartigen GXP-basierten Architektur und stellen einen wesentlichen Ausblick und Input für die zukünftige Forschungsrichtung dar.

Abstract

Recently, it has been recognized that for the Internet to evolve, it is important that network operators be able to *virtualize* their physical infrastructure into multiple parallel virtual networks, which can support different protocols and services. Due to its inherent flexibility, optical networking technology has been identified as a key enabler for a *new breed* of efficient virtualization at the *physical network layer* or Layer 1 (L1). The major role in this approach play mechanisms and protocols for the *dynamic and automatic* network control, the most prominent of which have been proposed within the IETF Generalized Multi-protocol Label Switching (GMPLS) framework. Motivated by the need for efficient resource management in virtualized optical networks, this thesis develops the *GMPLS network federation architecture* and the *Infrastructure Service* model based on two novel concepts: (1) the dynamic interconnection of GMPLS domains is facilitated with the introduction of the novel *GMPLS Exchange Point (GXP)*, which is a physical layer equivalent of the Internet Exchange Point (IXP), and (2) the resource allocation and sharing of resources among infrastructure services are controlled by means of a new *resource visibility attribute*, which represents service-specific resource usage policies within the control plane. Based on these concepts, this thesis develops and evaluates in a simulation study dynamic traffic routing and topology engineering (RToE) methods for the infrastructure services. The results obtained demonstrate significant performance benefits of the visibility-enabled resource control and GXP-based architecture, and indicate important directions for further standardization and research work.

To my parents
JOSIPA and ZDRAVKO

Acknowledgment

I gratefully acknowledge my advisor Prof. Admela Jukan for her constant support and encouragement. She was an indispensable source of personal and professional inspiration and ever-lasting enthusiasm. My thanks also goes to Prof. Harmen R. van As for giving me the opportunity to conduct a part of this research at the Institute for Broadband Communications (*Institut für Breitbandkommunikation*). I also gratefully acknowledge Prof. Christoph Mecklenbruker for his involvement as second examiner (*Zweitgutachter*), and Prof. Dimitra Simeonidou for acting as an external examiner. Many thanks go to Das Forschungszentrum Telekommunikation Wien (ftw.) where I completed this work.

I have been supported and encouraged in my work by a number of people from the optical research community, who I was privileged to meet in many occasions, and particularly during my involvement in the EU Network of Excellence e-Photon/ONE. It is a great pleasure that I can here express my gratitude for all of them, and specially to Prof. Tibor Cinkler, Dr. Dave Faulkner, Prof. Biswanath Mukherjee, Prof. Fabio Neri, Prof. Josep Sole-Pareta, Dr. Richard Rabbat, Prof. George Rouskas, and Prof. John Senior.

The greatest and warmest thanks goes to my husband Dusan and to our son Tim, for their enduring love, patience, and cheer. I love you!

Vienna, August 2007

Slobodanka Tomic

Contents

List of Figures

List of Tables

1 Introduction

As a growing number of industrial, scientific and business applications are benefiting from the global Internet based on broadband optical transmission, several major networking trends emerged. First, networking communities in GRID computing [Trav06], User-Controlled Lightpath Architecture (UCLP) [StAr06], and the newly promoted initiative towards Global Environment for Networking Innovations (GENI) [geni06] of the National Scientific Foundation (NSF), recommend that for the Internet to evolve, it is important that network operators be able to configure multiple parallel topologies over the common physical (L1) infrastructure, as a means to open it for new protocols and services. Second, large corporations as well as municipalities are buying or leasing physical resources, including fiber links, switches and routers, in order to build and operate their own networks, which are more flexible and of lower-cost than those offered by the carriers today. Finally, not only the research communities, but also regulatory and industry parties have realized that for the growth of economies and even societies, it is of a significant importance to advance network technologies towards carrier-neutral, open infrastructure [inec06].

In light of these developments, which clearly show the need for a new breed of L1 services, the approach to unified and automatic network control is of particular interest and merit. In fact, it has been proposed and standardized within Generalized Multiprotocol Label Switching (GMPLS) framework of the Internet Engineering Task Force (IETF), and the Automatic Switched Optical Network (ASON) framework of the International Telecommunication Union (ITU-T). Unification of different parties' control tools, e.g., of different technology layers, different carriers, and different network services, is an important step in the evolution of global transport networks. In the ultimate vision, this evolution shall result in an *autonomic network* capable to *reconfigure* when needed, and *self-adapt* to the changing traffic or business needs of its users.

Two scenarios in a transport network evolution that we identified as the most promising to address future business needs, are network federation and network virtualization [Tom02a]. They provide strong motivation for the architecture and service concept proposed and evaluated in this thesis. In the *federation scenario*, multiple administrative domains - i.e., multiple carriers - interconnect and create a control-plane federation, in order to dynamically trade resources and offer global services by means of protocols and mechanisms in the control plane. The points of physical federation - i.e., of traffic exchange - can be determined based on the physical topology of existing networks, and the availability of collocated facilities. The challenges in this scenario are to design a federated topology for optimized network utilization, and to establish a control process that offers benefits of the multi-domain access to the users, and, on the other hand, accounts for operational policies - i.e., objectives and constraints - of the connected carriers. In the network *virtualization* scenario, one large network domain virtually splits into smaller operational entities - i.e., L1 Virtual Private Networks (L1 VPNs)- which use resources of the same physical infrastructure. The major challenge here is to develop methods for efficient dynamic allocation of resources to virtual networks, and for collaborative control of shared resources. To address these needs, this thesis proposes architecture for federation of heterogeneous optical transport networks, based on the ***GMPLS exchange point*** as its central novel element, and introduces the ***Infrastructure Service*** as a special type of L1 VPNs for dynamic virtualization of GMPLS networks.

The GMPLS Exchange Point (GXP) is conceptually similar to the Internet Exchange Point (IXP) [Metz01]; just as how the IXP manages interconnections of autonomous systems (AS) in the Internet by means of the Border Gateway Protocol (BGP) [Rekh06], GXP manages dynamic interconnections of multiple carriers, employing the GMPLS control framework protocols. As such, it supports interconnected carriers in making their physical resources automatically available for dynamic allocation.

The infrastructure service is defined as an atomic operational entity dynamically "instantiable" within the common physical infrastructure, spanning multiple layers and multiple carriers. Similarly to the IETF L1 VPN Service [Take05] the infrastructure service uses the GMPLS-based routing, call admission, and connection management. It has a direct control over the network resources that it dynamically allocates or releases, in order to optimize utility of its virtual topology. Infrastructure services share resources: a virtual link allocated within one service may appear in the virtual topology of a set of other services enabling them to time-share unutilized configured bandwidth. In order to realize this important feature, the infrastructure service model uses a new TE attribute, referred to as the resource visibility attribute, which is assigned to each physical resource and defines whether and how it can be used by different services. The infrastructure service concept is inclusive for very diverse environments with carriers offering dark fibers, SONET/SDH-based TDM containers, or wavelengths. Hence, the examples of resources include transmission, multiplexing and switching functions of switches and routers, within the optical and digital domains, including WDM layer and TDM layers of different granularity. In our service model, network resources may present different visibility to different services. For example, an optical router can be partially visible to one set of infrastructure services, and not visible to other. Together, the concept of GMPLS Exchange Points and the concept of resource visibility enable novel strategies for accommodation of infrastructure services in multi-layer and multi-carrier GMPLS networks.

1.1 Thesis Contribution

The contribution of this thesis relates to two broader research topics: network federations and infrastructure service provision. The first addresses the requirements of the multi-domain heterogeneous optical networks, where we proposed and studied the concept of the federated architecture with the novel ***GMPLS Exchange Point***, [Tom02a], [Tom02b], [Tom03b]. The second deals with the resource management for L1 VPN services by means of GMPLS-enabled dynamic traffic grooming, based on a single-layer or multi-layer routing. Here we proposed a model of extensible ***Infrastructure Services*** with the resource allocation and sharing based on the novel visibility concept, and the dynamic routing and topology engineering (RToE) methods, [Tom03a], [Tom03a], [Tomi04c], [Tomi07].

In the late 1990s and early 2000, relatively low attention was paid to the topic of multi-carrier optical networking both within the research community and standardization groups. Luckily, in the years in which the optical grid was emerging [StAr01], the issues of cooperative (federation-like) networking emerged, and we proposed in [Tom02a], [Tom02b], an architecture for federation of all-optical wavelength switched networks, based on the optical exchange point as well as the trusted control overlay with a global path computation function. With GMPLS rapidly entering the scene [Bane03a], [Bane03b], we extended this concept for GMPLS-enabled grooming networks with the transport and switching hierarchy including optical wavelength division multiplexing (WDM), and the time division multiplexing (TDM) [Tom03b]. We proposed the control model for GMPLS-XP [Tom03b], and analyzed the architectural and performance differences between the GMPLS-IX architecture and the architecture with multiple UNI/NNIs. The gird community was embracing the concepts of exchange points and the GMPLS-based optical exchange points were later demonstrated in a testbed presented in [Dijk04], [Prev06].

Basing our work on the essential capabilities of grooming networks, within the framework of network federations we further introduced the model of an extensible L1 infrastructure service, and within it, the resource allocation based on the visibility parameter. We studied the methods for dynamic resource management of the service, including topology adaptation and traffic routing. Our approach is inspired by MPLS-based traffic engineering concepts [Awdu99], which conceived and stimulated significant amount of research work in the last couple of years. The problem of connection provisioning in multi-granular optical networks is broadly studied with particular emphasis on graph methods for efficient traffic grooming, e.g., [Cink00], [Zhu03e], [Zhu04]. Similarly, in our approach to modeling network and service resources, we proposed a graph model based on the requirements of generic multi-layer network modeling [G805], and developed resource management strategies based on graph theory. Our graph model captures resource visibility, and can represent resources at inter-domain exchange points with service-specific features in mind. An important motivation for our research came from the results which demonstrated in several studies that multi-layer routing outperforms the single-layer approach, e.g., as in [Sabe03]. In our service model, we proposed two methods for routing and topology engineering (RToE): one using the single-layer approach, and the other using the multi-layer routing approach. In our performance study, we evaluate their benefits and limitations under dynamic traffic conditions in single and multi-service scenarios.

Continuous efforts of the optical community to establish optical virtual private network services, recently resulted in L1 VPN standardization framework [Take05]. This framework is currently under study in both IETF and ITU-T. A L1 VPN classification is provided to cover a broad spectrum of different provision, control and management scenarios, borrowing the general features of Layer 2 and Layer 3 VPN services, in particular from BGP/MPLS VPN [Rose06]. While our notion of infrastructure services belongs to a special class of Routing and Signaling VPN services, it features in addition a specific, dynamic topology maintenance as well as resource sharing. Concerning the practical advances in VPN provisioning, one of the most widely used tools today for experimentation in network topology configurations is the UCLP Version 2 (UCLPv2) management tool. This tool allows for the creation of arbitrarily, user-defined topologies called Articulated Private Networks (APNs) and is defined in [StAr06]. APNs allow many dynamically created networks to share a common network infrastructure consisting not only of links, routers, and switches, but also of instruments and end-user devices. Our approach to create infrastructure services aided with GMPLS TE mechanisms does not duplicate, but complements the UCLP approach. While UCLP is a management tool enabling service configuration in the management plane, the infrastructure service uses GMPLS control plane mechanisms for the dynamic resource management. Thus, they could co-exist and operate at different time scales. Recently, GENI also proposed a similar approach for virtualization with the focus on Layer 3 virtualization [GENI06]. Our approach can also be used in combination with GENI, as it merely controls the lower layers of an infrastructure and as such can facilitate virtual services in upper layers.

1.2 List of Publications Supporting this Thesis

[Tomi07] S. Tomic and A. Jukan, Performance Analysis of Infrastructure Service Provision with GMPLS-based Traffic Engineering, Special Issue on Multi–Layer Traffic Engineering, IEEE Journal on Selected Areas in Communications, July 2007, pp. 881–894.

[Tomi07b] S. Tomic, Control Plane Support for Traffic Grooming, Chapter II, in Traffic Grooming for Optical Networks, Editors: R. Duta, A. Kamal, G. Rouskas, Springer, to appear in 2007.

[Tomi05] S. Tomic and A. Jukan, GMPLS-based exchange points: architecture and functionality, Chapter VIII, in Emerging Optical Network Technologies Architectures, Protocols and Performance, Editors Krishna M. Sivalingam, and Suresh Subramaniam, Springer, 2005.

[Tomi04] S. Tomic, Issues of resource management in two–layer GMPLS networks with virtual network services, IEEE Global Telecommunications Conference, GLOBECOM'04, vol.3, Dec. 2004, pp. 1803–1807.

[Tomi04b] S. Tomic, B. Statovci-Halimi, A. Halimi, W. Mllner, J. Frhwirth, ASON and GMPLS, Overview and Comparison, Photonic Network Communications, vol. 7, no. 2, 2004, pp. 111–130.

[Tomi04c] S. Tomic and A. Jukan, Extensible Service Provision in Grid Networks: A Case For Resource Visibility and Inter-Domain Exchange, First International Workshop on Networks for Grid Applications, San Jose, USA, 29 October 2004.

[Tomi03a] S. Tomic, A. Jukan, Dynamic Provisioning over Multi-provider Interconnected GMPLS-enabled Networks, European Conference on Optical Network Design and Modeling ONDM 2003, Feb. 2003, pp. 365-383.

[Tomi03b] S. Tomic, A. Jukan, GMPLS-based Exchange Points: Architecture and Functionality, Workshop on High Performance Switching and Routing, HPSR'03, 24-27 June 2003, pp. 245–249.

[Tomi03c] S. Tomic, Issues of Two-Layers Routing in GMPLS Networks with Virtual Network Services, 8th European Conference on Networks & Optical Communications 2003 (NOC'2003), Vienna, Austria, July 2003, pp. 114-121.

[Tomi02b] S. Tomic and A. Jukan, MPFI: The multi-provider network federation interface for interconnected optical networks, IEEE Global Telecommunications Conference (GLOBECOM'02), vol. 3, 17-21 Nov. 2002, pp. 2365–2369.

[Tomi02a] S. Tomic, A. Jukan, Policy-based lightpath provisioning over federated WDM network domains, Optical Fiber Communication Conference and Exhibit, OFC'02, 17-22 March 2002, pp. 499–501.

[Tomi02c] S. Tomic, A. Jukan, Optical Virtual Private Network Services in the Multi-Provider Network Environment"; IP over WDM international conference, IPoW'02, Paris, France; June 2002; pp. 1- 19.

1.3 Thesis Organization

The thesis is structured in six Chapters. Following the Introduction, Chapter 2 provides an overview of the standards for automatic control of resources in multi-layer networks developed under ITU-T ASON and IETF GMPLS frameworks. Chapter 3 describes the proposed network federation architecture with GMPLS Exchange Points (GXP), highlighting the differences between the traditional architecture with static interconnections and the novel GXP-based architecture. A control model and functionality for GMPLS Exchange Points is presented, followed by our network graph model and the description of characteristic multi-domain provisioning scenarios in heterogenous architecture with both static interconnections and GXP. We further use the network graph model to state the GXP location problem and propose a heuristic to solve it. In Chapter 4, we present the Infrastructure Service model, and the concept for resource allocation and sharing based on the resource visibility attribute. We describe two graph methods for traffic Routing and Topology Engineering (RToE), which correspond to single-layer and multi-layer TE methods, and complementary stand-alone topology engineering (ToE) strategies for virtual link release. In Chapter 5, we present our simulation study and numerical results that quantify the impact of the proposed concepts, including the resource visibility and the GXP architecture. The conclusions and directions for further work are summarized in Chapter 6.

2 Control Plane for Virtualized Optical Networks

2.1 Introduction

Today, multi-wavelength optical networks build the foundation for the overall telecommunications landscape. Indeed, the advent of the wavelength division multiplexing (WDM) and optical switching have changed the way we think about bandwidth. With WDM, a large number of concurrent wavelengths, each capable of transporting a high-capacity traffic flow, can be established within the same optical fiber, immensely increasing the bandwidth of transport networks [Mukh97]. At present, systems with up to several hundreds of wavelengths are tested in labs; standard commercial systems offer the capacity of 40 wavelengths, with a single wavelength bandwidth as high as 40 Gbit/s [Rama06]. The innovations in optical signal switching that followed the advances in point-to-point WDM transmission essentially redefined the role of the optical technology from a point-to-point physical layer to a networking stratum with capability to provide low-cost high-capacity switched wavelength services - i.e., lightpaths - [Chla92]. Particularly the fully transparent all-optical network without wavelength conversion, which represents the most cost-effective wavelength service platform, but in turn imposes a challenging wavelength continuity constraint on the wavelength assignment and routing, was in a focus of extensive studies. The problem of the wavelength assignment and routing in all-optical networks without wavelength conversion capabilities has been shown to be NP-complete [Chla93], which stimulated work on numerous practical approaches [Rama95], [Mukh96], [Subr97], [Baro01], [Juka00], [Zang00], [Rous01]. Extensive attention was also given to the problem of extending the general model for blocking probability in circuit switched networks [Gira90] to wavelength switched networks without or with sparse wavelength conversion [Barr96], [Birm96], [Yate97].

On the other hand, due to the concept of the digital wrapper [G709], the opaque networks with full wavelength conversion gained a lot of attention, as they largely simplify network resource management, and efficiently deal with signal impairments inherent to the optical transmission.

The demonstrated cost-benefits of optical networking, and on the other hand, specific bandwidth and quality requirements of today's applications such as data transport and real-time multimedia streaming, is what further motivated the work on sub-wavelength switching technologies. Sub-wavelength services bridge a gap between a huge bandwidth of a wavelength (in terms of Tb), and the bandwidth that applications require, typically ~100Mbit/s up to 1Gb/s. Only high capacity grid applications, such as those that need to collect measurements from distributed high-resolution sensory equipment produce traffic flows in terms of several Gbit/s [Karm05].

Within the optical layer, three major switching technologies capable of supporting sub-wavelength services compete for primacy: the *connectionless* optical packet switching (OPS) and optical burst switching (OBS), and the *connection oriented*(CO) switching of time division multiplexed (TDM) signals [Mukh97]. With OPS, the IP packets are mapped into optical packets which are statistically multiplexed and switched by optical routers [OMah01]. Presently however, a number of issues, such as processing of packets at the line speed and implementation of optical buffers for replacing currently used optical delay lines, used to store the payload information during processing of the packet header, are still to be resolved in order to make the deployment of optical packet switching networks technically and economically viable and deployed beyond the demonstrators and test beds [Rama06]. The burst switching paradigm is another promising approach for IP-optical integration [JJue05]. In OBS networks, the payload data, e.g. IP packets, are mapped within optical bursts and the control data related to each burst is sent on the optical channel preceding the burst. The payload burst is delayed relative to the control information at the edge of the network so that each OBS switch along the path can process the control data and configure the optical switching fabric for the appropriate next-hop connection before the payload burst arrives. With the optical burst switching the connection is set up link-by-link as a control packet progresses towards destination. The traffic burst follows the control packet using configured fabric connections; upon the successful burst transmission, each connection can be immediately released. OBS technology promises lower cost as the signal is kept in the optical domain, and a new QoS-enabled paradigm [Yoo01]. However, it seems that to make OBS a carrier-grade technology, some practical aspects will need more attention and better understanding [Rama06].

Both OPS and OBS technology have been shown to have significant potential [Klon05], [DeLe06], [Qiao06], particularly OBS in the context of the optical GRID [Sime05]. At present, however, the sub-wavelength services of the strongest practical interest in carrier and grid applications, e.g., [Veer06], are those offered with the optical transport network architecture (OTN) [G872] that combines WDM transmission and optical wavelength switching with the digital grooming and switching. OTN merges the traditional digital transport network technology, such as Synchronous Digital Hierarchy (SDH) [G803] and Synchronous Optical Network (SONET) [T1.105], with the reconfigurable optical layer, and offers flexible mapping, transport and switching of different bandwidth granularities in a new opto-digital hierarchy. The optical transport network technology also includes innovative concepts for service end-to-end monitoring and control based on the dedicated digital overhead.

The cost benefits of hybrid connection-oriented multi-layer networks, which combine WDM switching with traffic grooming technology, e.g., the IP or TDM layer, provide strong motivation for broad research work on optimal traffic grooming [Dutt05]. Two classes of traffic grooming problems have been addressed: the static grooming and the dynamic grooming. Static grooming is an optimal network design problem, where for known traffic requirements the cost of the physical topology, the deployed grooming capacity and physical links, are minimized. The dynamic grooming aims at optimal routing of generally unknown and dynamic traffic in a network of known physical topology and capacity. The dynamic grooming problem can be stated as follows [Zhu02a]: For a network characterized with the physical topology of transmission links and nodes with specific multiplexing and switching capability at the wavelength and sub-wavelength layer, including, the number of transceivers at each node, number of wavelengths on each fiber, and the capacity of each wavelength, map a given set of connection requests with different bandwidth granulations, such as OC-12, OC-48, etc., into a minimal set of lightpaths terminated at the transceivers of the grooming nodes.

Central to the grooming architecture are hybrid nodes that can terminate or switch either lightpaths or SONET/SDH traffic, or both, over single or multi-hop paths. Hence they are classified into four classes: single hop grooming OXC, which can multiplex TDM signals into a wavelength but can switch only at WDM layer, multi-hop partial grooming OXC, with WDM and TDM switching fabrics, de-multiplexing and dropping only a limited number of wavelengths into a grooming fabric for further signal switching or dropping at the TDM layer, and multihop full grooming OXC with opaque technique which drops all wavelengths and switches only at the TDM layer [Zhu03c]. Static grooming problem has been shown to be an NP-hard problem, as it includes three correlated problems, including NP-hard virtual topology design ([Mukh97]), NP-hard lightpath routing and wavelength assignment (RAW) ([Chla93]), and the multi-commodity-flow-based routing of traffic on the virtual topology. It was first formulated and studied in the context of reducing the cost of ring networks [Gers99], [Gers00], [Berr00], [Zhan00], [Chiu00]. [Wang01], [Wang02], [Dutt02]. With the transition of the core networks from the ring topologies to mesh topologies, the mathematical approach devised for the problem of grooming in ring networks has been extended for general topologies [Zhu02a], [HuJQ04], [Yao05c], [Huan06].

The emerging of the control plane for optical networks, first under the term Multi-protocol-λ-Switching [Awdu01], and than under the term Generalized MPLS [Mann04], was an important milestone towards new optical network *services*. Generalizing the MPLS model of a *label* and a *Label Switched Path (LSP)* for different technologies, and introducing the LSP hierarchy within which multiplexing of signals of different types and granularity can be represented, has opened a new field of application for standard MPLS signaling and routing protocols extended for new requirements. Facilitating this synergic effect of unification and reuse, the proponents of the unified control plane concept have demonstrated that the traditional transport network management can be significantly simplified and improved with the real-time resource control, which also enables a new breed of dynamic and *automatically switched* connection services. The control plane approach to service provisioning opened the scene for dynamic traffic grooming and multi-layer traffic engineering methods, which aim at solving the grooming problem for one connection request at a time. With these methods, connections are dynamically routed on a resource graph that captures resources at different layers of the grooming hierarchy [Cink00], [Zhu03e], [Sabe03], [Vigo05]. Essential to the graph-based approach is modeling of transceivers at each node, wavelengths on each fiber-link, wavelength-conversion capabilities, and grooming capabilities. Therefore, the control plane for grooming networks needs to support such resource modeling and the creation of a resource graph for the on-line resource management.

The work on the automatically switched transport network concept is still strongly driven by the need to reduce the network operation cost (OPEX), and the potential of the GMPLS control plane to act in the "optimizer" role. However, emerging of new bandwidth-greedy applications, such as those which today transport masses of visual data on the scientific grids, will more and more put the stress on the service-enabler role of the control plane [Ghan03], and on services and their requirements [Alan04], [Alan06]. With the growing base of the "prosumers" (the private content producers and consumers) on the net, these applications are expected to grow in number and kind, and will change the landscape of existing communication services. Ultimately, the combination of large-capacity, flexible granularity, and intelligent control of L1 services, which the new optical transport networks will stand for, shall be a foundation for a novel cost-effective networking, supporting many different closed-groups or communities, with different L2 and higher protocol preferences, transporting traditional voice, Internet data, voice-over-IP, transactional traffic, and others. In other words, the automatic switched transport networks will provide a re-configurable and "virtualizable" physical layer for higher layer services and applications, with novel business models including virtual providers, and carrier carriers models.

Migration of the traditional L1 services, the most basic of which is a static leased line, towards novel L1 virtual private network (VPN) [Take05] that establish and dynamically adapt bandwidth and connectivity of their virtual topologies by setting-up or releasing switched end-to-end connections over the physical infrastructure, bears many new challenges. We see these challenges related to two important requirements: first, to provide architecture and mechanisms for establishing an automatically controlled common L1 infrastructure in which the resource management of each connected carrier relies on the common set of automatic control protocols, and second, to design efficient virtualization mechanisms operating on the common infrastructure. These two requirements were also the motivation for the work presented in this thesis. It should be noted, however, that the challenge of creating a common infrastructure has different complexity in a *collaborative* environment such is a scientific GRID, and in a *competing* environment of telecom providers. The collaborative environment is based on the symmetry between the total benefits, and the total cost for each of the participant. As the total benefit may increase with the integration, and the total cost may be improved with GMPLS-based network control, there is a significant immediate interest in GRID community for advancing this technology [Veer06]. In the competitive carrier environment, on the other hand, a requirement for integration of infrastructures and support for global services is stated within the standardization framework on Automatic Switched Transport Networks (ASTN) of ITU-T; however, advancing the understanding of new practical business models for the GMPLS-based integrated network is still a long-term objective.

The remaining of this Chapter provides an overview of the transport and control plane enablers for virtualization that exploits the opto-digital circuit switching. Section 2.2 provides a brief overview of the traffic grooming technology. Section 2.3 reviews the major concepts of the control-plane centric optical network control and the concepts of Layer 1 Virtual Private Network, relevant for the definition of infrastructure services are revisited. Section 2.4 concludes the Chapter.

2.2 Traffic Grooming Technology

Optical transport network technology that combines wavelength switching with the sub-wavelength switching based on the time division multiplexing (TDM) and the digital electronically processed frame, offers today an efficient and quality-guarantied networking solution for a global network in which the core, the metro, and the access areas have inherently different requirements regarding carried bandwidth. Because a single wavelength can transport capacity of an order of 40Gb/s, pure wavelength switching can be seen as a well positioned in core networks, which carry large aggregates of traffic. Metro networks, on the other hand, must offer services of finer granulation. Presently, two network technologies that combine WDM with TDM are considered as candidates for the high performance multi-service network. These two technologies primarily differ in the way in which client signals- primarily SONET/SDH - are mapped into the optical layer. The first technology directly transports SDH/SONET hierarchy over the wavelength switched network; the second one, developed under the ITU-T Optical Transport Network (OTN) [G872] framework uses the digital wrapper - i.e., a new digital layer - to map SONET/SHD and other clients within the optical signal. The benefits of the second approach are numerous. The dedicated overhead of the new digital layer supports client-independent monitoring, which allows for the transparent transport of all clients without the need to reuse the parts of the client signal overhead. In terms of bandwidth management, OTN borrows from the concepts introduced with the new generation SDH, and extends them for the optical layer. In SDH and SONET hierarchies, connections of finer granularity - i.e., lower speed - are multiplexed together to support services of courser granularity and higher speed. Traditional *contagious concatenation* scheme supports bundling contagious containers according to the hierarchical multiplexing schemes of SONET/SDH into the pipes of courser granularity and higher speed. This approach, designed for voice services, renders inefficient for the transport of data services. To support requirements of data traffic new virtual concatenation and inverse multiplexing [G7043] have been introduced and form foundation for next-generation SDH/SONET standards. By means of *virtual concatenation*, containers of higher granularity can make a group of a flexible size. With the *inverse multiplexing* a high-speed connection is split over the lower speed channels and concurrently transported along several paths from the source to the destination. At the destination, the differential delay between the sub-frames transported over different paths needs to be accounted for, by buffering and by re-assembling the signal based on the enumeration of the sub-frames. A new Link Capacity Adjustment Scheme protocol, [G7042], supports flexible management of a group of virtually connected containers (connections) forming a single TDM-based call. With LCAS, the capacity of a call can be changed by adding or deleting connections in a dynamic way. LCAS includes procedures for extension and reduction of the call capacity and automatic state update. VCAT/LCAS provide for a number of new network features. These are flexible concatenation of containers at the right granularity (e.g., 100 Mb/s Ethernet is equal to VC-3-2v, or STS-1-2v), virtual grouping of component links possibly realized over multi-hop paths taking different routes, bandwidth on demand and IP traffic engineering where new capacity is added to the congested IP links keeping the IP layer topology and routing stable. In addition, a new component link can be routed and set up along a new path before the old component is released. This is denoted as painless re-grooming [Bern06]. Complemented with LCAS, VCAT can support realization of new forms of protection/restoration and graceful degradation based on dynamic repair.

Although new generation SONET/SDH technology also uses WDM links for cost effective transport, it does not support cost effective switching of high capacity aggregates. As a solution to this problem, ITU-T OTN specifies a new digital layer frame (the digital wrapper), in which a variety of data services, including SONET/SDH, Ethernet, ATM, IP can be mapped. OTN defines new hierarchy based on 2.5Gbit/s, 10Gbit/s and 40Gbit/s signals, and specifies their flexible hierarchical multiplexing. The OTN digital hierarchy includes three digital layers with distinctive functionality, the so-called optical channel payload unit (OPU), responsible for the mapping of the client signals into the payload containers, the optical channel data unit (ODU) which is responsible for the service monitoring, and the optical transport unit (OTU) responsible for the reliable transport of 2.5 Gbit/s,10 Gbit/s and 40 Gbit/s signals. With the switching of coarse granular aggregates OTN efficiently responds to the transport requirements in the core. On the other hand, finer granularity offered by SONET/SDH provides answers to particular service bandwidth requirements. Similarly to the next generation SDH/SONET, OTN incorporates virtual concatenation, inverse multiplexing, and flexible bandwidth adjustments scheme in both the digital and the optical layer. This means that the payload can be packed into a virtually concatenated group of digital signals, which may be transported on different wavelengths and along different paths. The OTN technology also introduces the capability of flexible adaptation where a the node can be flexibly configured for different signal mappings.

2.3 Control Plane for Grooming Networks

Novel transport networks with opto-electronic multiplexing and switching can use static or dynamic traffic grooming to cost-efficiently resolve a mismatch between the high capacity of wavelengths, and the low bandwidth requirements of predominantly IP or Ethernet services: The multiplexing and switching at different granularity layers in optical or digital hierarchy can be either *selectively deployed* to achieve cost-efficient network design for anticipated multi-granular traffic (static grooming), e.g., as discussed in [Ceru05], or *selectively used* to maximize network throughput for dynamically changing traffic (dynamic grooming) [Zhu03b]. A distinguishing feature of grooming networks is *flexibility* - only the lowest layer in the hierarchy has a static topology; at each higher layer, the topology may be virtual, and can be changed - i.e., *engineered* - by setting up or releasing end-to-end connections in the underlying layer.

In order to exploit this flexibility when the traffic is dynamically changing, topology engineering requires an automatic process that combines traffic monitoring and fast network re-configuration. Today, the tools for fast network reconfiguration are being standardized within the Generalized Multi-protocol Label Switching (GMPLS) [Mann04] framework of the Internet Engineering Task Force (IETF), and the Automatic Switched Optical Network (ASON) framework of the International Telecommunication Union-Telecommunication Sector (ITU-T) [G8080].

These two standardization frameworks define a new paradigm for service provisioning and resource management in multi-layer networks, the core of which makes a transition from a centralized management-plane (MP) approach to a control-plane (CP) approach to operating transport networks. CP-approach is based on automatic and distributed control functions, and therefore it promises higher scalability and robustness, and lower operational costs. In fact, OPEX reduction on the order of 50 % can be expected for most telecom operator models [Pasq05]. CP-approach also requires a unified resource control model applicable to multiple layers.

In this respect, GMPLS proposes a control plane model with a five-layer hierarchy of switching capabilities for *IP-over-WDM* converged infrastructure. The model includes packet switching (PSC), layer-2 switching (L2SC), time division multiplex switching (TDM), lambda switching (LSC), and fiber switching (FSC) [Komh05]. At the same time, ASON defines a model for optical transport network technologies with a "rich" TDM switching layer *(IP-over-OTN)*, including synchronous digital hierarchy (SDH) [G803], synchronous optical network (SONET) [T1.105] and optical transport network (OTN) [G872]. The different scope of these two frameworks resulted in several differences between their models. On the one hand, the optical channel (OCh) switching capability of the OTN hierarchy is mapped into the lambda switching capability (LSC) of GMPLS; on the other hand, switching at different path layers of SDH/SONET, as well as at the digital path layers of the OTN hierarchy maps into only one interface type being TDM switching capability. This has implications on network and service models, for example a topological representation of a TDM layer in GMPLS - a TDM Region [Komh05] - may include different SDH/SONET/OTN interfaces. These issues are thoroughly discussed in [Tomi04b].

GMPLS and ASON classify the core control plane functionality within three major processes, which support dynamic and automatic provision, and eliminate “human factor” often found responsible for misconfigurations and provision delays. These processes are:

- Connection and call control (signaling)
- Distribution of network information and path computation (routing)
- Automatic resource discovery and inventory

Within this scope, the ASON framework particularly focuses on the control plane architecture and protocol neutral functionality, including special purpose controllers and their relationship. GMPLS, on the other hand, defines protocols for realization of CP-functions. GMPLS extends two MPLS routing protocols, namely the Open Shortest Path First extended for Traffic Engineering (OSPF-TE) [Komp05a], and the Inter-System to Inter-System extended for Traffic Engineering (ISIS-TE) [Komp05b], and two signaling protocols, the Reservation Protocol extended for Traffic Engineering (RSVP-TE) [Berg03a] and the Constrained Label Distribution protocol (CR-LDP) [Ashw03].

Signaling extensions [Berg03b] support distributed configuration of circuits of different types, refereed to as *generalized label switched paths (LSP)*. This includes reservation of link resources allocated to LSPs (e.g., wavelengths or a time slots), realized by means of the link-local distribution of generalized labels, and configuration of switching fabrics along the path.

Routing extensions [Komp05e] support distribution of TE attributes for GMPLS TE links, e.g., by periodic advertisements, as a support for distributed path computation. Regarding routing and path computation requirements, a Path Computation Element Framework within the IETF currently defines a new architecture [Farr06] for a routing process decoupled from GMPLS protocols.

Link Management Protocol (LMP) [Lang05] supports the discovery and configuration of the transport and control plane interfaces and their mapping.

Within GMPLS the current work focuses on specific requirements of IP-over-OTN grooming networks, which are superior to IP-over-WDM networks in terms of scalability, flexibility, and robustness as shown in [Seng03]. The objective of this work is to fill in the gaps between the generic functionality supported in GMPLS so far, and the required functionality, often assumed in theoretic traffic grooming studies. This is an essential task, because the applicability of the traffic grooming methods, in particular, ***multi-layer routing algorithms*** using specific network graph models, such as those proposed in [Zhu03e], [Zhu04], and [Cink00], largely depend on the availability of the network state information used to construct the routing graph. The required availability can be assumed as fully given with a centralized management system; on the other hand, the distributed CP-based provisioning must rely on information appropriately aggregated for scalability.

2.3.1 Grooming with GMPLS

Grooming with GMPLS is based on mechanisms and protocols for resource-optimal configuration of flexible grooming capability, and for optimized traffic routing. Figure 2.1 depicts the required control plane functionality within the framework of three functional blocks, namely, the network resource control, call control and virtual network control. So far, while GMPLS protocols provide concrete tools for resource control, and call routing and control, the requirements for virtual network control are under study and consideration.

Figure 2.1 shows a general network scope for CP-based operation. This is a multi-layer multi-domain network, operated by different carriers, and providing services of different granularity. The ASON architecture makes a distinction between three types of characteristic CP-interfaces: the user-to-network interface (UNI) between a service user and a

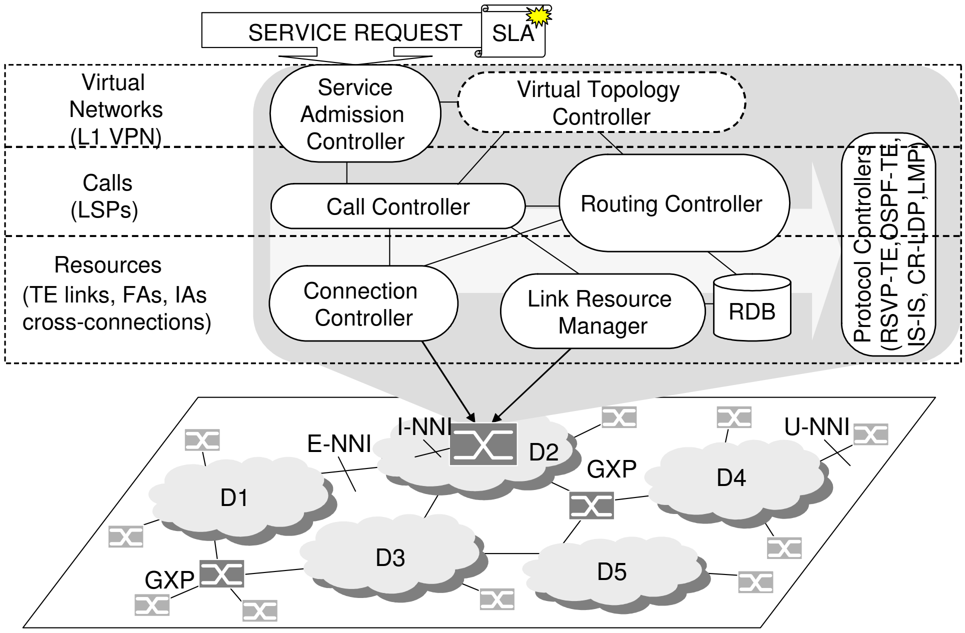


Figure 2.1: *Framework for GMPLS Grooming Support.*

carrier domain, the internal network-to-network interface (I-NNI) between the network elements within one carrier domain, and the external network-to-network interface (E-NNI) between the network elements of different carrier domains. These interfaces differ in the level of network information exchange they may support: while internal I-NNI supports full information sharing, UNI and E-NNI often require information hiding.

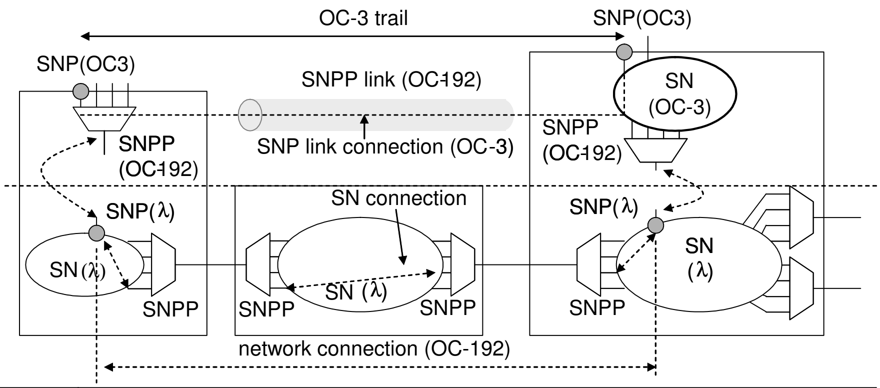
Conceptually these interfaces relate to the GMPLS models for the control plane integration which are the ***overlay model*** with no routing information exchange, the ***peer model*** with full information sharing facilitating integration of control plane instances, and the ***augmented model***, assuming the exchange of aggregated information [Mann04]. In Figure 2.1, we show the standard multi-domain architecture enhanced with challenging new elements, referred to as GMPLS-based exchange points (GXP), which enable both flexible inter-domain connections between different administrative domains and customers, and their GMPLS-based control. While the GXP concept is described in Chapter 3, the sections that follow review functionality inherent to different blocks of the framework.

2.3.2 Resource Modeling

Due to scalability and trust issues, service provisioning by means of the control plane has to rely on aggregated information, and the appropriate aggregated representation of resources is essential. GMPLS and ASON developed specific network resource models by using two different set of architecture tools.

ASON Model

The ASON resource representation is based on the generic functional model for transport networks (G.805), [G805], and is inherently "topological". A network is modeled with a number of layer topologies in a client-server relationship. An example in Figure 2.2 illustrates the ASON model for network with an optical channel (OCh) switching layer and the OC-3 switching layer. The relationship between GMPLS and ASON models shown in the table is briefly covered in the discussion that follows.



	Corresponding Artifacts in ASON and GMPLS models				
ASON	SNPP Link	SNP	SNP-SNPP Association	Network connection / Trail	SNP Link Connection
GMPLS	TE Link, FA	ISCD (External Interface)	IACD (Internal Interface)	LSP (Connection / Call)	Generalized Label

Figure 2.2: *Example of ASON Modeling and Relationship between ASON and GMPLS.*

Each layer topology provides a switched service between two client access points, which is routed over interconnected ***sub-networks (SN)***, where the switching matrix is represented as an SN as well. The interconnection between two SNs is a ***link*** with one or multiple ***link connections***. The end-points of link connections are subnetwork connection points (SNP). Logically coupled SNPs form a pool (SNPP), e.g., all OC-3 within a OC-192 multiplex are coupled within one pool, Consequently, a transport capacity between two SNPPs of two sub-networks, or between an SNPP on a sub-network and an access SNPP (see the case shown in Table 2.2), is referred to as an ***SNPP link***. An ASON service is an ***SNP trail***, which connects two client access points, and is routed over a number of SNP link connections or subnetwork connections. The association between an SNP in a server layer (optical channel - λ) and an SNPP (of OC-3 SNPs mapped within OC-192) in the client layer represents a specific ***termination and adaptation*** between the two layers in the multi-service switching node. With variable adaptations that can be flexibly activated several different associations can exist at the same time.

ASON is particularly designed for discrete bandwidth networks: an *atomic transport resource* is an ***SNP link connection***. Obviously, in terms of routing requirements, advertising SNP link connections would not scale, therefore SNPP link is to be advertised as an aggregate characterized by the number of available SNP LCs. An association SN-SNPP uniquely identifies an ***atomic grooming resource*** which must be allocated along the multi-layer routed path, and therefore, should also be advertised to support distributed path computation. As already mentioned, the ASON resource model is based on SDH/SONET/OTN functional modeling and extensions to GMPLS were needed to account for a rich TDM layer.

GMPLS Model

Several ASON subnetwork points (SNP) and/or point pools (SNPP) map to a GMPLS ***interface***, which constitutes the termination of a ***GMPLS TE*** link. A GMPLS interface can be either a simple interface (SNP) or a combination of component interfaces (SNPP). To cope with the multiple parallel interfaces between two nodes the interfaces can be unnumbered [Komu03]. For some technologies a component interface is an atomic resource, e.g., a wavelength, and is addressed with a ***generalized label*** which corresponds to an

SNP link connection (LC). For SDH/SONET and OTN digital hierarchy a component interface has a multiplex of smaller discrete resources, each of which is described with essentially more complex generalized label. One interface can therefore represent several SNPs or SNPPs at different ASON layers, each of which can be represented, within a GMPLS model, with the GMPLS Interface Switching Capability Descriptors (ISCD). ISCD includes a switching type (LSC or TDM) and maximum and minimum reservable bandwidth available, the latter depending on the adaptation configured at the interface.

Examples of ISCD usage are given in Table 2.4. An interface of a photonic cross-connect is fully described with LSC switching (ISCD-1). In addition to ISCD-1, an interface of a hybrid (SDH-WDM) cross-connect is described also with ISCD-2 and ISCD-3 (TDM switching) defining VC-4 and VC-3 as a minimum reservable bandwidth which can be allocated at this interface. However, if VC-4 is allocated ISCD-2 will disappear from the ISCD list.

The GMPLS interface model becomes rather complex when used to represent capabilities of multi-service switching nodes. The concept of a GMPLS TE link is not meant to be used to model internal multiplexing capability of nodes, although in fact, SN-SNPP association could be modeled as TE link with one interface in a client layer and the other in a server layer. Without such modeling, the available grooming capacity of a node cannot be used to calculate a TE path for a service. Within GMPLS, the concept of the *Interface Adaptation Capability Descriptor (IACD)* is currently under evaluation [Papm05] and experimental verification [Veer06]. IACD is a new attribute defined for an *internal interface* which could be advertised with GMPLS routing protocols, IACD describes an interface in terms of two ISC Descriptors (for both ends of the interface or link) with the standard bandwidth encoding depending on the ISC type.

Photonic XC	Hybrid XC	ISC Descriptor	ISC Type	Encoding	Max BW	Min BW
X	X	ISCD-1	LSC	SDH	STM-64	-
-	X	ISCD-2	TDM	SDH	STM-64	VC-3
-	X	ISCD-3	TDM	SDH	STM-64	VC-4

Table 2.1: Multiple Interface Switching Capability Descriptors.

Client-Server Resource Relationship

Both ASON and GMPLS define a *client-server relationship* between adjacent layers in their hierarchies. In the ASON model, the link in the client layer is realized over a network connection in the server layer. A client layer link is also a part of an end-to-end network connection supporting the client layer trail. A GMPLS client-server relationship associates a dynamic TE link in a client layer, referred to as a Forwarding Adjacency (FA), and a label switched path (LSP) in a server layer, established between two interfaces that implement client and server switching types. By using the dynamically established FA, a client layer LSP can embed itself within the server layer LSP. For routing scalability, all parallel resources with the same TE capability could be bundled into, and advertised as, one bundled TE-link or FA [Komb05]. Each TE link or FA is also assigned to some shared risk link group (SRLG): all links that fail together are assigned the same SRLG value. Also, each link is assigned a protection level such as extra traffic, unprotected, shared, and dedicated 1:1, or 1+1.

2.3.3 Resource Control

The functionality of the control plane processes within the functional block ***Resources*** can be shortly summarized with *"discover, advertise, reserve"*, the issues of which are described in the following sub-sections.

Resource Discovery

For automatic correlation of interfaces between either two neighboring nodes or neighboring technology layers within one node, an automatic discovery process is needed. This process can therefore support building up of the network topology graphs. Within the discovery process both transport connections and control channels are to be discovered and verified. A central component in the ASON discovery process is the link resource manager (LRM), which maintains the local inventory of links and updates a local routing database

(RDB) with configured TE links. Although, the routing database can be configured also through a management system, the complexity of its update procedures, best illustrated in the MIB document [Nada07], offers strong motivation for automation of the discovery process. Within the discovery process, the binding between the management plane (MP) names of the transport resources (modeled with G.805) and the control plane (CP) names of the same resources (SNPs) is established. The MP names and the corresponding CP names of potential link connections (SNP-SNP) between the neighboring nodes are logically associated together. A potential SNP-SNP link connection becomes an actual LC when a corresponding flexible adaptation is activated. In the GMPLS framework, the link management protocol (LMP) [Lang05] supports the resource discovery process with four basic functions for a node pair. The control channel management establishes and maintains connectivity between adjacent nodes. The link verification procedure verifies the physical connectivity. The link summary messages are exchanged to correlate link properties between adjacent nodes, first when the node is being brought up and then periodically when a link is up and not in the verification procedure. Finally, LMP provides a mechanism to isolate link and channel failures in both opaque and transparent networks, independent of the data format.

Resource Advertisement

The advertisement of network topology data - i.e., the TE link information - that is maintained within the routing database (RDB), is a process that enables distributed path computation for requested services. For the purpose of scalability and topology hiding, ASON proposes that at each node, the TE link information, acquired either through the advertisement process (external TE links), or established by the discovery process (internal interfaces), be organized within multiple ***routing areas***, or views, which filter data based on different operational constraints.

The filtering could be based on a technology layer (e.g., TDM, LSC), administrative aggregation level (inter-domain, intra-domain), shared risk link group (SRLG) identifier, or some other type of administrative grouping. Filtering can also be subject to operational constraints related to a type of an interface (e.g., UNI, I-NNI, E-NNI) and the CP-integration models (overlay, peer, augmented). Regarding the organization of the CP-functions, the advertisement process at a particular hybrid node, may involve a number of routing controllers logically associated with different routing areas and collocated at the same node.

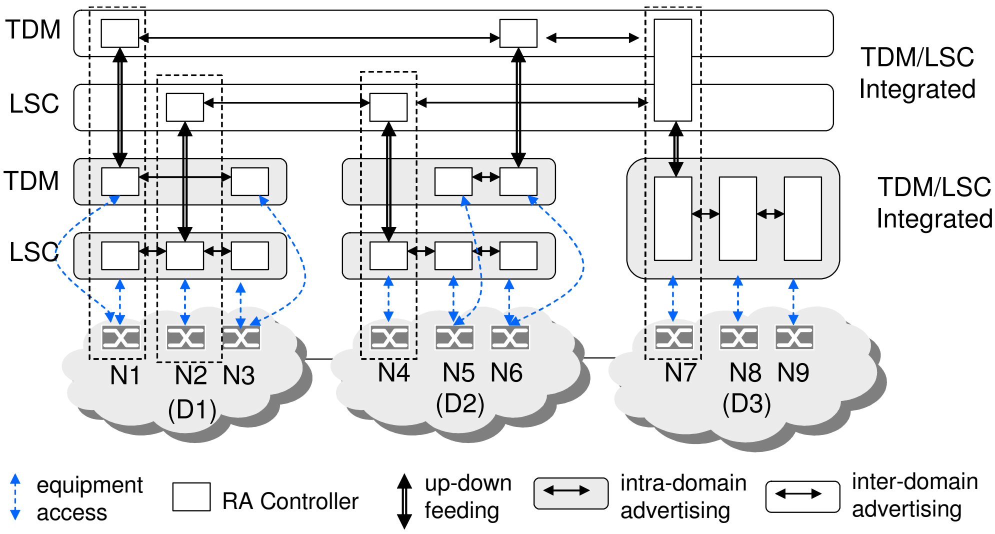


Figure 2.3: *Hierarchical Routing Areas in Multi-Domain TDM-LSC Network.*

For example, Figure 2.4 shows the hierarchical architecture of the advertisement process in a network of three interconnected domains that implement the LSC and TDM switching. In domains D1 and D2 advertisement of the TE attributes of the LSC and TDM layer is separated (overlay model) and in D3 integrated (peer model). The controllers at nodes N1, N6 and N7, and N2, N4 and N7, take part in the inter-domain information exchange at the TDM and LSC layers, respectively. Controllers at the same node exchange routing information by feeding-up or feeding-down data; each of them also communicates with its intra-domain peers located at other nodes. The vertical information flow may support different routing paradigms, such as the hierarchical, and the source-based routing. For *hierarchical* routing only feeding up is necessary because the path is calculated top-down, starting from the highest possible level. For the source-based routing, feeding down is needed as the routing decisions are made at the lowest level of the routing hierarchy.

Driven by ASON routing requirements, proposals extending the OSPF-TE and ISIS-TE protocols for hierarchical operation are currently under study [Papr07], with feed-up and feed-down capabilities between the hierarchically organized OSPF-TE and ISIS-TE areas. Hierarchical routing in ASON can also be supported by routing protocols that do not belong to the GMPLS framework. For example, the extension of the PNNI protocol, originally defined for ATM networks was proposed in [Sanc03].

An important aspect related to the TE link advertisements in the dynamic GMPLS networks is that the advertised changes not only should reflect the update of the TE link attributes due to resource allocation - i.e., residual bandwidth - but they also need to show the creation or deletion of dynamic TE links (FAs), and to alow for links with zero bandwidth. However, as currently defined, a virtual link (FA) is always allocated a bandwidth. We recognized this problem in our study and introduced the virtual link as a link that can have no bandwidth at all. Recently, this approach, has been also proposed within the GMPLS framework, and is currently under study [Papm05].

Link bundling and advertisement of bundled links is also a challenging task because of the parallel links with different TE features. Therefore, taking into account the dynamics and the nature of advertised changes, the process of advertising in grooming networks needs new consideration. The latency and accuracy of data updates, the issue of information hiding, the impact of distributed resource reservation and others, are related issues of interest, as also reflected in recently published studies, e.g., [Szig04], [Szig05].

Distributed Resource Reservation

With GMPLS the resources on the TE links along a particular service path - i.e., labels - are selected and allocated in a link-local hop-by-hop distributed resource reservation (signaling) session. In an RSVP-TE session, Path message is sent from the LSP ingress controller to the LSP egress controller, carrying objects, which define connection end-to-end requirements, such as the signal type and required bandwidth, and objects that have link-local meaning and support resource/label selection process. The explicit route object (ERO), which is provided by the path computation controller at the source of LSP, is in general also a link-local attribute. In other words, it reflects only the knowledge of the specific node that sends this message, and could be changed at each intermediate controller. In this way, a number of controllers can be involved in a hop-by-hop distributed path computation.

On each link, the label selection can be supported with *the explicit label* object, where an upstream node provides a label preference to a downstream node, to be accepted or rejected. The upstream node can start configuring its hardware with the proposed label in advance, which in case of accepted label can reduce the setup latency, particularly important when restoration LSPs need to be rapidly established. With *a label set* object an upstream node can suggest several preferred/acceptable resources for connection. In the optical domain, by means of a label set, a common available wavelength along the whole path may be determined. An *Upstream label* is used in the allocation of bi-directional links for bi-directional LSPs, with both directions following the same path and having the same traffic engineering attributes, including the protection and restoration level. The two directions of a link are allocated at the same time and therefore the setup latency and the control overhead are equal to those of a unidirectional LSP setup.

The allocation of time slots within SDH/SONET and OTN hierarchies is supported by new GMPLS extensions which encode generalized labels and describe both the multiplexing structure and the position of the signal in the multiplex [Mann06f], [Papa06f]. With these extensions to GMPLS, one SDH/SONET/OTN LSP can require allocation of several labels, by specifying a number of multiplex signals and the structure of this elementary multiplex, given with a number of elementary signals either contiguously or virtually concatenated. With fixed adaptations a signaled request may result into a number of LSPs set up first in the lower layers (e.g., ODU3) and then in the requested layer (e.g., ODU1).

However, configuration of flexible adaptations is not supported in a straightforward manner since a label request alone does not specify the multiplexing structure. Even assuming that the internal adaptation capabilities are advertised and used to create a routing graph for ODU1, ODU2 and ODU3 layer, transforming a request into a specific ERO and activating flexible adaptations is a challenging task, which is still not fully supported.

2.3.4 Service Control

Similarly to other connection oriented networks, the automatic switched optical network offers a connection service, which is a *logical* association - i.e., a call - established between two customers, *physically* realized by means of ***connections***. A call can be associated with zero or a number of connections that may be established, released, modified, and used for different purposes. In particular in transport networks where restorability and bandwidth adaptability are *service properties*, the control plane support for bandwidth modification, and protection/restoration of a call are of major interest. Accordingly, the separation between a call and a connection is identified as a major requirement, and extensions to GMPLS that support a call with no connections are under consideration. Regarding bandwidth modification, the aggregation of multiple connections within one call needs to be implemented within the transport plane; on the other hand, the control plane must be capable of discovering and controlling this capability. Figure 2.4 illustrates this issue. A call in the topology shown can be provisioned over contiguously concatenated signals

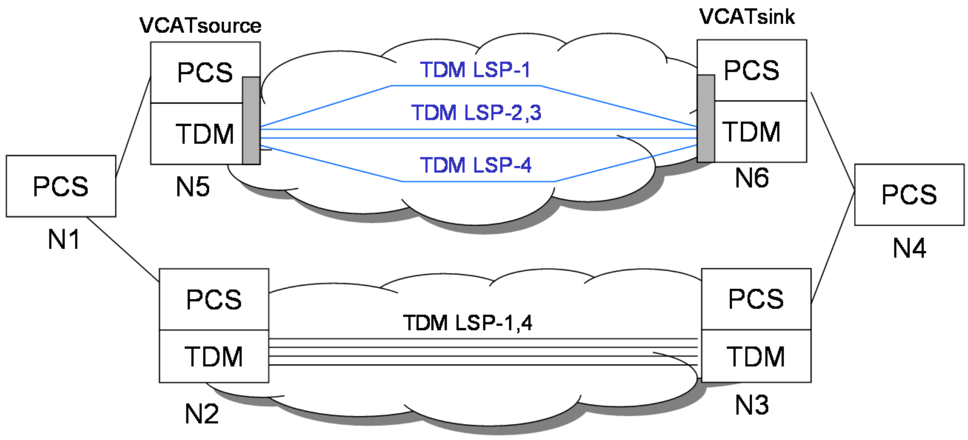


Figure 2.4: *Path-dependent vs. Path-independent Routing of Call Components.*

over a path N1 - N2 - N3 - N4, or as a virtually concatenated signal split over different paths between N5 and N6. If the availability of source and sink VCAT functions can be discovered, the path-independent routing of call components could be supported. Here, path-independent routing refers to routing of different connections along different paths. In the context of path-independent component routing, there is no single path associated with a call. In general, any call components, can traverse two UNI interfaces, and one or several different administrative domains, i.e., E-NNI interfaces. In grooming networks, work on enhancing GMPLS/G.ASON for VCAT/LCAS has recently been undertaken [Bern05], including the discovery of VCAT and LCAS capable path termination sources and sinks, as well as the VCAT group identification.

Distributed call signaling and connection path computation are tightly intertwined: how and where a path is calculated depends on the organization of the routing process and the corresponding availability of routing information. As already mentioned, ASON routing assumes hierarchically organized routing areas (RA) which feature hierarchical advertisement of information between different layers of aggregation (feeding-up and down) and thus support hierarchical inter-domain routing.

Routing controllers within routing areas at the different levels of hierarchy build a topology of abstracted resources - i.e., virtual links and nodes - and can provide path computation services to call controllers at the customer side, according to a hierarchical, source-based or hop-by-hop routing paradigm.

The setting up of a call over several domains, which includes setup of several call segments, as well as the error recovery by means of crank-back, must be carefully orchestrated. To this purpose the GMPLS Notify message (RSVP-TE) is used to report failures related to LSPs to the ingress node or some other nodes responsible for error recovery. The node resolving a failure may in the future also perform crank-back. Modification of a connection, the flexible association of calls and connections, as well as the crank-back capability, are advanced features, which are still not fully covered within the GMPLS standard framework [Pap05a].

2.3.5 L1 VPN Service Support

Automatically controlled grooming networks can efficiently support the creation of multiple virtual topologies over common physical infrastructure. Such *virtualization* of the physical infrastructure is recognized as an important enabler both for distributed high-speed applications of the GRID community [Trav06],[Veer06], and for research on novel networking protocols and mechanisms [geni06]. This provides a strong motivation for research and standardization work on the Layer 1 Virtual Private Network (L1 VPN) service paradigm. L1 VPN should offer a new set of tools allowing clients to manage their logical topologies directly, by means of GMPLS, without the need to owe or manage any physical infrastructure.

The standardization framework for L1 VPN services is currently under development within both IETF and ITU-T Study Group (SG) 13, [Take05]. L1 VPN service borrows the concepts from widely deployed Layer 2 and Layer 3 VPNs. Reusing terminology from Provider Provisioned Virtual Private Networks (PPVPN) [Ande05], L1 VPN model includes customer edge (CE) nodes, provider edge (PE) and provider nodes (P). CE nodes are customer devices, e.g., Time Division Multiplexing (TDM) switches, routers, and layer 2 switches, which receive, and either switch or terminate, the layer 1 signal. One CE device connecting to at least one PE device, can be either a host, or a border node of a customer site, connecting to other customer devices (C). A PE node, e.g., a TDM switch, an Optical Cross-Connect (OXC), or a Photonic Cross-Connect (PXC), provides L1 VPN service to the customer, and connect to other PE devices, or provider devices (P nodes) - i.e., TDM switches, OXCs, or PXCs - that do not serve any customers.

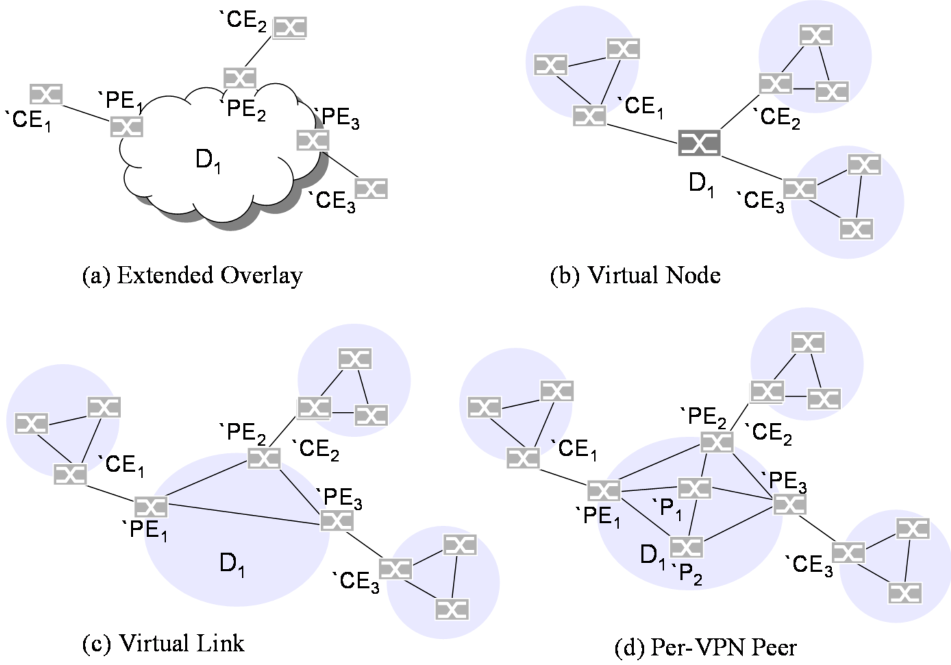
In general, *connectivity services* offered at layer 1 may feature control capabilities of different complexity, varying from low flexibility, where a pair of customer edge devices (CEs) is connected through a *permanent* connection, i.e., a traditional leased line statically configured by the provider, to high flexibility, where customers dynamically establish either *dynamic switched* connections, or customer-controlled *soft permanent connection* between CE devices. A soft permanent connection includes two permanently configured CE-PE links and a switched connection between the PE devices. Multiple permanent point-to-point connections among a set of CEs make a *static* L1 service; a *dynamic* L1 service consists of switched point-to-point and customer-controlled soft permanent connections.

Within general L1 connectivity services, Layer 1 VPN belongs to a class of *dynamic services* with a particular CE-connectivity defined as VPN membership, and with per VPN control and management functions, distributed among provider and customers control and management systems. The advanced novel features of L1 VPN are the capability to use traffic engineering tools to route service traffic and to modify VPN topology when needed. L1 VPN can support multiple types of (client) transport technologies, such as IP, and Ethernet, and different types of customers, in terms of the administrative relationship. The customers may be either "internal" and "external", such as service departments within one carrier's network which use the L1 VPN to provide different kinds of higher-layer services with payloads at any layer (e.g., asynchronous transfer mode [ATM], IP, or TDM), or other carriers that use another carrier's layer 1 VPN service, respectively. Depending on the administrative relationship, the topology information provided at the service demarcation points may be more or less limited. For example, "external" customers of layer 1 VPN service would receive only a limited view of the layer 1 VPN provider network and the limited control over the layer 1 VPN provider network resources. The level of control and management may be seen as a part of the service requirements and accordingly service layer agreement. At the service interface, or service demarcation point between CE and PE, the information exchanged may include signaling messages, membership information, TE information of the associated CE-PE TE links, customer network routing information (reachability and TE information), and provider network routing and TE information. The information sharing is highly dependable on the trust relationship between the customer and the provider.

In the transport plane, L1 VPN model supports sharing of transport resources, which are allocated and released by multiple VPNs in a time-sharing manner. Resources can also be configured as dedicated to - i.e., for exclusive use by - a VPN. Resource availability information provided to customers is in general abstracted: *a virtual link* is a provider network Traffic Engineering (TE) link that is a part of the routing information advertised to the customer, and *a virtual node*, is a provider network logical node advertised to customers, representing either a single physical node, or multiple physical nodes and the links between them. By using this advertised information, customers can specify explicit links in a request for a layer 1 connection setup, deletion, or modification between CEs, and will be given a notification of connection rejection with a reason when the provider cannot complete a service request.

Regarding the control and management information exchanged over the customer interfaces L1 VPN framework envisages four service models: the management-based model, basic signaling, overlay signaling, and signaling and routing model. In the management-based model the customer management systems and the provider management systems communicate to establish the service. With the basic signaling model connection setup is initiated by a customer over the CE-PE signaling link (GMPLS LSP signaling), however, the distribution of routing and membership information in the control plane is not supported. The overlay signaling model provides the customer interface based on the GMPLS UNI Overlay [Swal05]. The membership information is configured on PEs, and distributed between PEs, so that the PE that receives a Path message from the ingress CE can identify the remote PE connected to the egress CE. The customers can perceive the provider network from the path attribute specified and recorded in signaling messages. The signaling and routing model, as one of the most advanced VPN service paradigm, supports limited exchange of information between the provider control plane instance (CPI) and the customer CPI to facilitate discovery of customer network routing information - i.e., reachability of the TE information in remote customer sites - or parameters of the part of the provider's network dedicated to the customer.

The complexity of signaling and routing model depends on the service virtual topology. L1 VPN framework envisages four signaling and routing L1 VPN models, illustrated in Figure 2.5), defined as follows.



(a) Extended Overlay (b) Virtual Node

(c) Virtual Link (d) Per-VPN Peer

Figure 2.5: *Routing and Signaling L1 VPN Models.*

- **Extended Overlay Model** makes the CE-PE links visible to customers. A list of CE-PE TE link addresses to which a customer can request a VPN connection - i.e., membership information - and an additional information concerning these TE links (e.g., switching type)is provided at the CE-PE interface.

- **Virtual Node Model** presents the whole provider network to the customer as a single virtual node to which all other CEs in the VPN are connected. It is also referred to as a Generalized Virtual Private Cross-Connect (GVPXC). The CE receives routing information about CE-PE links and customer network (i.e., remote customer sites).

- **Virtual Link Model** constructs virtual links between PEs, each of which is assigned to a VPN. The CE receives routing information about CE-PE links, customer network, and virtual links assigned to each VPN. For each virtual link, the provider network allocates resources for the exclusive use of each virtual link. The TE attributes of a virtual link are determined according to resources allocated to it.

- With **per-VPN Peer Service Model** the provider partitions the TE links within the provider network into virtual links and nodes per VPN. A CE receives routing information about CE-PE links, customer network, and partitioned portions of the provider network. At least one virtual node corresponding to P devices (one single P or a set of Ps) must be visible to customers.

In addition to the information exchanged, service models significantly differ in the way in which transport resources - i.e., data plane resources - are allocated for each VPN. Virtual link and Per-VPN models support only dedicated transport plane; all other models support both dedicated usage of resources per VPN and sharing of resources between VPNs.

The Infrastructure Service model has many features of the per-VPN Peer service model, however, we extended it concerning the approach to the transport plane sharing. In our approach, the resource usage policies are defined to govern appropriate sharing of resources between deferent infrastructure services competing for the resources of the common infrastructure. With our visibility concept, described in Chapter 4, multiple services established over the shared infrastructure are *selectively offered* specific grooming and switching capabilities, based on which, each service's virtual topology may adapt to specific traffic and connectivity requirements of applications using it.

To enable customers to exploit the topology flexibility of grooming networks and dynamically set up L1 virtual networks, appropriate support in both the control and management plane is needed. As compared to statically provisioned L1 services, operational benefits of the L1 VPN with topology adapting to traffic changes are cost saving for the customer, and the higher network throughput - i.e., higher profit - for the provider. GMPLS/ASON mechanisms today only partially cover functionality that advanced L1 VPN services would need, but new protocol extensions could address new requirements. A related important question and an issue for further research, is whether the control plane support for virtualized networks needs new protocols - i.e., currently not within the GMPLS framework - to orchestrate distributed topology engineering, or could this be done by means of signaling protocol extensions? The control plane support for different type of services may differ, and in fact, CP functionality available to a specific type of services could be assumed as part of the service model, based on which a service level specification (SLS) can be defined. With a service level agreement (SLA), a guaranteed level of the service between the service user (who defines requirements) and the service provider (who offers capabilities) can be further detailed along the three components:

(1) VPN Traffic Requirements: These include the *connectivity and traffic/bandwidth requirements*, and the requirements regarding availability and restorability, as subject to monitoring and policing. For example, performance monitoring data, fault information, e.g., failure notification (by RSVP-TE), data plane alarm notification through the management plane, notification of connection setup rejection causes, could be given to the customers of L1 VPN service.

(2) The Transport Plane capabilities: These may include, e.g., granularity of virtual links, use of VCAT, use of grooming at different layers, resource sharing with other VPNs, and resilience support.

(3) The Control/Management Plane Capabilities: These specify the service interfaces in the control and the management plane and the information exchanged over the service interface. These may be classified according to a standard FCAPS model of ITU-T. Within, the **F**ault management has a role in monitoring and verification of correct operation, failure correlation and detection of customers affected by the failure. Resolving of failures and the reporting to the customers would depend on the agreement between a customer and provider. The **C**onfiguration management includes functions for configuration and verification of virtual node and virtual link, allocation and recovery policies, extra traffic policies, membership configuration/discovery, constraints on the call set-up time, the rate at which the topology can change, supported control plane interconnection model, measurement thresholds and triggers, resource sharing and visibility policies. **A**ccounting management functions need to record usage of VPN connections. **P**erformance management assists in monitoring of parameters related to the Service Level Agreement (SLA), and performance monitoring and analysis of other parameters such as network resource utilization.

The capabilities of the control plane play important role in the service model. For example, the call set-up time and service availability can be used to classify services into a small number of classes (such as gold, silver, bronze) as proposed in [Fawa04]. Service differentiation based on the capabilities of the control plane is also demonstrated in the CHEETAH network [Veer06] where the SONET path is used as a backup service for Internet transfer and its selection depends on the expected delay of the path setup and the expected duration of the transfer.

2.3.6 Outlook for the Control Plane Development

The control plane support for dynamic grooming decisions must meet two basic requirements. Firstly, there is a need for efficient modeling of network capability beyond capacity of links at different switching layers. This new grooming-specific TE functionality includes fixed and flexible inter-layer adaptations and VCAT/LCAS source and sink ends. Secondly, efficient distributed coordination of the path computation and the resource allocation is needed in order to deal with the restricted availability of the TE information at the routing controllers in the network. The current work on GMPLS and PCE (path computation element) addresses these requirements. Further extensions of CP support will play decisive role in a novel and challenging scenario of application-optimized network virtualization. Just as optimized selective deployment of grooming nodes can provide trade-off between bandwidth efficiency and network cost for expected traffic in static grooming networks, selective access to traffic grooming capability facilitated with the selective visibility of network resources, can support dynamic creation of the optimized topologies for L1 VPN services. This type of dynamic virtual topology engineering requires especially effective orchestration of routing and signaling, as a new virtual link is first routed, then signaled, and then made visible by the TE information update. Topology sharing between services is another challenge for the coordination of different control functionalities. Compared to future expectations, it could be observed that GMPLS/ASON currently provide a basic set of functionalities enabling automatic service provisioning that still needs to be extended to answer some of the requirements of traffic grooming. However, proposed standards have reached some maturity, successfully demonstrated in experimental inter-operability and feasibility tests [Jone04], [Jone05]. Furthermore, aiming at proof of new concepts several demonstrators and test-beds have been recently established with different network architectures for traditional provider or grid-centric network setups [Cava05], [Pina06], [Zhou06], [Papa06], [Veer06], [Lehm06], [Muno05], [Fois05], [Habi06]. These experimental studies shed light on a number of practical GMPLS aspects and on open gaps within the standardization framework. They also provide platforms for further studies on different service scenarios and new features that are essential for further improvements in the control plane.

2.4 Conclusions

This Chapter provided an overview of the grooming technology and the control plane standards related to the contribution of this thesis. The review of the GMPLS/ASON control plane, in particular its support for traffic grooming, highlights the approach to resource modeling, and the most important tools and concepts of the distributed resource management. These also provide a base for the federated architecture and the infrastructure service model presented in next two chapters.

3 GXP-based Architecture

3.1 Introduction

Generalized Multi Protocol Label Switching (GMPLS) provides a means for transition from centralized, manual network management to automatic and distributed network control. The benefits of such a transition include improved flexibility and scalability, reduced operational cost, and increased service adaptability. The original aim for GMPLS was to support unified, intra-domain multi-layer traffic engineering; however, this goal is currently evolving to include multi-area, multi-carrier aspects or routing. Learning from the case of the Internet, the proliferation of which was possible because it can function as a network of networks, the transport networks community identified the need for inter-domain control mechanisms in multi-carrier networks with Layer 1 services. Consequently, some aspects of multi-domain networking and traffic engineering have been addressed in the scope of inter-domain routing in optical networks [Bern02], inter-area routing with traffic engineering (TE) [Komp02], and BGP extensions for routing over multiple concatenated provider-domains [XuBa02]. These efforts, however, assume that global connectivity is achieved by means of statically configured inter-domain links. We believe that, because of its constrained flexibility, the existing interconnection approach offers only limited support for some challenging inter-carrier business-models, e.g., L1 bandwidth trading, which GMPLS could, at least in theory, support. As an alternative, we propose a novel transport network federation architecture [Tom03b] to exploit the reconfigurability of the optical layer, the flexibility of the dynamic grooming, and the benefits of GMPLS control.

This chapter describes the architecture for a federation of GMPLS networks, in which different administrative domains collaboratively integrate resources within a common infrastructure, making them subject to trading by dynamic allocation. The key elements of the proposed architecture are a GMPLS-enabled interconnection facility, referred to as the GMPLS Exchange Point (GXP), and a trusted control overlay, referred to as the Multi-Provider Edge (MPE). We propose these two concepts for the first time in the context of interconnected all-optical networks [Tom02b], [Tom02a], and then for federating heterogenous GMPLS networks [Tom03b].

With the GMPLS Exchange Point (GXP), we put the concept of the Internet Exchange Point (IXP), or the Internet Business Exchange (IBX) [Metz01], into the realm of the control-plane-centric (CP-centric) optical transport networks. Considering the importance of the role that IXPs play in supporting the global Internet, we see GXP as a significant but still-missing link in the big picture of GMPLS-based operations. In future multi-carrier heterogeneous transport networks, GXP could enable provider domains to change their interconnects automatically, using its switching capability in the optical and digital transport layers. To exploit the flexibility achieved with GXP, a trusted control overlay (MPE) is proposed, which can act either as a routing proxy or a routing service. As a routing proxy, MPE receives traffic engineering (TE) information from different administrative parties, from which it performs its policy-based filtering and re-distribution. As a routing service, MPE directly responds to all path computation requests without the need to re-distribute any TE information to other parties.

A design problem that arises in the GXP architecture is the GXP location problem. The problem belongs to a general class of facility-location problems which have been shown to be NP-complete, and for which a multitude of heuristics and approximation methods have been proposed [Drez95]. Focusing on a realistic scenario in which GXP nodes are incrementally deployed, we propose a simple, greedy heuristic for GXP location, based on the topological characteristic of interconnected networks [Tom02a].

This Chapter is organized as follows. Section 3.2 gives a brief overview of existing interconnection architectures and discuss the requirements for a new architecture. Section 3.3 describes the GMPLS Exchange Point (GXP) concept, the GXP resource control model, and the federated architecture with GXP and MPE. A network graph model that represents resources in a multi-layer, multi-carrier network with GXPs is described in Section 3.4. Based on this model, a classification of different service-provisioning scenarios is given in Section 3.5. Section 3.6 states the GXP location problem and describes a heuristic which incrementally selects a constrained number of GXP locations out of the set of potential ones. Section 3.7 concludes the chapter.

3.2 Existing Inter-domain Architectures

Internet Exchange Points and the associated mechanisms for inter-domain inter-working play crucial roles in supporting the growth of the global Internet. In the conventional wireline IP network, at the Internet Exchange (IX) or Internet Business Exchange (IBX) [Metz01], Autonomous Systems (AS) are connected either directly or through a Layer 2 switch. Inter-domain operation is enabled through a process in which network routes - i.e., the reachability information of the end-systems - are exchanged among the domains. In the multi-domain Internet today, this process is governed by the Border Gateway Protocol (BGP) [Rekh06] used by the autonomous systems to advertise their local routes, which are then distributed and installed according to specific policies. As a support for inter-domain traffic engineering for each of the imported routes, the end-systems learn which autonomous systems are traversed on this route. However, the routing information internal to each autonomous system is not propagated to other ASs. In this way, internal topology is effectively hidden and a global routing at the higher hierarchical level - i.e., the AS level - is achieved.

The topic of bandwidth trading with Internet Exchanges gained a great deal of attention during the late 1990s because of increasing interest in scenarios in which resource prices - i.e., the resource cost - and, consequently, the best traffic path, can change quickly [Semr00]. In such scenarios, resource trading can be organized by means of online auctions [Laza98], [InvH02]. A bandwidth trading solution based on these concepts and implemented within a management-system uses Pooling Points [Chel00] as infrastructure facilities, which connect and place bandwidth sellers and bandwidth buyers, physically and legally, in the market. The pooling point architecture offers a management-system interface for resource trading: When a request for the resource is processed, the system produces the management scripts and service can follow. In this system, the pooling point operator is responsible for the whole process, which requires special operational agreements.

A similar management approach is proposed for optical networks [Bisw03], which introduces a "bandwidth exchange layer" that collects information from the element and network management in order to react to new connection requests. In the architecture of all-IP wireless (or mobile) networks, the concept of the exchange has its representation in GPRS Roaming Exchange Points (GRX), where the providers of the GRX service are interconnected according to GRX peering agreements [Blyt01]. GRX plays a crucial role in users' roaming, enabling not only global connectivity, but also new mobile VPN services. In the optical domain, the "distributed exchange" concept, based on the optical BGP, is successfully deployed in the CA*net4 research network [StAr01], [StAr03], [WuJ04]. This concept supports establishment of dynamic IP-peering sessions over the global optical fiber network. In the CA*net4, the institutions interested in interconnecting acquire dark fibers from different carriers and connect to the optical cross-connects of the optical core in the CA*net4 network. The optical BGP distributes the reachability over the optical core. When one institution wants to establish a direct peering with another reachable institution, an IP BGP session is initiated and, within this session, a direct lightpath connecting them is established by the Lightpath Route Arbiter component. After the direct lightpath is established, the interconnected institutions can start BGP peering sessions between them. In this way, the optical core acts as a re-configurable, distributed exchange point.

The application of exchange points has also gained attention in the MPLS community. The exchange architecture, based on MPLS technology called MPLS-IX, was proposed in

[Naka02]. MPLS-IX is data-link independent and can unify two IX architectures prevailing today, viz. Local Area Networks (LANs), such as Fiber Distributed Data Interface (FDDI), Ethernet or Gigabit Ethernet, and Permanent Virtual Circuits (PVC) ATM. The management system can use MPLS mechanisms to configure virtual back-to-back links or back-to-back LSPs, inter-connecting the domains through MPLS-IX. Over those virtual interconnections, traditional bi-lateral peering agreements can be deployed. MPLS-IX is an important architectural advance, and by substituting MPLS with GMPLS, a similar control approach can be adopted for the optical IX. We refer to this kind of extension of the MPLS-IX approach into the GMPLS realm as a *static XP*. The GMPLS exchange point (GXP) proposed in this thesis is more flexible than the static XP, as it does not rely on a management system to set up the XP-internal interconnections, but integrates their set-up within the end-to-end automatic network control.

3.2.1 Requirements for a New Interconnection Architecture

The requirements for a new interconnection architecture based on GXP are further discussed using the example from Figure 3.1, which shows a multi-domain GMPLS network with four different carrier domains interconnected by static bi-lateral interfaces through a static XP. Like the VPN taxonomy, there is a distinction between the customer edge (CE) devices and the provider edge (PE), internal (P), and border (B) devices-all of which are generally referred to as Label Switched Routers (LSR). Each LSR can deploy any of the opto-electrical-optical (OEO) cross-connects (XC), all-optical wavelength XCs, all-optical waveband, fiber XCs, or their various combinations within hybrid nodes. Consequently, some of the domains may be all-optical, transparent, clouds, or deploying all-optical switching and routing without any electro-optical conversion within the optical nodes.

Within each domain, GMPLS TE techniques may be used for the intra-domain LSP control that is based on the available TE information about the transmission, multiplexing, and switching resources. In this multi-carrier interconnection architecture, the CE-PE interface

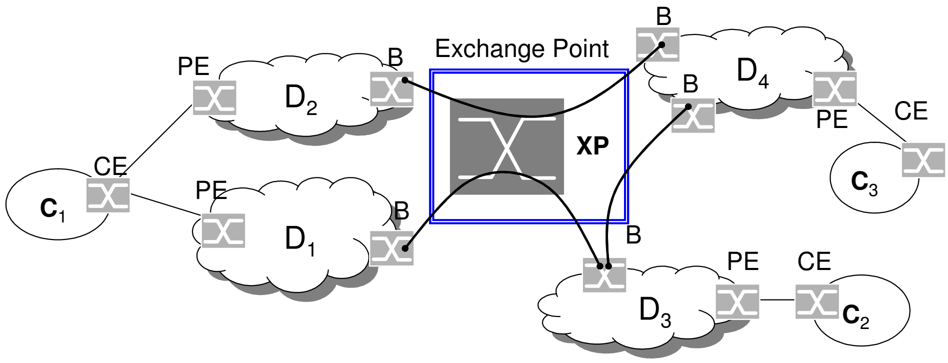


Figure 3.1: *Multi-provider, multi-domain architecture thee static inter-domain links.*

and the B-B interface, also denoted as the User-Network Interface (UNI) and the Network-Network Interface (NNI) [G807], are static layer 1 (L1) connections. In the control plane, UNI and NNI are bilateral interfaces with specific syntax and semantics for the exchange of control information. A specific set of policies or a specific agreement governs the exchange of the control messages and data traffic between the connected parties. This set of policies or agreement also applies to the exchange of topology information as a prerequisite for distributed path optimization. In this architecture, which we also refer to as the UNI/NNI architecture or traditional architecture, a static XP may be realized as a centralized multi-service switching node or a distributed multi-service switching network. In the example in Figure 3.1, two interconnections are established through a switch, and one is a direct, back-to-back link. Between two connected parties, e.g., between site c_1 and domain D_1, or between D_1 and D_3, GMPLS can support different interconnection models, either overlay, augmented or peer. The clients in this example can get the TE information from their direct-access providers by means of routing protocols (i.e., BGP, or OSPF or IS-IS extended for GMPLS) and use it for path computation and for connection-signaling (i.e., with RSVP-TE or CR-LDP extended for GMPLS). Regarding the availability of interconnections, such as the one between D_2 and D_4, this model assumes that these are statically provisioned through a management system based on bandwidth requirements and are often an outcome of bi-lateral negotiations [Gibb04]. Once an interconnection exists, domains can exchange their routing information, e.g., by using extension of BGP, as proposed in [XuBa02], or [Yang03].

The Problem with the TE Information Inconsistency

Because interfaces and agreements are bi-lateral, the problems of scalability and information integrity naturally arise in the traditional architecture as a N^2 problem. This is particularly the case in networks with multi-homing, where either a customer connects to multiple providers or a second-level provider connects to multiple core providers, which allows a choice of different routes over which to direct the traffic. The issue of making the best choice in this situation is illustrated in the Example in Figure 3.1, where we assume that a multi-homed user c_1 initiates the set-up of a generalized label switched path (LSP) to c_2, which may involve different domains-e.g., either D_1 and $D4$ or D_2, D_3 and D_4. Hence, c_1 may either use the TE information to determine the end-to-end path or select to which of the two access domains to delegate the path computation task, as discussed in [Komp02]. In general, the availability of the TE information or of a path computation service could be subject to a contract between connected parties, i.e., a kind of a *control level agreement* within a general Service Level Agreement (SLA). Such control-level agreements could also specify the supported level of TE information sharing, e.g., either full (peer) or none (overlay). For example, the interconnected parties from the Figure 3.1 example may settle for different levels of trust and interconnect their control-plane instances according to different interconnection models, as shown in Figure 3.2.

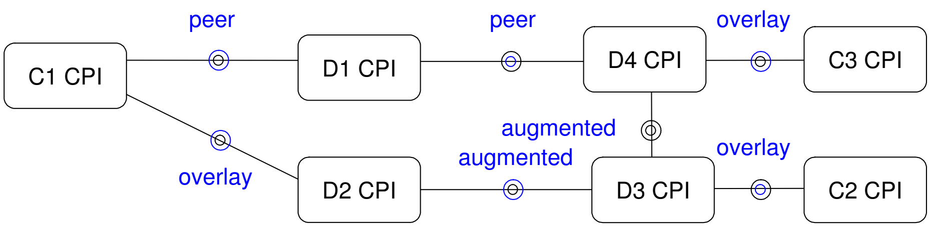


Figure 3.2: *Control plane interfaces between the routing controllers of control plane instances (CPI) of carrier domains and client sites.*

A control plane interconnection model on the interface between c_1 and D_1 is configured as a peer model and, on the interface to D_2, as an overlay. This means that D_1 provides full topology information to c_1, and D_2 provides none. Because of a high level of intra-domain information-hiding, which is currently a prevalent network operation policy, the

TE information received by c_1 will strongly depend on the bi-lateral agreements between even those domains that are not directly connected to c_1. For example, the topology information of D_4 learned by D_1 may either be transitive, in which case c_1 can receive it from D_1, or intransitive, in which case c_1 needs a direct bi-lateral agreement with D_4 to acquire it. The TE information may, in general, be inconsistent, so the traffic engineering based on it will lead to unavailable-or at least sub-optimal-routes. Whilst more frequent updates can provide better accuracy, they may incur a significant signaling burden. Consider the case of Amsterdam Exchange Point (AMS-IX) alone, which processes 170 Gbits per second of data and interconnects around 250 Internet providers. A control plane message using a TLV value of 41 bytes to update all bi-lateral agreements periodically (a conservative assumption) would result in control plane traffic of around 1.2 Gbyte per second.

As a way to approach this problem, we introduced a new global routing mediation functionality (MPE) in the GXP-based architecture. With MPE, the overhead for topology exchange can be reduced significantly, e.g., down to 20 kbyte on the cost of processing time at the MPE control engine in a previous example.

3.3 From the Static XP to GXP

3.3.1 The Introduction of a Trusted Control Overlay

The first step in the transition from the static XP to GMPLS Exchange Point (GXP) addresses the problem of bilateral CLA agreements by deploying a *trusted distributed routing service* similar to the concepts of a newly started path computation element initiative (PCE) [Farr06]. Figure 3.3 illustrates how such a routing service, referred to in our architecture as the Multi Provider Edge (MPE), can be offered in statically interconnected L1 networks, including static exchange points (XP). As the control flow shows, the MPE imports rules and agreements that each domain defines based on its resource management policies. The correlation of rules can now be a functionality of the overlay, which reduces the impact of modifications. Carrier domains can use MPE either as a routing service or as a routing proxy and advertise full routing information into it, together with the topology-masking - i.e., aggregation - policies and routing preferences. MPE, in turn, may either compute the best path for an LSP request or export the customized topology information to the connected parties, enabling them to compute paths locally. Each single bilateral agreement, now migrated into the trusted control overlay, can be modified in a more scalable way because new policies are advertised at only one interface and are directly used and applied only within the MPE routing service.

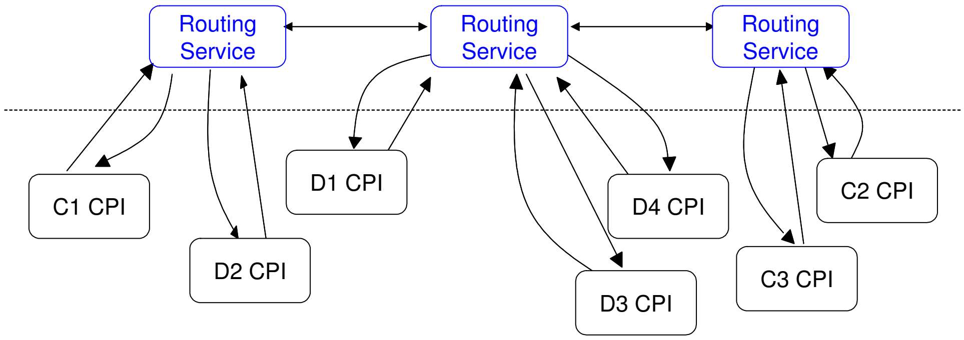


Figure 3.3: *Control flow between the routing service overlay and the control plane interfaces (CPIs) and of carrier domains and customer sites.*

3.3.2 Introduction of GXP Facilities

Even improved with MPE, a traditional network architecture would still suffer from low flexibility of the static CE-PE and B-B interconnections. GXP, on the other hand, can leverage a multi-layer exchange facility to offer dynamic interconnections of different types. In general, interconnections may be established through fiber switching, wavelength switching or digital-level TDM switching fabrics. Figure 3.4 illustrates how different domains can interconnect at a multi-layer node and use interfaces of different switching capabilities: D_1 and D_4 connect at the lambda switching capable (LSC) interface, D_6 and D_5 at the TDM switching capable (TDM) interface, and D_2 to D_3 at the fibre switching capable (FSC) interface.

Following the basic idea of GMPLS, GXP enhances the static XP with the capability of

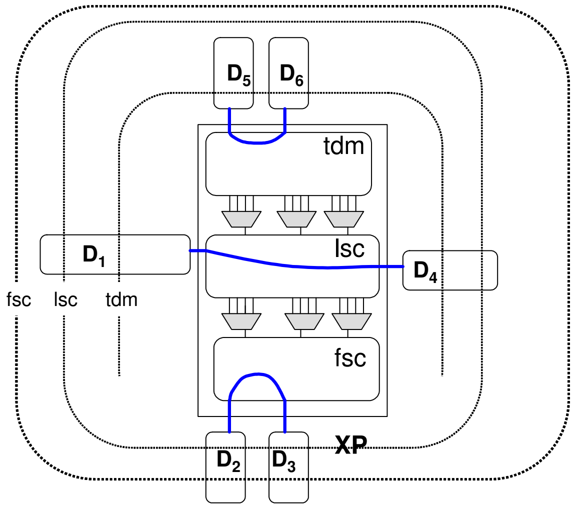


Figure 3.4: *Logical domain interconnections over the static XP. Domains D_5 and D_6 support switching at the TDM layer, domains D_1, and D_4 support switching at the LSC (wavelength) and TDM layers, domains D_2 and D_3 support dark fiber switching (FSC) and wavelength switching (LSC).*

dynamically changing domains' interconnection as a result of a *control-plane decision*. Let as assume that D_1 and D_4 and D_2 and D_3 internally include interfaces of the TDM and LSC type, and that the traffic that they exchange is a multiplex of TDM basic signals within one wavelength or a WDM multiplex within one fibre. Consequently, domain D_1 can elect to establish parallel TDM LSP connectivity to D_6 and D_4 during the morning hours and connect to D_2 with the extended bandwidth during evenings. The proposed GMPLS Exchange Point concept adds to this flexibility by leveraging re-configurability of the multi-layer switching core, which is composed of an optical and a digital cross-connect, and by supporting automatic control over exchange point facilities based on the GXP control model.

3.3.3 Control Model for GXP

A GXP control model is a topological representation of its internal capabilities. Just as a conventional Internet Exchange Point can be configured as one Autonomous System, a GXP can be one GMPLS administrative domain composed of a number of LSP Regions (technology layers). The multiplexing and switching capability of the domains' access and internal interfaces can be logically mapped to LSP Regions as well, based on the switching type. In fact, the domains connecting to GXP are logically connected to an appropriate LSP region, depending on the type of access link and the configuration of ingress and egress interfaces. Figure 3.5 shows a model of GXP that comprises TDM, LSC and FSC LPS Regions, as modeled with vertices x_{tdm}, x_{lsc}, $andx_{fsc}$, and interconnects several domains.

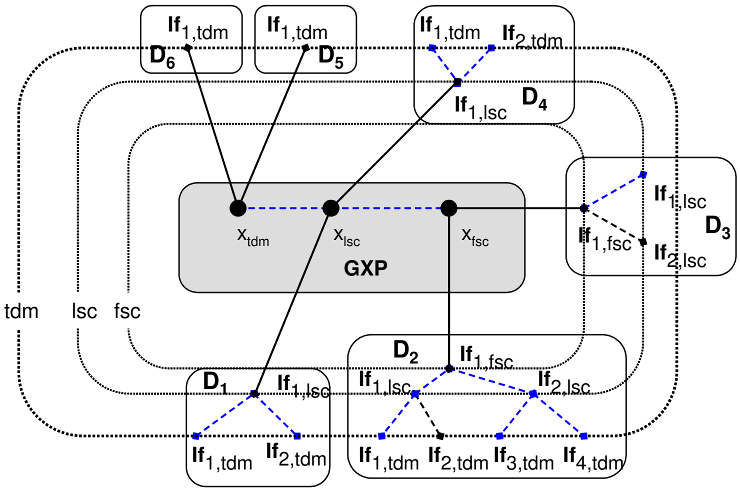


Figure 3.5: *GMPLS-XP Internal Topology Model.*

Domain D_1 is connected to GXP over a wavelength switching capable interface ($if_{1,lsc}$). This link is represented as an edge connected to a GXP-internal vertex x_{lsc}, representing the LSC Region. Let us focus on one wavelength component of this link: At one point in time, this wavelength may be switched to the interface $if_{1,lsc}$ of the domain D_4 while, at another point in time, D_1 may require interconnections to D_5 and D_6 at the TDM level. In this case, the GXP-internal interface between the GXP-internal TDM and LSC Regions will be used, that is, the edge between x_{tdm} and x_{lsc} and the TDM components will be accordingly de-multiplexed and switched at the TDM layer. Similarly, D_2 can connect to D_3 at the FSC layer or to D_4 and D_1 at the LSC layer. An important enabler of this kind of GXP control is the resource discovery and self-configuration process. As a part of the GXP control plane functionality, both the resource discovery and the self-configuration process are needed to create the topological TE representation of GXP resources according to this internal control model.

3.3.4 Federated Architecture

The GMPLS Exchange Point and a trusted control overlay MPE are the key elements of the federated architecture that we propose in this thesis. GXP physically interconnects the border equipment of different customers or carrier networks and supports connected parts to act as a provider-a party that advertises available resources-and/or as a customer-a party that initiates signaling of LSP requests. GXP also advertised its own available resources-d its own topological representation-either into the control plane instances of the connected domains or into the trusted control overlay (MPE), as illustrated in Figure 3.6.

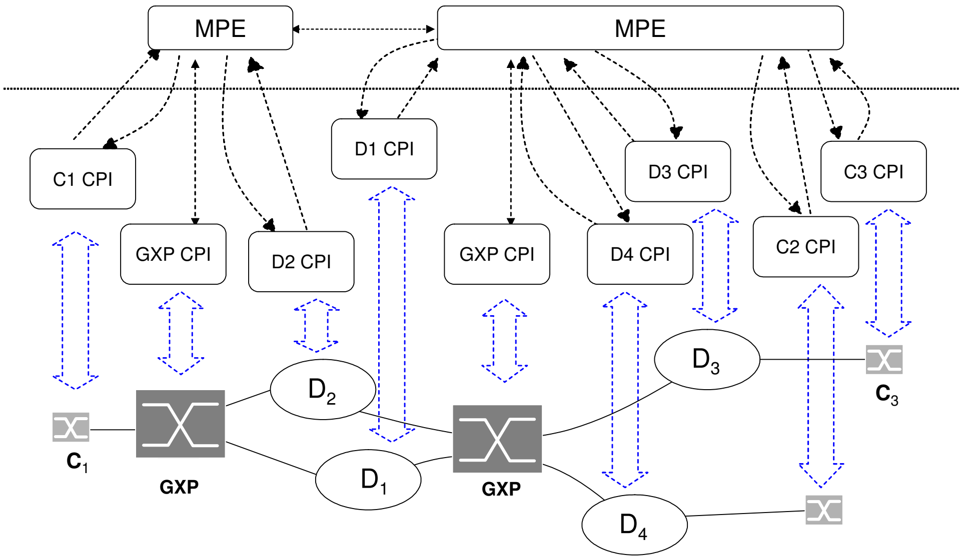


Figure 3.6: *Control flow in the federated network with GMPLS-XP and the Multi-provider Edge (MPE) control overlay.*

As a trusted routing service, MPE acts as a routing peer for carrier domains and customers' sites connected at GXPs and collects complete routing and policy information, including the metric for which the least-cost is to be achieved and the preference list of domains. Accordingly, GXP and MPE can enable functions of dynamic and carrier-neutral resource trading, a missing functionality in today's systems [Prev06]. Figure 3.6 shows the control flow in the federated architecture based on GXPs with the MPE control overlay. When connected to GXP, a client network c_1 can choose between the two provider domains D_1 or D_2 via one physical connection established with a GXP. Which domain would be used to establish a connection can be completely resolved by MPE. MPE and GXP are the binding elements in the federated architecture, whose aim is to offer a view of a ***common infrastructure*** to services provisioned over it. Hence, we also refer to a federated network as *the infrastructure*, to the resources integrated within it as the *infrastructure resources*, and to the services provisioned over it as the *infrastructure services*.

3.4 GMPLS Federated Network Model

This section presents a graphic model that captures transmission, switching, and layer adaptation (multiplexing) resources in a federated multi-layer GMPLS network. The model is defined based on the requirements of ASON/GMPLS resource modeling (described in Chapter 2) and includes the extended representation of the inter-layer adaptation capabilities. Let the *carrier domains* in a federated network be represented with a set $\mathbf{D} = \{D_k\}$, $k = 1, ..., |\mathbf{D}|$, the *customer sites* be represented with a set $\mathbf{C} = \{c_m\}, m = 1, ..., |\mathbf{C}|$, and GXP instances be denoted as $\mathbf{X} = \{X_u\}, u = 1, ..., |\mathbf{X}|$. Let the *granularity layers* be given with a set $L = \{l_g\}, g = 1, ..., |L|$, where a lower granularity layer has a higher index, e.g., $L = \{l_1 = FSC, l_2 = LSC, l_3 = TDM\}$. Following the concepts of GMPLS, the maximum capacity T_{g-1} of a signal multiplex in layer l_{g-1} can be represented as a group of U_g containers (labels) of capacity b_g in layer l_g; hence, $T_{g-1} = U_g \times b_g$. The values of U_g and T_g are encoded in the routing protocol extensions documents, e.g., OSPF-TE [Katz03]. We refer to U_g as a multiplexing factor between adjacent layers l_g and l_{g-1}.

Within each layer $l_g \in L$, let the switching capability[1] at the customer sites be represented with a set of vertices $V_g^c = \{v_{g,m}^c\}, m = 1, ..., |\mathbf{C}|$. At each granularity layer, let a customer site $c_m \in \mathbf{C}$ be represented with, at most, one vertex $v_{g,m}^c \in V_g^c$. A superscript c is used to annotate *customer* resources. Let the switching capability of the provider domain D_k be represented with a set of vertices $V_{g,k}^p = \{v_{g,k,i}^p\}, i = 1, ..., N_k$. A superscript p indicates that this is a *provider* resource. The vertices representing switching fabrics at different layers deployed within the same node share the same index; for example, $v_{g,k,j}^p$

[1] In ASON terminology: SNP subnetworks

and $v^p_{g-1,k,j}$ represent switching capability within the same node n_j of ^D_k at layers l_g and l_{g-1}, respectively.

Let intra-domain links in a provider domain D_k be represented as a set of unidirectional edges $E^p_{g,k} = \{e^p_{g,k,i,j}(v^p_{g,k,i}, v^p_{g,k,j})\}$. A superscript p indicates an intra-domain link. Let access links between customer sites and a domain D_k be represented with a set of unidirectional edges $E^c_g = E^{cp}_g \cup E^{pc}_g$, where $E^{cp}_g = \{e^{cp}_{g,k,m,j}(v^c_{g,m}, v^p_{g,k,j})\}$ and $E^{pc}_g = \{e^{pc}_{g,k,j,m}(v^p_{g,k,j}, v^c_{g,m},)\}$. The superscripts cp and pc indicate that the corresponding edges have a direction from a customer to a provider, and vice versa. Similarly, inter-domain links between domains D_k and D_n are given as a set of unidirectional edges $E^{pp}_g = \{e^{pp}_{g,n,k,i,j}(v^p_{g,n,i}, v^p_{g,k,j})\}$. Here, a superscript pp indicates that the corresponding edge connects two provider domains.

Let internal multiplexing and switching capacity between layers l_g and l_{g-1} be represented with a set of unidirectional edges: $E^{cm}_g = \{e^{cm}_{g,g-1,i}(v^c_{g,i}, v^c_{g-1,i})\} \cup \{e^{cm}_{g-1,g,i}(v^c_{g-1,i}, v^c_{g,i})\}$, for customer nodes and $E^{pm}_{g,k} = \{e^{pm}_{g,g-1,k,i}(v^p_{g,k,i}, v^p_{g-1,k,i})\} \cup \{e^{pm}_{g-1,g,k,i}(v^p_{g-1,k,i}, v^p_{g,k,i})\}$ for provider nodes. The total set of all edges is denoted as $E^{um}_g = E^{cm}_g \cup_k E^{pm}_{g,k}$. Superscripts um, cm, and pm indicate that the corresponding edge represents an inter-layer (multiplexing) link.

Every GXP instance X_u is represented with a graph $G^x_{g,u}(V^x_{g,u}, E^x_{g,u})$ abstracting the switching capability and connections enabled by this instance. A superscript x indicates GXP-related edges and vertices. A set of the granularity layers supported at X_u is given by L^x_u. Correspondingly, the switching capability at X_u is given with a set of vertices $V^x_{g,u} = \{v^x_{g,u,0}\}, l^g \in L^x_u$. Inter-layer capacity between $v^x_{g,u,0}$ and $v^x_{g-1,u,0}$ in X_u is modeled with the unidirectional edges $e^{xm}_{g,g-1,u}$ and $e^{xm}_{g-1,g,u}$. The nodes connected to X_u are modeled with a set of vertices denoted as $V^x_{g,u} = \{v^x_{g,u,i}\}$, where $i = 1, ..., N(X_u)$. Connections at the granularity layer l_g in X_u are given as edges $E^x_{g,u} = \{e^x_{g,u,i,0}\} \cup \{e^x_{g,u,0,i}\}$. A GXP can connect to carrier or customer nodes, so, assuming at most one vertex per carrier domain and granularity at each GXP, $V^x_{g,u} \subset V^c_g \cup_{\{D_k \in \mathbf{D}\}} V^p_{g,k}$.

In each layer l_g in which dynamic, virtual, links can be created, the set of graph edges

E_g–which includes the access, inter-domain, intra-domain, and GXP edges as defined-has to be extended with the inter-customer edges E_g^{cc} and inter-GXP edges E_g^{xx}. These are the unidirectional edges defined with $E_g^{cc} = \{e^{cc}_{g,i,j}(v^c_{g,i}, v^c_{g,j})\}$ and $E_g^{xx} = \{e^{xx}_{g,i,j}(v^x_{g,i,0}, v^x_{g,j,0})\}$.

Generally, each edge in a network graph $e_g \in E_g$ is characterized with the total bandwidth $B(e_g)$, and the residual bandwidth $R(e_g)$ is expressed as a discrete numbers of labels at layer l_g. Each edge is also characterized with some administrative edge cost $A(e_g)$ and the visibility attribute $\vartheta(e_g)$. The usage of these attributes is described when the infrastructure service model is described in Chapter 4.

Figure 3.7 shows a graph of a federated two-layer network (L=$\{l_1$=LSC, l_2=$TDM\}$), introduced in Figure 3.6.

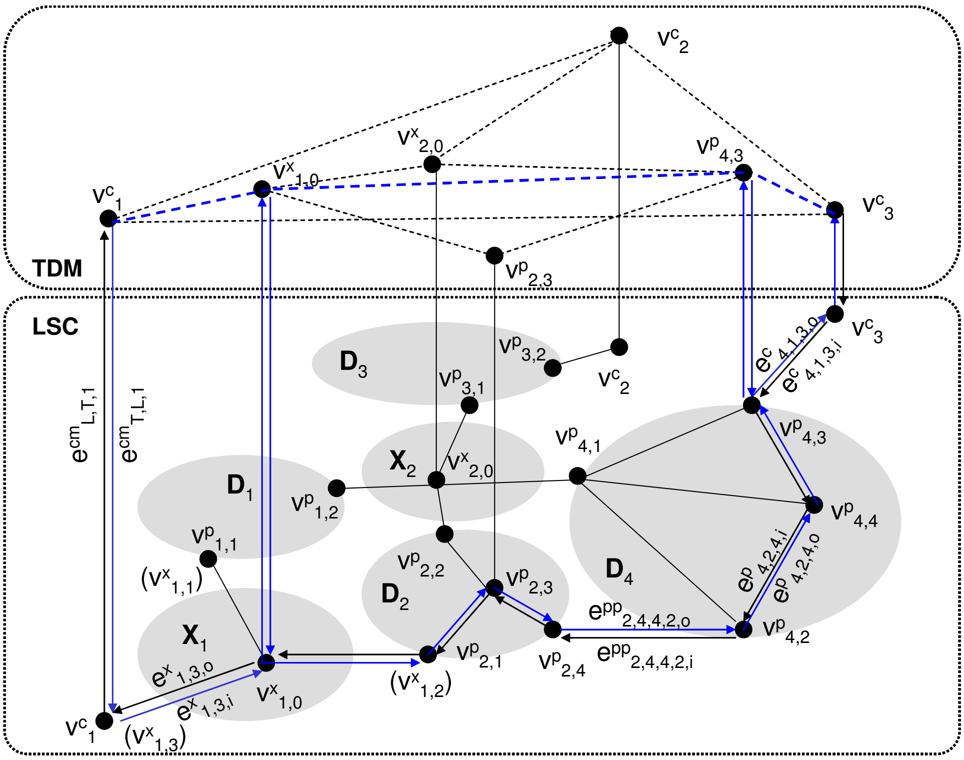


Figure 3.7: *Two-layer network graph model.*

For simplicity, we omit layer indices (TDM, LSC) in the names of vertices and edges at the corresponding layers and use abbreviations "T" and "L" only for inter-layer edge indexing. Each edge is comprised of two directions, which are explicitly shown only for a subset of characteristic edges. Inter-layer edges are shown for three nodes at customer sites v_1^c, v_2^c and v_3^c; provider nodes $v_{2,3}^p$ and $v_{4,3}^p$ in D_2 and D_4, respectively; and two GXPs, v_1^x, v_2^x. Note that, in the physical topology, GXP interconnections captured in E^x (GXP) are created on-demand (in ASON terminology SNP subnetwork connections), whereas the edges captured in E_g^{cp} (customer access links, e.g., between V_2^c and $v_{3,2}^p$) and E_g^{pp} (back-to-back domain interconnections, e.g., between $v_{2,4}^p$ and $v_{4,2}^p$) are static and always present. Figure 3.7 illustrates that GXP x_1 connects customer $c_1(v_1^c)$ and domains $D_1(v_{1,1}^p)$ and $D^2(v_{2,1}^p)$.

The network multi-layer graph captures all network resources and is a basis for a routing graph. In the case of routing over the multi-layer graph, we assume the GMPLS/ASON model of client-server inter-layer coordination, in which a virtual link in a client layer (e.g., TDM) is created by routing a server layer LSP (e.g., WDM), and a client layer LSP is routed using the client layer (virtual) links. For example, to route a TDM LSP from c_1 to c_3 (dotted), the corresponding virtual links in the TDM layer must first be established. Figure 3.7 shows one possible topology in the TDM layer and depicts resources (in blue), both in the LSC layer and between the TDM and LSC layers, that are allocated when three uni-directional virtual links in the TDM layer are established.

3.5 Multi-domain Service Provisioning Scenarios

The federated architecture featuring GXP and the multi-provider edge (MPE) allows for a number of provisioning scenarios involving carrier domains, customer sites, GMPLS Exchange Points, and the MPE routing service. With a simple, single, point-to-point connectivity service, a number of characteristic provisioning scenarios can be identified, which we describe in the example of a multi-carrier network with both static interconnections and GXPs, depicted in Figure 3.8.

To model a service path computation performed on behalf of a customer by either a particular access domain or by MPE, we use the following notation. For a domain D_k, the network graph capturing the actual (accurate) resource state is denoted as $G^k(V, E)$, and the state of the other part of the network, as observed by the domain D_k, is denoted as $\widetilde{G^k(V,E)}$. This means that $\widetilde{G^k(V,E)}$ is created based on the TE information that D_k receives. For MPE, we assume that the observed state of the whole network is the actual state, i.e., $G(V,E)$. Here, the scenarios are defined either from the viewpoint of a domain D_1 (shown in blue in Figure 3.8) or from the viewpoint of MPE. We can make a distinction between different types of customers and domains: Customers c_1 and c_2 and a domain D_5 are directly connected to D_1; through GXP X_1, the domain D_1 connects to D_2, D_3 and c_6; and through X_2 and X_3, the domain D_1 connects to D_2 and c_4 and c_3, D_3 and D_4, respectively. Depending on the party that computes a service path, we classify services into intra-domain, inter-domain, domain-neutral, and bandwidth trading. These different classes are illustrated in the example depicted in Figure 3.8.

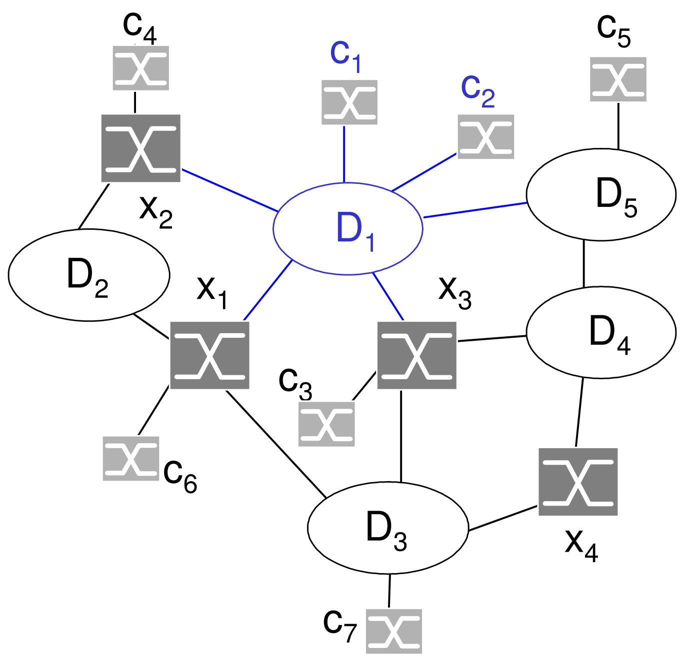


Figure 3.8: *Example of a federated network with five carrier domains, four GXPs and seven customers connected either directly or through GXPs.*

- **Intra-domain Service:** Directly connected customer c_1 initiates an LSP to another directly connected customer c_2. Domain D_1 computes the path to c_2 on behalf of c_1. Typically, only domain-internal resources will be used for intra-domain requests. The routing objective is to minimize the cost of the used links, i.e., to find the shortest path of all the paths $P^1_{c_1,c_2}$ in the graph $G^1(V,E)$ between v^c_1, and v^c_2.

- **Inter-domain Service - without MPE:** Directly connected customer c_1 initiates an LSP to an external customer c_5 of a directly connected domain D_5. Hence, the service has two segments, one within D_1 and one within D_5. Either access domain D_1 will use the routing information provided by the domain D_5, as integrated within $\widetilde{G^1(V,E)}$, to compute an end-to-end path $P^1_{c_1,c_5}$ in the graph $G^1(V,E) \cup \widetilde{G^1(V,E)}$, or each domain autonomously will resolve the sub-path for its segment, i.e., $P^1_{c_1,D_5}$ and $P^5_{D_1,c_5}$.

- **Inter-domain Service - with MPE:** Directly connected customer c_2 uses MPE service to get a route for an LSP to another customer c_6. This scenario is different from the previous scenario because MPE, not D_1, is responsible for finding the route to c_2. In this case, P_{c_2,c_6} is computed on the graph $G(V,E)$.

- **Domain-Neutral Service - with MPE:** Customer c_4 connecting over GXP X_2 uses MPE to find the path for an LSP to another customer c_7, which is directly connected to D_3. Here, the routing can reflect all the policies imported into MPE. Again, P_{c_4,c_7} is computed on the graph $G(V,E)$.

- **Bandwidth Trading Service:** Domain D_1 either obtains sufficient routing information through MPE or uses MPE to find a route for an LSP between two of its nodes connected over GXPs. This results in establishing intra-domain link over external resources. Correspondingly, MPE computes a path P_{X_1,X_3}, P_{X_2,X_3}, or P_{X_1,X_3}.

The MPE control overlay provides the interface for all connected parties through which the resources can be offered, by injecting the related TE information and by requesting LSP setup over them. In computing paths, MPE relies on the TE information it acquires; however, this information does not have to reflect the carriers' actual topologies explicitly, as we previously assumed, but can be an aggregated, simplified graph. Based on the approach proposed in [Tom02a], one possible aggregated topology model with GXPs is illustrated in Figure 3.9. The topology model is devised for the network shown in Figure 3.8. Figure 3.9 represents a single granularity layer l^g; in a multi-layer scenario, it could be simply extended by adding additional layer graphs and interconnecting them with inter-layer edges.

Omitting the layer indices, the switching capability at GXP X_u is represented with a vertex $v^x_{u,0}$; each connected carrier domain D_i is represented with one vertex $v^x_{u,i} \in V^x_u$. Note that, in the previously presented (non-aggregated) model, the vertex $v^x_{u,i}$ represents an LSR. If a domain D_k is connected to GRXs X_i and X_j, an edge between $v^x_{i,k}$ and $v^x_{j,k}$ is inserted in the graph to represent resources of D_1 between X_1 and X_2. For each domain D_z which connects to only one or no GXPs, a separate vertex v_z is included in the graph (e.g., for D_5). Direct connectivity between domains is modeled as edges between the corresponding vertices. For example, D_5 connects to D_1 and D_4, so the edges between v_5 and $v^x_{i,1}$, (i=1,2,3) and between v_5 and $v^x_{j,4}$ (j=3,4) are inserted in the graph. Each customer is represented with a vertex which is connected to either a corresponding GXP vertex (e.g., c_6 to $v^x_{g,1,0}$) or to all vertex representations of carrier domains (e.g., c_1 to D_1 and c_5 to D_5).

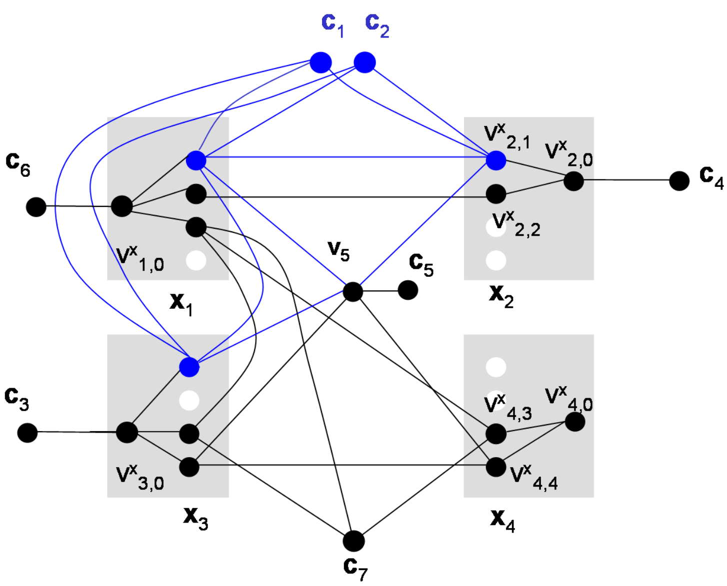


Figure 3.9: *Aggregated graph for a federated network.*

This graph, which can be used for computing service paths by MPE, assumes that each carrier domain has advertised the bandwidth and the cost of the relevant aggregated edges in the graph.

3.6 GXP Placement Problem

The most important problem related to the federated GMPLS architecture is that of a migration from the statically interconnected topologies towards a network with GXP facilities. A provider of GXP facilities would need to determine placement of one or a number of GXPs so the extended topology improves the network performance for the parties connected over it. In a realistic networking scenario, where there is an existing multi-carrier network with established points of collocation, several locations can be identified as potential candidates for deploying GXPs. We assume that the GXP installation incurs costs and that the monetary budget for capital investments in GXPs is constrained (B). Without loss of generality, we also assume that each potential location requires the same investment for installing the GXP. Considering the fact that GXP is a multi-layer switching node, a GXP location problem can be formulated as an optimization problem on the multiple layers graph representing the network resources. For example, if we assume that GXP has a WDM switching and TDM switching and grooming capability, both TDM and LSC layer topology can be improved by introducing an GXP. The GXP placement problem belongs to a group of facility-location problems. In the IP network, the optimal design of peering locations between autonomous systems was considered in [Awdu98] and in a similar gateway location problem in [Lian89]. In a hybrid optical network, a problem which is most relevant to the GXP location problem is the one in which, starting from the original network topology of wavelength switching nodes, a set of nodes which upgrade with the digital grooming capability can optimize the network cost or throughput, as proposed in [Ceru03] or more recently in [Che05a], [Che05d].

In our approach to designing GXP locations, a crucial assumption is that the carriers establish trust relationship with MPE, which acts as a routing control overlay, even with no GXPs deployed in the field. Based on the information provided to MPE, the interconnection topology of the domains could be discovered. Similarly, the TE information which is regularly made available to the MPE can be used to discover the traffic patterns in network. Based on this information, the MPE operator could solve the GXP location problem, including the positions and also, potentially, the dimensioning and the routing over the GXP network. On the other hand, the GXP provider can also consider a simplified problem based on the assumption that the routing will be dynamically governed by different policies and that the topology design and the dimensioning of GXP nodes' grooming capability can be separated. Thus, the locating GXPs should improve the static physical layer topology. (In a TDM/WDM network, this is the WDM layer.) We use this approach to define the design objective as: *Find locations for GXP such that the best possible improvement of a global network utilization is achieved compared to the original topology, while keeping the total installation cost constrained.*

Let I be the set of indices of all possible GXP locations, and let I^i be one possible set of indices of installed GXPs. Depending on the GXP nodes in the topology, let the edge-set E of the network $G(V, E)$ be represented with $E(I^i)$. This means that, for the starting UNI/NNI network, we have $E(\emptyset)$. By installing GXP, the MPE operator would try to increase the network throughput for the traffic generated between the nodes that can request an LSP in the lambda switching layer (LSC LSP). This can be achieved by using shorter paths enabled by connections through GXPs. We can assume that the traffic is known to the MPE provider and is characterized by the average traffic matrix $T = \{t_{od}\}$, with t_{od} as the average expected traffic between each pair of v_o and v_d) and the hop-count matrix $H = \{h_{od}\}$, where h_{od} is the length of the shortest path between v_o and v_d. Consequently, in the starting topology, each domain can be characterized with the average weighted number of hops $\overline{h}$ (Equation 3.1) for the current traffic, which is one possible representative of the network utilization [Leon00].

$$\overline{h} = \frac{1}{t_{tot}} \sum_o \sum_d h_{od} t_{od} \tag{3.1}$$

where t_{tot} is total average (expected) traffic in the network:

$$t_{tot} = \sum_{o} \sum_{d} t_{od} \tag{3.2}$$

Let $h_{od}(I^i)$ be the hop count of the shortest path for a request between v_o and v_d in a network with installed GXP instances with indices in I^i. The problem of network utilization improvement can be stated as:

$$find\ I^* \subseteq I\ so\ that\ \overline{h}(I^*) = \min_{I^i \subseteq I} \overline{h}(I^i) \tag{3.3}$$

Based on this rationale, we present a simple greedy heuristic solving the problem Equation 3.3. Given a set of candidate placements of GXP instances (e.g. existing domain collocations), this heuristic starts with an empty set of installed GXP instances and finds in each iteration one GXP instance (X_i) that most improves the weighted hop count for the topology. A GXP is included in a topology by inserting its corresponding vertex and edges, which serve as the input for a new iteration. As depicted in Figure 3.10, we assume that all the collocated nodes connect to a newly inserted GXP. The cost of the deployed GXP (p_m) is added to the cost of the total solution.

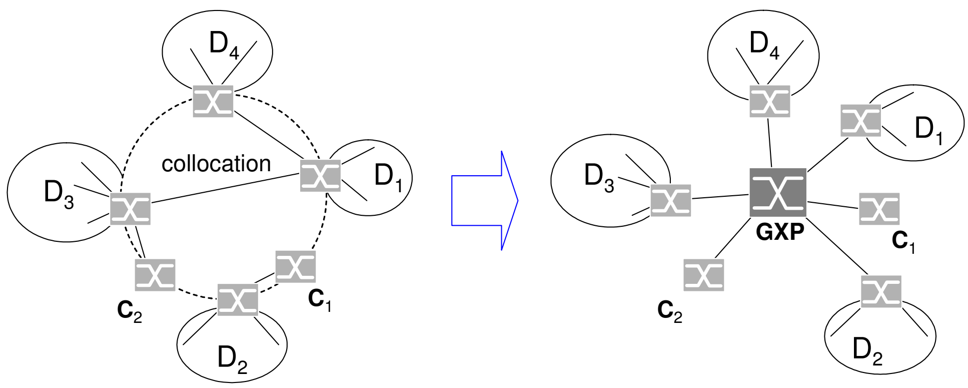


Figure 3.10: *Extending interconnections with GXP.*

The algorithm runs until no further improvement is possible or the budget for GXP installation has been exhausted. The hop reduction factor, ρ_i, is the metric used to find the best GXP in each iteration. The metric is computed for each GXP instance X_i which is a candidate for installation in a particular iteration as a sum of differences between the weighted number of hops for each (o, d) pair computed in the current topology (μ_{sd}^{k-1}) and in the i-test topology $(h_{sd}(I_{k-1}^* \cup \{i\}))$, if positive; otherwise, it is set to zero (Equation 3.4). Here, the i-test topology is the topology which includes links of the candidate i^{th} GXP instance under test (X_i). We select the location to install an GXP instance as the one with the maximum hop reduction factor, and the output of the iteration is the topology where the edges related to this GXP instance are added for a new iteration.

Heuristic for GXP placement (GXPp)

Initialization: Let $k = 0$, $I_0^* = \emptyset$

for each $c_o, c_d \in C$ set $\mu_{od}^0 = h_{od}(I_0^*)t_{od}$

Check Cycle:

$k = k + 1$,

for each i not in I_{k-1}^*, set

$$\rho_i = \sum_o \sum_d max(0, \mu_{od}^{k-1} - h_{od}(I_{k-1}^* \cup \{i\})) \tag{3.4}$$

find $m \notin I_{k-1}^*$, so that

$$\rho_m = \max_{i \notin I_{k-1}^*} \rho_i \tag{3.5}$$

set $\rho^k = \rho_m$

if $(\rho^k \leq 0)$ and $(|I_k^*| \geq 1)$ then go to End.

if $(p_m + |I_k^*|p_i) > b$ then go to End.

let $I_k^* = I_{k-1}^* \cup \{m\}$

if $|I_k^*|p_i \geq b$ then go to End.

for each o, d, $c_o, c_d \in C$, set $\mu_{od}^k = h_{od}(I_k^*)t_{od}$

go to *Check Cycle.*

End: Calculate

$$\overline{h} = \frac{1}{t_{tot}} \sum_{o} \sum_{d} h_{od}(I_k^*) t_{od} \tag{3.6}$$

The complexity of these heuristics is $|M|^2(|D|^2N^2 + C)$. Here, $|M|$ is the number of all GXP candidate instances, $|D|$ the number of all domains, N the number of nodes per domain, and $C = |D|^2N^2$ the complexity of the Dijkstra's shortest path algorithm [Dijk59].

For our simulation study, we apply the simplified form of this heuristic, which does not account for the traffic. Alternatively, the load at the LSC layer can be measured while simulating in the UNI/NNI topology, and this observed value can be used in the proposed heuristics. However, for the traffic patterns that we use in the study, there is basically no difference in the results of the GXP location heuristic.

3.7 Conclusions

For the last several years, the focus of the GMPLS community has been predominantly on mechanisms for the unified control of the multiple-layer networks, while the issues of inter-domain control have attracted limited attention. However, inter-domain networking issues are of critical importance for the operation of the global, automatically controlled transport infrastructure and for dynamic provisioning of global services. This Chapter introduced the GMPLS Exchange Point (GXP) as a functional equivalent of the Internet Exchange Point (IXP) which, in a global Internet, has an enabling role in managing business relationships between interconnected domains and in inter-domain IP traffic routing. Similarly, GXP can dynamically interconnect different administrative GMPLS domains on different hierarchy levels. We proposed a TE information model for GXP switching capability, a topological representation of LSP Regions within GXP to be advertised within the global routing process. We also introduced a Multi-Provider Edge (MPE) as a trusted control overlay facilitating GXP infrastructure and providing a trusted global routing service. A federated architecture with GXP and MPE has several advantages over the architecture with multiple User Network Interfaces (UNI) and Network-Network Interfaces (NNI). First, assisted by flexible access interfaces and the multiple switching capability of GXP, the interconnection capacity among the domains can change. This has a significant practical relevance, particularly with an increased traffic dynamic. Today, the domains' interconnections are mostly dimensioned empirically-that is, based on measurements-and are statically configured; hence, they cannot adapt to some transient changes. Second, for the domains that need to engage in inter-domain traffic exchange, MPE can provide relevant TE information that is customized according to a control agreement between the domains. Third, with the increased number of domain interconnections which can come to hundreds or thousands, an MPE control interface can reduce the signaling overhead. Therefore, by giving the user GMPLS-based control over the exchange facility, the standard multi-domain network architecture, based on the traditional bi-lateral exchange mechanisms and agreements between domains, as is inherent to BGP and the static XPs, can be made flexible. In the network architecture with MPE and GXPs, traditionally off-line network engineering concepts can be newly positioned for dynamic, on-line, adoption: Combined, the trusted control overlay and the flexible physical layer connectivity can support scalable exchange of the routing information, policy-based routing and bandwidth trading.

4 Infrastructure Service Model

4.1 Introduction

This Chapter describes the Infrastructure Service model, which we proposed for dynamic virtualization of heterogenous optical transport networks. Infrastructure services are virtual networks, running their own instances of the GMPLS control plane, within the common physical infrastructure spanning multiple layers, multiple carriers, and, in a federated architecture, multiple GXPs. The key property of the Infrastructure Service model is resource sharing: a virtual link configured by one service, may be used by other services. This enables services to time-share un-utilized configured bandwidth. To instrument this feature, this thesis proposes a concept of the resource visibility, which includes a new TE attribute assigned to each physical and logical resource, defining whether, and how, it can be offered to different services.

Regarding the technological realization of infrastructure services, we focus on two-layer hybrid optical networks, with the label switched routers (LSR) preforming lambda switching and TDM switching and grooming. At the customer interface, we assume that customer equipment can multiplex TDM containers into wavelengths and switch at both TDM and WDM layer. Therefore, at the customer side, the setup of lambda switched LSPs (LSC LSPs), as well as the TDM granular LSPs embedded within wavelengths, can be initiated. The infrastructure service model is general, and makes no distinction between "simple", and "complex" customers, for example, between a home user requesting a video-on-demand bandwidth pipe, and a campus network operator requesting a "user-owned" network akin to UCLP. For example, a service S_1 may be used by a regional broadcaster (RB) who wishes to link all his regional offices via an optical network, and run an IP based network infrastructure over it [Prev06]. Similarly, infrastructure service S_2 may interconnect the sites - i.e., the scientists - of three National Research Experimental Networks (NREN), to support joint execution of bandwidth rich experiments in astrophysics [Karm05]. A customer can also be another provider's network, which uses a service to temporarily *extend* its coverage, or capacity.

The *extensibility* is a major property of infrastructure services: each service can dynamically add or release virtual links, changing the connectivity and bandwidth of its virtual topology. This is done either in order to accommodate traffic or to optimize the utility of resources temporarily bound within the virtual topology, or both. This functionality is implemented within the service control plane, the core of which is the service topology engineering (ToE) function, which either reacts upon the resource monitoring events, such as low utilization or outage, or directly assists traffic routing as an atomic Routing and Topology Engineering (RToE) action. The control plane instance (CPI) of each infrastructure service interacts with other services' CPIs, with CPIs of the carrier domains, or with the MPE control overlay. This interaction is based on GMPLS control protocols, which are the basic tools for performing topology engineering (ToE) actions. For example, setting up, or releasing a virtual link is done by means of signalling along a specific path. Depending on the availability of the TE information, a path may be computed in a centralized or distributed way by different CPIs involved in the virtual link setup. In this respect the infrastructure service belongs to a class of Routing and Signalling L1 VPNs [Take05]. Akin to L1 VPNs, the provisioning of infrastructure services includes several phases, such as resource negotiation, membership learning and topology maintenance. We assume that the interplay of the data plane, the management plane, and the GMPLS-based control plane, assures capability that service are instantiated "on-fly", and that the corresponding physical resources are configured as visible. With this assumption in mind, our work focuses on dynamic topology engineering, and dynamic traffic routing during the maintenance phase.

This Chapter is organized as follows. Section 4.2 provides a brief overview of the related work in virtual topology design and reconfiguration. The infrastructure service concepts of resource visibility and resource sharing are presented in Section 4.3. Section 4.4 presents the modeling approach, including the *Service Visibility Graph (SVG)*, and its transformation used for traffic routing and topology engineering. Topology Engineering (ToE) methods are described in Section 4.5. Section 4.6 presents two methods for traffic Routing and Topology Engineering (RToE), and illustrates them in a case study with 2-layers hybrid optical networks. Section 4.7 concludes the Chapter.

4.2 Related Work in Virtual Topology Design and Reconfiguration

The past research work on virtual (logical) topology design and reconfiguration in WDM and grooming networks was initially motivated with the benefits of topology extension in IP-optical integrated networks [Chan98]. In this scenario IP-optical routers are capable of setting up lightpaths accross the underlying wavelength-switching infrastructure, dynamically increasing the bandwidth of the IP layer links when needed. In a data network which traffic can be described with a traffic matrix $T = \{\lambda_{sd}\}$, for each pair (s, d), ideally there would be a direct lightpath - i.e., a direct virtual link - transporting this traffic. Due to the physical constraints, such as the number of available transceivers that can terminate lightpaths at each network node, only a fraction of the traffic can be transported on the single-hop virtual links; the remaining must be routed over multi-hop paths in the virtual topology. The seminal work presented in [Rama96] stated a virtual topology design problem as a mixed-integer linear program (MILP), which input is a given traffic matrix and the constraints of the nodes, and the output is a set of virtual links which minimizes the highest logical link load, also refereed to as congestion. This formulation, included here for the illustration of the problem, use the following model.

The virtual topology is described with binary variables $b_{ij} \in \{0, 1\}$, where $b_{ij} = 1$ if there is a logical link from node i to node j in the logical topology, and $b_{ij} = 0$ otherwise. The integer values of variables b_{ij} are the output of the optimization. The propagation delay on a logical link (i, j), is denoted with d_{ij}, and is determined by the actual routing over the physical topology as a sum of physical link delays. The maximal propagation delay between any $s - d$ pairs is denoted as d^{max}, and the maximum permissible average delay as αd^{max}. A common assumption is that the shortest propagation delay is achieved by routing the traffic on the shortest path in terms of number of physical hops. In the model λ_{ij}^{sd} denote the arrival rate of packets from customer s to d on link (i, j), λ_{ij} the total arrival rate of packets on link (i, j) from all $s - d$ pairs. The maximum load on any logical link, or the congestion, which is subject to minimization, is denoted as λ_{max}.

The problem is stated as follows:

$$\min \lambda_{max} \tag{4.1}$$

subject to the following.

Flow conservation at each node:

$$\sum_j \lambda_{ij}^{sd} - \sum_j \lambda_{ji}^{sd} = \begin{cases} \lambda^{sd}, if\ s = i, \\ -\lambda^{sd}, if\ d = i,\ for\ all\ s, d, i \\ 0, otherwise \end{cases} \tag{4.2}$$

Total flow on a logical link:

$$\begin{aligned} \lambda_{ij} &= \sum_{s,d} \lambda_{ij}^{sd},\ for\ all\ i, j \\ \lambda_{ij} &\leq \lambda_{max},\ for\ all\ i, j \\ \lambda_{ij}^{sd} &\leq b_{ij}\lambda^{sd},\ for\ all\ i, j, s, d \end{aligned} \tag{4.3}$$

Weighted average delay constraint for each s-d pair:

$$\sum_{ij} \lambda_{ij}^{sd} d_{ij} \leq \lambda^{sd} \alpha d_{max} \tag{4.4}$$

Degree constraints and the logical link capacity (C) constraint:

$$\begin{aligned}\sum_i b_{ij} &= \Delta_l,\ for\ all\ j \\ \sum_j b_{ij} &= \Delta_l,\ for\ all\ i \\ \lambda_{ij}^{sd}, \lambda_{ij}, \lambda_{max} &\geq 0,\ for\ all\ i, j, s, d \\ b_{ij} &\in \{0,1\}\ for\ all\ i, j \\ C &\geq \lambda_{max}\end{aligned} \tag{4.5}$$

In this formulation d_{ij} are input values which are obtained by finding the shortest path between node i and j over the physical topology. Δ_l is a logical degree of a node, i.e., a node can terminate Δ_l input and output logical links. This model does not allow multiple links between the same nodes, which can be accounted for by a new set of variables ($b_{ijk}, k = 0, 1, ..., \Delta_l$-1). Traffic is assumed to be spread across multiple routes.

Smaller instances of this problem can be solved with some commercial optimization tool, for example CPLEX, [Cplex]. For larger instances the numerical intractability of the MILP formulation requires efficient heuristics. Heuristic approaches to solving the virtual topology design problem, are based on a specific heuristic for selecting the virtual topology combined with the optimal flow routing, the problem which can be formulated and solved using methods for the multi-commodity flow problem. Heuristic approaches for selecting the virtual topology can be divided in four classes: the heuristic solutions to the MILP problem, maximization of the single-hop traffic flows, maximization of the single-hop and multi-hop traffic flows, and embedding of the regular logical topologies. References [Leon00] and [Dutt01], provide an extensive overview of the major approaches. Here we include several heuristics, which inspired many further work and also provided important guidance for our approach.

Heuristic selection of virtual links can be done following different objectives. The Single-hop Maximization Logical Topology Design [Leon00], formulates the design problem as a maximum weight matching problem. The network is presented as a bipartite graph and the uni-directional traffic requirement $t_{i,j}$ between each nodes i, j, is taken as a weight for the edges on the bipartite graph. The problem can be stated as:

$$max f = \sum_i \sum_j b_{i,j} t_{i,j} \tag{4.6}$$

subject to

$$\sum_j b_{i,j} \leq \Delta_O^i \tag{4.7}$$

$$\sum_i b_{i,j} \leq \Delta_I^i \tag{4.8}$$

Solution to the problem can be obtained, e.g., by transforming it into a max-flow problem [Bert98]. This solution however, includes only a single virtual link between two nodes. To obtain a solution with parallel virtual links, which would require a model of larger complexity, two heuristics are proposed [Leon00]. Increasing Multi-hop Logical Topology Design (I-MLTD) heuristics, starts from the unconnected topology and proceeds in a number of iterations. In each iteration a matching problem is solved and a corresponding set of virtual links is inserted in the topology. Initially the edges of the bipartite graphs weights are proportional to the traffic demands; this results in inserting direct links for the maximal traffic requirements first. Next, traffic routing over the inserted links provides a new weight: the traffic weighted hop count $f_{i,j} \times h_{i,j}$ calculated for pairs of nodes that still have available logical terminations is used as weight on a bipartite graph. Accordingly, if a high traffic demand between two nodes is routed over several hops the cost of the edge will be high, and in a new iteration of solving maximum weight matching problem, it will be assigned a direct link. Again, all new links are inserted into the topology, and a new iteration starts. The procedure ends when there are no more available logical interfaces. An alternative approach is taken in the Decreasing Multi-hop Logical Topology

Design (D-MLTD), which starts from the fully connected topology and proceeds with weighting the edges in a bipartite graph proportionally to the weighted hop count. The solution to the minimum matching problem identifies virtual links that should be deleted from the topology. However, only those lightpaths are deleted that do not disconnect the topology. Several iterations of traffic routing, edge weighting, minimum matching, and link deleting may follow. As the algorithm starts with the violation of the logical degree constraints, i.e., all nodes are connected into a full mesh, it stops when non none of the degree constraints is violated. One lower complexity heuristics is Heuristic Logical Topology Design Algorithm (HLDA) [Rama96], also referred to as a greedy logical topology design algorithm (GLTDA). HLDA is also an iterative heuristics, which in each cycle first orders all (s, d) pairs according to their importance, and then inserts virtual links for (s, d) pairs starting from those with the highest importance. In the first cycle the importance is equal to the traffic requirement; for each (s, d) for which in this iteration a virtual link is successfully inserted in the topology a new importance is calculated as the difference to the importance of the (s, d) pair with the next lower importance. For an s, d pair for which a link cannot be inserted because either s or d node have no more available logical terminations (logical topology degree constraint), the importance is set to zero. This is repeated in all further iterations. The algorithm terminates when the importance of all s, d pairs is set to zero. The last step includes random insertion of virtual links between the remaining available logical link terminations, until the total od $N\Delta_l$ edges are established.

Still another alternative approach, more concerned with the resource usage in the physical layer, is a Minimum-Delay Logical Topology Design Algorithm (MLDA) [Rama96]. This heuristics creates a pair of edges for each physical edge, i.e., it replicates a physical topology in the logical layer, and the remaining edges are added as in HLDA. This approach is possible only is the logical degree of nodes is larger than the degree of the physical topology, a node should also be able to terminate at least one single-hop lightpath on each of its physical (WDM) interfaces. This heuristic combines a traffic independent procedure, i.e., replication of the physical topology, and a traffic dependant procedure, i.e., HLDA. A fully traffic-independent approach is taken with the Traffic Independent Logical Topology Design Algorithm (TILDA) [Rama96]. The idea is to create virtual topology of short virtual links, by first replicating a physical topology, and then adding two-hop links, three hop links, etc., until all lightpath terminations are used. In the performance studies presented in [Leon00] and [Rama96], the performance of traffic routing over topologies created with these heuristics are compared to the performance obtained over random topology.

The results of the performance studies have shown that none of the topology design heuristics shows consistently better performance for different traffic patterns and different nodes' and links' configurations. Important result is that the optimal routing plays the dominant role, and is the most important part of the problem. With the optimal routing even randomly created logical topology show performance not more than 10% lower than more complex approaches. The multi-hop approach based on the deleting logical links was shown to be superior in many situations. Regarding less complex heuristics the MLDA heuristics, which replicates the physical topology, and than inserts direct links for high traffic requirements (as HLDA), was shown to often perform better than pure HLDA.

The topic of virtual topology reconfiguration also provides an important input for our work. The reconfiguration of the static virtual topology is a reaction to a new traffic state in the network. General assumptions is that a new traffic can be expressed with a new traffic matrix: considered in time at the beginning at each new interval T the input to the reconfiguration problem is the old virtual topology and the new traffic matrix. The reconfiguration process hence transitions the old virtual topology into a new topology by adding new lightpaths or deleting lightpaths. The reconfiguration process may be characterized with the cost of equipment configuration and is constrained with the need to perform actions with the lowest impact on the traffic that is already in the network, i.e., the number of deleted or added lightpaths should be minimal. Reconfiguration is addressed in a number of studies, in which a specific mixed integer linear programming problem formulation is combined with the heuristic approach for practical optimization in larger network topologies. In [Bana97] a MILP based approcah is proposed that combines, in the first step the virtual topology design for a new traffic matrix, and in the second step minimization of the number of added or released lightpaths, for the constrained maximal congestion. In [Rama00] a similar alternative approach is presented which in the second step constrains both the number of changes in lightpaths and in the physical layer, and minimizes the weighted hop-count in the network. In [Genc03] a MILP formulation is given for a one-step adaptation of the virtual topology, which solution shows a lightpath that should be either deleted or added. By regularly solving the MILP, the virtual topology is transitioned from one into another state, one step in time. In addition, a heuristic approach was proposed which monitors new traffic requirements and the load on network links. Again, in each iteration either one virtual link is added or one is deleted from the topology. Adding a new link happens either if there is a new traffic requirement, the highest of which is then taken and accommodated on a new virtual link, or if for a multi-hop traffic over a highly loaded link a new direct link can be opened. Deleting a link happens if a traffic can be rerouted from the link with the lowest utilization. Recently, numerous further approaches are proposed in the literature, which either focus on more elaborate monitoring, e.g., [Bhan05], or on a particular methods to solving the NP-hard problem, e.g., as genetic programming [Saha05].

The RToE methods for infrastructure service provisioning, can most benefit from relatively low-complexity heuristics, such as MLDA and HLDA, because of the particular service requirements that we describe next.

4.3 Infrastructure Service Model

The Infrastructure Service is a L1 virtual network which implements mechanisms for the dynamic virtual topology creation and maintenance. These mechanisms are inspired by the methods for virtual topology design and reconfiguration which periodically re-configure the virtual topology and set up the traffic routes for a known traffic matrix. However, the infrastructure service resource management must account for different operational assumptions and constraints: (1) The service traffic is comprised of TDM layer LSPs, which setup or release can be dynamically requested, (2) service bandwidth requirements, or bandwidth matrix, cannot be fully specified, and (3) services are of shorter duration, with rather dynamic traffic changes.

These constraints require dynamic, instead of static, approach to the service topology creation. The difference between the static approach and the dynamic infrastructure service approach can be described as follows: With a static approach we can use some estimation of the traffic - i.e., bandwidth - requirements between the service members of all known VPN services, and design and configure a dedicated set of lightpahs per VPN that satisfy these requirements, e.g., as proposed [Rama02]. With the dynamic approach infrastructure services create, use and reconfigure their topologies directly reacting on the traffic demand between the service customers. Such dynamic services does not require a fully dedicated virtual topology, on the contrary, because of the anticipated dynamically changing traffic we can assume that such topology would have low utilization, hence too high a cost. On the other hand, performing too many configuration actions, i.e., configuring and releasing lightpaths (virtual links), is not desirable in the transport network. The time needed for the configuration of a lightpath is not negligible, and can even be longer than a required connectivity. The infrastructure service model addresses these requirements by introducing the concept of *resource visibility*, which enables time-sharing of physical and logical resources between a customizable set of services. The goal of our approach is to achieve high throughput, and less reconfigurations for network carriers who provide transmission, multiplexing and switching resources on the one hand, while facilitating efficient bandwidth use, i.e., high utility of purchased bandwidth for dynamic services on the other.

The visibility concept is a control plane concept: it extends the GMPLS TE information with an attribute which describes service-specific resource usage policies, and at the same time makes resources *visible* to one or more services. Within the control plane context, we refer to it as the *Resource Visibility Attribute.*

In a network graph that captures the traffic engineering properties of infrastructure resources, the visibility attribute applies to each vertex and edge, and may be simply a set of service identifiers. Accordingly, resource visibility is defined as follows.

Any arbitrary transport resource r, is **visible** *for a service S_s, if for a resource attribute $\vartheta(r)$ it holds that $S_s \in \vartheta(r)$.*

The concept of visibility extends the TE paradigm as we know it today. Today, TE attributes and link routing weights are modeled after the resource-specific TE attributes, such as link residual bandwidth [Katz03], or topological features regarding distribution of available paths [Bout02], [Kar00], [Bagu04]. With the visibility attribute, however, each network interface is additionally tagged with a set of services that can use it, and accordingly appears only in the routing graphs of these services.

4.3.1 Algorithmic Support

The visibility concept requires two types of algorithms in the control plane, which operate on different time scales.

To the first group belong the algorithms that compute *the amount and type of physical resources* which should be made visible to each service, at the time of its creation, and later during its lifetime. In the control plane these methods can be implemented within the CPIs of the physical resource owners, and the algorithms for the virtual topology design and reconfiguration can be adopted for this purpose. With this approach such algorithm will take as an input the estimation of service bandwidth requirements, but the output of the algorithm will not be used to physically configure virtual links, i.e., lambda switched paths (LSC LSPs), but to set the visibility attribute of physical resources along the calculated paths. Repeating this process for different infrastructure services, as they emerge, results in making some resources visible to several services. As an illustration, Figure 4.1 shows a network partitioned into three physical topologies which are used by several services. When a new service is created, e.g., service S_3, the visibility of both T_1 and T_2 is enhanced with service S_3, hance S_3 will time-share physical resources with S_1,S_2,S_4 and S_5. During

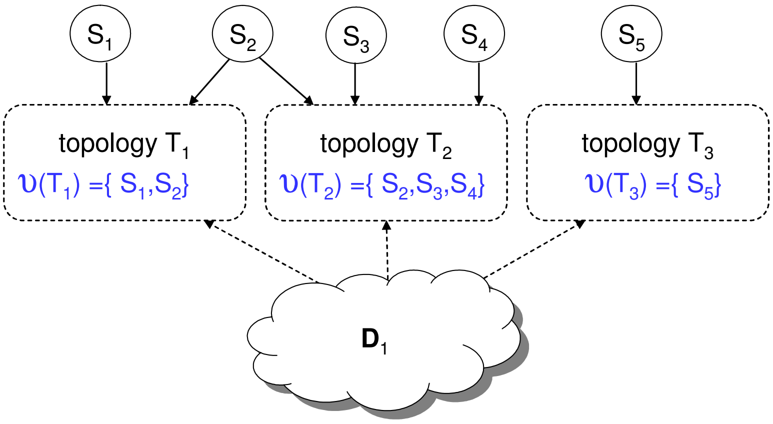


Figure 4.1: *Partitioning of resources based on the visibility attribute.*

a service lifetime, a set of visible *physical* resources can be increased or decreased, as a result of service performance and resource monitoring.

To the second group belong the methods that dynamically *maintain service virtual topology*, using the visible physical resources. We refer to these as service topology engineering (ToE) methods. With ToE methods, *logical* resources - i.e., virtual links - are created, by physically configuring lightpaths (LSC LSPs). These LSPs can be also dynamically released reducing the bandwidth bound within the virtual topology. For the time in which a virtual link is active - i.e., configured - it also has specific visibility. In this sense, ToE methods dynamically create set of *visible logical resources*. Based on the configured visibility, the bandwidth of each virtual link, i.e., an LSC LSP, can be used by a number of services, i.e., it can carry TDM granular LSPs of these services. In the control plane, ToE methods involve service CPIs, MPE control overlay, and the CPIs of carrier network, and can assume different level of control coordination, and TE information advertisement, between different CPIs.

System featuring these two groups of methods can largely benefit from an *automatic control loop*, in which traffic and resource usage are continuously monitored, and a system changes resource visibility to adapt to new service requirements and achieve more optimal resource usage. Such *autonomic management* approach is considered to be a promising, but also a challenging task [Herr05]. In this thesis, we focus only on methods for traffic routing and topology engineering for infrastructure service provisioning. However, in the simulation study we show a simple example in which changing physical resource visibility based on monitoring improves the performance of the service. More advanced autonomic optimization approaches for dynamic visibility adaptation are out of the scope of this work, yet show a promising direction for future work. In future studies we may include dynamic changes in physical resource visibility, resulting from performance monitoring, that occur on a time scale larger than that of the dynamic service topology engineering.

4.3.2 Infrastructure Service Resource Sharing Concept

Based on our technological assumption that at the customer side, the setup of lambda switched LSPs (LSC LSPs) as well as the TDM granular LSPs can be initiated, the service resource model captures two-layer resources that can be jointly used - i.e., shared - between a number of services. The two-layer resources relate to each other according to a simple client-server principle: Between two interfaces with TDM and lambda switching (LSC) capability, a server-layer LSP - i.e., an LSC LSP - appears in the client layer (TDM) as a dynamic TE link, i.e., a forwarding adjacency (FA), with the bandwidth reflecting the multiplexing capability of the interface (e.g., for OC-192, 192 $\times$ OC-1 basic signals are multiplexed within one wavelength). A client layer LSP - i.e., a TDM LSP - is routed over TE links in the TDM layer, if these are available. If the client layer topology is not connected, the missing TE links will need to be established by setting up new server layer LSPs between the unconnected interfaces. The complexity of this apparently simple relationship between the dynamic client-layer TE link and the underlying server-layer LSP, increases when services are multi-granular, and several grooming layers are supported.

Figure 4.2 illustrates a simple scenario in which a TDM LSP is established over two virtual links (TDM FAs), which on their own are a representation of LSC LSPs in the TDM layer. By using the visibility concept, an infrastructure service S_s can make a virtual link (LSC LSP) that it configures, visible to other services, e.g., S_z. This link can therefore be used to route traffic flows (TDM LSPs) of either service S_s or S_z, or both. In the example in

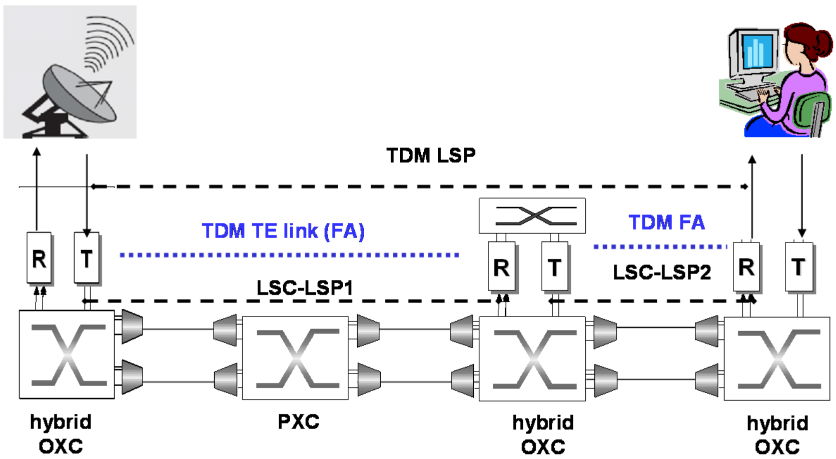


Figure 4.2: *GMPLS Client-Server Layer Relationship*

Figure 4.2, assuming that there are three services, S_1, S_2 and S_3, which share resources among them, LSC LSP1 may be set up by S_1, LSC LSP2 by S_2, and TDM LSP may transport the traffic from service S_3. Sharing means that the bandwidth of an LSC LSP which is temporarily not used by a particular service who set up this LSP may be used by a customizable set of other services.

4.3.3 Infrastructure Service Topology Model

In a network with TDM and lambda switching capability (LSC), the visibility of TDM grooming resources has significant impact on the flexibility of service topology, and consequently on the resulting service performance. The topology model established for grooming networks, which makes distinction between the physical and the logical (virtual) topology is relevant for the definition of the infrastructure service, and is illustrated in Figure 4.3.

In this two-layer network, the topology in a lambda switching (LSC) layer is physical, hence static; the TDM layer has flexible (virtual) topology, with static, semi-static, and dynamic TE links. For example one virtual link may be established over a single hop WDM link between one CE and one PE node, and the other one may be established by configuring a multi-hop LSC LSP.

A virtual link can have a static bandwidth, or can be bandwidth-extensible. For example

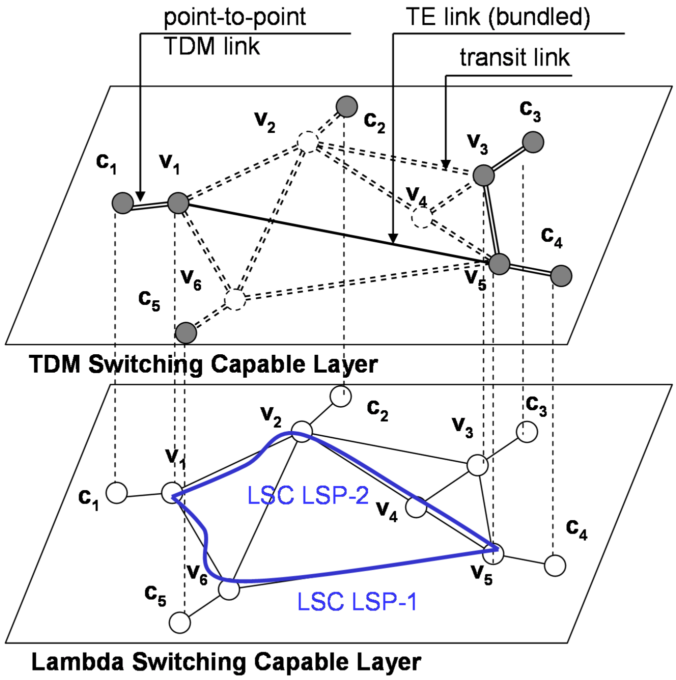


Figure 4.3: *LSC and TDM Layer Topologies. LSC topology is static. TDM topology includes static links, dynamic links and semi-static links.*

a single hop WDM link with 8 wavelengths each carrying SONET OC-192 can be configured as one virtual link with static bandwidth equal to 8×192 OC-1. An example of a bandwidth-extensible link is shown between PE nodes v_1 and v_5. It has two *component links*, the one established over LSC-LSP-1 and the other over LSC-LSP-2. A bandwidth-extensible link uses GMPLS bundling methods to dynamically aggregate its component links. Regarding different switching capability, Figure 4.3 shows nodes with switching capability in the TDM layer as full-line drawn vertexes in a TDM layer topology, and the others are drown with dashed lines. For example, PE nodes v_2 and v_6, and a P node v_4, do not support grooming at the TDM layer, and are therefore not visible (shown dashed) at a TDM layer. By applying the visibility concept, even some existing capability, e.g., TDM layer grooming as some nodes, may not be visible to a set of services. Hence, nodes shown dashed may be TDM grooming capable, but temporarily "invisible" nodes.

Transit Edge Model

The resources that infrastructure services use to accommodate traffic are on the one hand *visible physical resources*, such as LSC layer links and switching capability, and on the other *visible logical resources*, i.e., dynamically configured virtual TE links (FAs) between interfaces with TDM switching capability. To create a joint topological representation of physical (LSC layer) and logical (TDM layer) resources, we introduce the notion of a *transit edge* and a *transit path*. In this joint topological representation hybrid nodes with both LSC and TDM layer switching capability are shown as vertices drown in full line, and the optical nodes with only LSC interfaces are shown in dashed line. A transit edge represents an LSC layer link (WDM link) between lambda switching capable nodes, either hybrid or optical. A connected path of transit edges between two hybrid nodes is a transit path. A transit path can be used to set up an LSC LSP. Accordingly, a hybrid node can terminate a transit path, i.e., be an ingress-, or an egress-node for an LSC LSP, or be an intermediate node on the path, i.e., switch LSC LSP in the LSC layer. We denote an LSC LSP that is switched in the optical layer of a hybrid node as a flow-through LSP for this node, and the node itself as a flow-through node for this LSP.

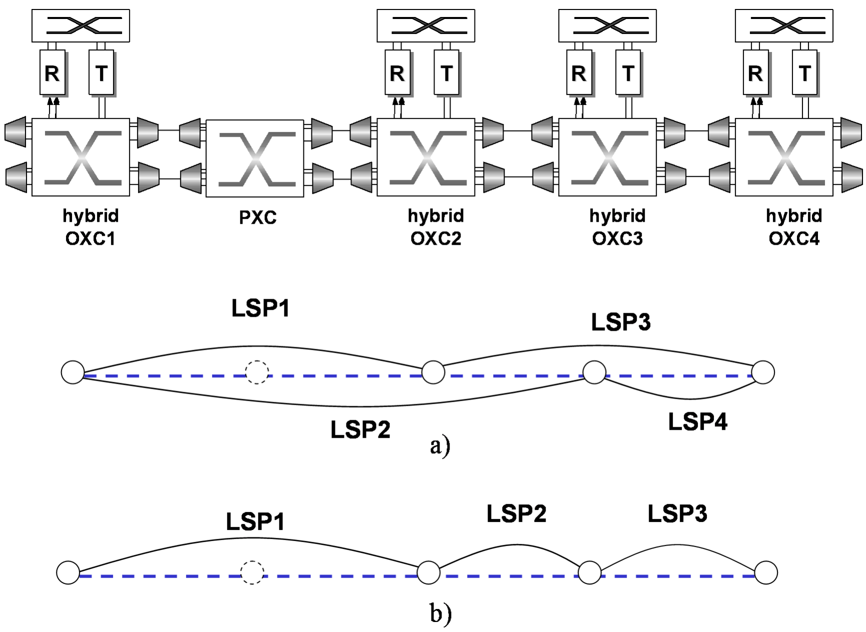


Figure 4.4: *Transit Edges. (a) Two different configurations of connected transit paths: (1) LSP1 and LSP3, (2) LSP2 and LSP4. (b) Configuration with shortest connected transit paths.*

This concept is further illustrated in Figure 4.4, where four hybrid OXCs can either switch LSC LSPs in the optical domain, or terminate them in the TDM layer. Two scenarios show LSC LSPs which may be established in order to connect OXC1 and OXC4. In the scenario (a) two possible combinations are shown: either LSP1 and LSP3 are established, where LSP1 is routed over a two-hop transit path with one intermediate photonic crossconnect (PXC) - i.e., a pure transit node - and LSP3 is a two-hop path with one flow-through node (OXC3), or LSP2 and LSP4 are established, where LSP2 is a three-hop transit path with one transit and one flow-through node, and LSP4 is a single hop between OXC3 and OXC4. In the scenario (b) LSPs are established over the shortest possible transit paths - i.e., hybrid OXCs - are not used as flow-through nodes. We use the concept of transit links and transit paths, when we define the Routing and Topology Engineering (RToE) strategy for infrastructure services that computes the path in the *combined topology view* comprising virtual and transit links (Section 4.5).

Virtual Topology Constraints

Different TDM layer visibility can result in different service layer topology. Figure 4.5 shows three examples of service layer topology, for the network domain of Figure 4.3. In the first example, (1), the TDM grooming resources in a provider domain are not visible to a service. A TDM topology may therefore be any subset of edges from the full mesh between the CE nodes. In the second example, (2), TDM grooming is fully deployed and visible, and a service topology can be constructed including CE-PE links and PE-PE links. Additional constraint may be that only direct CE-PE links are allowed (e.g., c_1-v_1, but not c_1-v_5) and all PE-PE dynamic links are allowed. Third example, (3), shows a constrained virtual topology, in which only a pre-defined set of edges can be dynamically established and extended during the service life time. Accordingly, service layer topology may either be designed for, or, adapt to the traffic and bandwidth requirements of service users. For example, a bandwidth matrix specifying maximum bandwidth requirements between different service users can be used as an input to a design algorithm. A virtual topology design heuristics may be used to find a set of virtual links, such as the one in the example (3). If such information is not available dynamic adaptation of topology is needed to cope with the changing demands. The infrastructure service model belongs to a class of extensible services which dynamically adapt its topology to the traffic needs, without assuming any pre-established virtual topology. The only constraint for the virtual topology in our model is therefore the visibility of grooming resources.

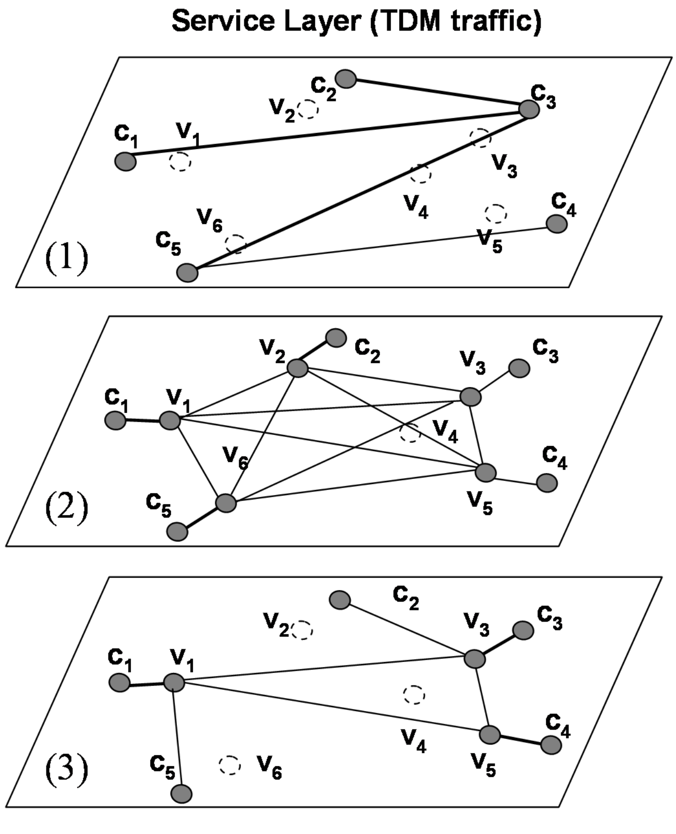


Figure 4.5: *Service layer topologies for different grooming visibility levels.*

4.3.4 Visibility and Control Plane Interconnection Models

It is important to notice that the resource visibility concept is different from the GMPLS control plane interconnection models, i.e., peer, overlay, or augmented [Mann04]. Interconnection models define the relationship between two control plane instances in terms of the advertised topology information. Resource visibility, on the other hand, is a service-specific resource usage policy. Resources available for service accommodation, i.e., the visible resources, may either be advertised into the corresponding service control plane instance (peer/augmented model), or used for service accommodation within the provider control plane, without being advertised to a service CPI (overlay model).

This concept is illustrated in Figure 4.6, where resource r_1, of a domain D_1, can be used by services S_1, S_2 and S_5, and resource r_2 is visible only to service S_1. In case of a peer/augmented interconnection model the resource visibility is used to customize the advertised TE information. As shown in Figure 4.6 (a) the information advertised to CPIs of S_1 and S_2/S_5 describes only visible resources, denoted as $\Sigma(r_1, r_2)$ and $\Sigma(r_1)$, respectively. Based on this information, service CPIs can create and maintain, their own Service Visibility Graphs (SVG). In case of an overlay model, the CPI of D_1 creates and maintains SVG of supported services (Figure 4.6 (b)). In other words, in the latter case, the owner of the service visibility graphs is a provider domain's CPI.

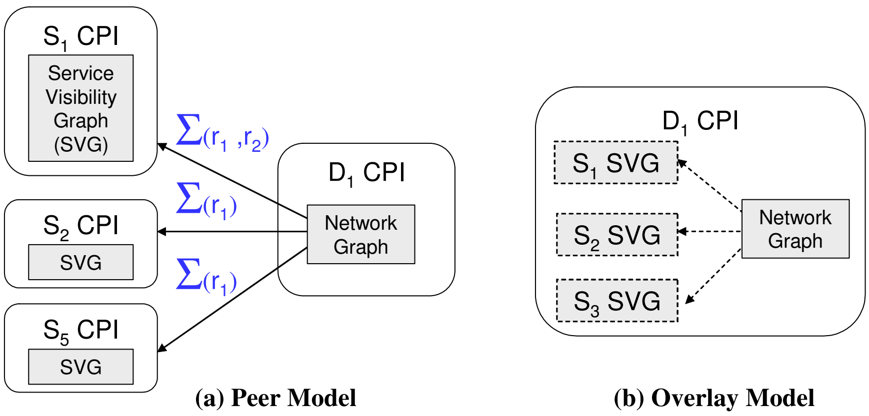


Figure 4.6: *Resource Visibility Example:* $\vartheta(r_1) = \{S_1, S_2, S_5\}$, $\vartheta(r_2) = \{S_1\}$. *(a) With the peer model the visibility information is advertised between the D_1's CPI and the services' CPIs. (b) With the overlay model visibility is used internally to the D_1's CPI.*

In both cases, flexible virtual topologies with virtual links terminated either in the customer edge nodes (CE), or in the provider domain (PE or P) nodes can be supported, by configuring the corresponding logical layer grooming and switching resources as visible for service accommodation. Which party, i.e., which CPI will actually allocate visible physical resources during the RToE procedure, depends on the control plane interconnection model. As depicted in Figure 4.6 (b), with the overlay model the visibility of P resources is not advertised into the service's CPI. Therefore, when a request to set up a new LSC

LSP is created, the corresponding path for that LSP will be computed within the network CPI (here D_1's CPI). With the peer mode, on the other hand, the TE information of all visible intermediate (PE/P) grooming interfaces will be advertised into the corresponding services' CPIs, which will use them to compute the best path for new traffic LSPs (see Figure 4.6 (a,b)).

4.4 Service Visibility Graph (SVG)

In the infrastructure service model, service provisioning is based on the methods for routing traffic LSPs on the service resource graph, which we refer to as Service Visibility Graph (SVG). We proposed two Routing and Topology Engineering (RToE) methods, which include specific SVG transformation, the single-layer or multi-layer routing on the transformed graph, and the SVG extension, i.e., update.

As depicted in Figure 4.7, SVG is a customized view of a network graph: for each transmission, multiplexing, or switching resource represented within the network graph, a ***visibility attribute*** ϑ defined as a set of service identifers, specifies a set of services which may use the resource. Consequently, for each service S_s, an SVG, denoted as $G^s(V^s, E^s)$, is created to model resources visible to S_s. SVG is maintained during the service lifetime, and is updated after each dynamic topology engineering (ToE) action, and after each accommodated traffic LSP.

The infrastructure service provisioning is modeled with a set of procedures that operate on different graph structures.

- Procedures *Create-*G^s and *CreateLayer-*$G^s(V^s_g, E^s_g)$ translate the network graph into an SVG;

- Procedures $TransfLBL_g()$ and $TransfCMB_g()$ transform the SVG into a routing graph W_g upon a particular traffic LSP request;

- Procedure $getA()$ translates the cost of resources associated with each edge in the

SVG graph into the cost of the corresponding edge in the routing graph W_g;

- Procedures $RouteLBL()$ and $RouteCMB()$ search for the shortest path on the routing graph W_g by running the Dijkstra algorithm [Dijk59];

- Procedures $InsBund_g()$ and $Update_g()$, and $DeBund_g()$ and $Release_g()$ update the SVG resources to account for a new configured LSP, or for a released LSP, respectively.

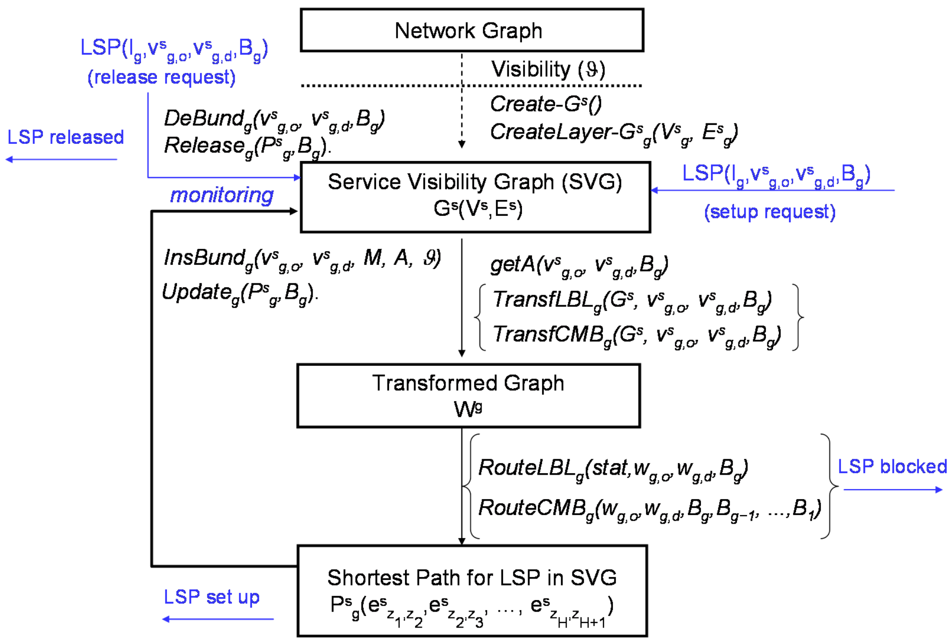


Figure 4.7: *Provisioning model for infrastructure services*

Infrastructure service provisioning combines traffic-driven RToE methods for traffic LSP accommodation and virtual topology extension, and ToE methods for link-state dependant link/bandwidth release. A requested traffic LSP can either be blocked or successfully routed, either on the existing resources or upon a successfully performed topology extension. A request for a release of an LSP will be also followed by the corresponding update of the service graph.

The service visibility graph and the procedures operating on it are defined using network graph notation described in Chapter 3, extended with several service specific definitions. We denote a set of services supported in a network as $\mathbf{S} = \{S_s\}$, a set of granularity layers supported within service S_s as L^s, $L^s \subseteq L$, and a set of domains that make some of their resources visible to S_s as D^s, $D^s \subseteq D$. We further denote a set of GXPs that are involved in provisioning S_s as X^s, $X^s \subseteq X$. For each granularity layer $l_g \in L^s$, the network resources visible to S_s are modeled with a set of vertices V_g^s representing the visible switching capability at the customer sites, in provider domains and GXPs, and a set of unidirectional edges E_g^s representing visible intra-provider links, inter-provider links, access link, GXP links, inter-customer links, inter-GXP links, and inter-layer links, i.e., the multiplexing/adaptation capabilities in the CE nodes, PE/P nodes, and GXPs.

4.4.1 Representing Aggregated Links, Component Links, and Labels in SVG

The major feature of the infrastructure service is a GMPLS-based dynamic topology adaptation: a service visibility graph becomes dynamically updated when either virtual links are added/deleted, or extended/reduced in bandwidth. The infrastructure service resource management uses GMPLS notion of link bundling that models the aggregation of parallel component links between the same interfaces. This is a means to achieve routing scalability: the parallel component links are aggregated within one bundled virtual link, and only the TE information of the bundled link, such as the available bandwidth and the aggregated routing cost, is distributed between the routing controllers and used for the path computation.

Consequently, in the service visibility graph, the bundled (aggregated) links and the component links have different representations: while bundled TE links are represented as graph edges, the component links are represented as objects associated with these edges. The component links comprise atomic discrete bandwidth blocks, referred to as *labels*. Correspondingly, for each edge $e_g^s \in E_g^s$, the information maintained in SVG includes the total bandwidth $B(e_g^s)$, given as a number of bundled labels at layer l_g, the residual bandwidth, given as the number of available labels $R(e_g^s)$, the visibility $\vartheta(e_g^s)$ and the edge weight (cost) $A(e_g^s)$. Within each aggregated link, the state of its component links and labels is also maintained in SVG. For example, each component link e^{comp} has a bandwidth

$B(e^{comp})$ given as a number of the contained labels. This number depends on the multiplexing capacity of terminating ports between two layers: a wavelength switched path (LSC LSP) established between the ports capable of multiplexing and de-multiplexing four TDM containers can be represented as a TDM component link with the capacity equal to four labels. The weight of the component link may depend on the length (in hops) of the underlying server layer LSP, and hence component links routed over different paths and bundled together within one virtual link (one edge in SVG) may all have different weights. Each label can either carry traffic or not, and have state either "used" or "unused", accordingly.

4.4.2 Locating SVG within the Global Control Process

In the infrastructure service model, the Service Visibility Graph is a data structure that can be established in different instances of the control plane. Before describing the routing and topology engineering methods on the SVG graph we briefly tackle the issues of SVG location within the global control process, and the issues of SVG update.

As previously mentioned, an important aspect of infrastructure service provisioning is the availability of the TE information and the dynamics of updates between the control plane instances involved in provisioning. Figure 4.8 illustrates interworking between the control plane instances, in a network with three domains D_1, D_2 and D_3, one GXP, and two customers c_1 and c_2 of infrastructure services S_1, and S_2. Here, PE and P devices may be PXC or OXCs, and GXP is a grooming node (OXC), hence, TE information is maintained for resources at the TDM and the LSC layer. The S_1 CPI (at c_1) interconnects with D_1 CPI, to either get the network information for the source based routing, or to get a route directly. The S_2 CPI (at c_2) uses MPE (through GXP) acting either in a role of a routing proxy, or a routing service. CPIs of S_1 and S_2 can also exchange information directly, for example if they share some common resources.

Regarding the maintenance of the service visibility graph, the location of SVG depends on the CPI interconnection models, i.e., whether or not the TE information is advertised, and

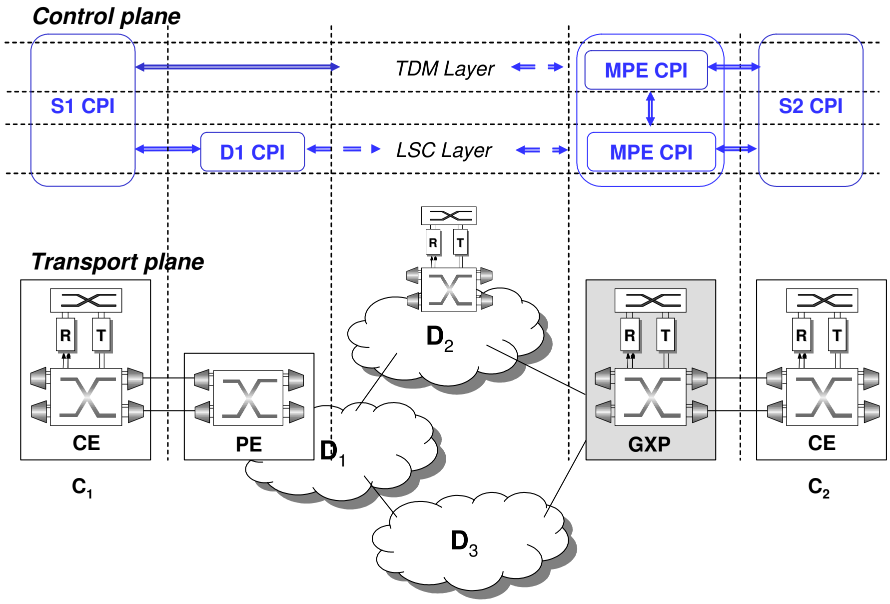


Figure 4.8: *Architecture of the SVG update process.*

the resulting routing model, i.e., a source-base routing, hop-by-hop routing, or a routing service. If a global routing service is deployed, an SVG can be centrally maintained. With the source based routing, an SVG will be crated at the service CE nodes, and updated based on the advertised TE information. The location of SVG, and the scope and the rate of its update, will affect its accuracy and the performance of service provisioning. Different possible scenarios are described as follows:

(1) SVG at the Service Member / Source Routing: An SVG is created at each customer edge node, and is updated periodically with a given rate. Hence, topology information is accurate only immediately after update, and the accuracy gradually degrades until the next update. Between updates, the path computation process at the source CE uses the static information provided on the last update. This information can be enhanced with the information about the dynamic links terminating at this particular source nodes.

(2) SVG in the Access Domain / Local Routing Service: This model assumes a routing service implemented in the access domain and the periodic TE information update between domains. Therefore, the path segment computed within an access

domain is always accurate, which is not the case for other path segments as they are computed based on the static network information of the last update.

(3) SVG in the MPE / Global Routing Service: In this case SVG is a centralized, always accurate, representation of service resources, hence the exact topology information is always used for computing the path. This is an ideal centralized approach which provides the upper bound of the performance, as none of the LSPs is blocked due to the stale TE information.

The study on how stale information affects the performance of provisioning, and whether the updating process should be periodical, or on a specific event, are out of the scope of this thesis. However, the SVG model shown in Figure 4.7, may be extended to include two instances of SVG, the "actual SVG" and the "recent update". This is illustrated in Figure 4.9.

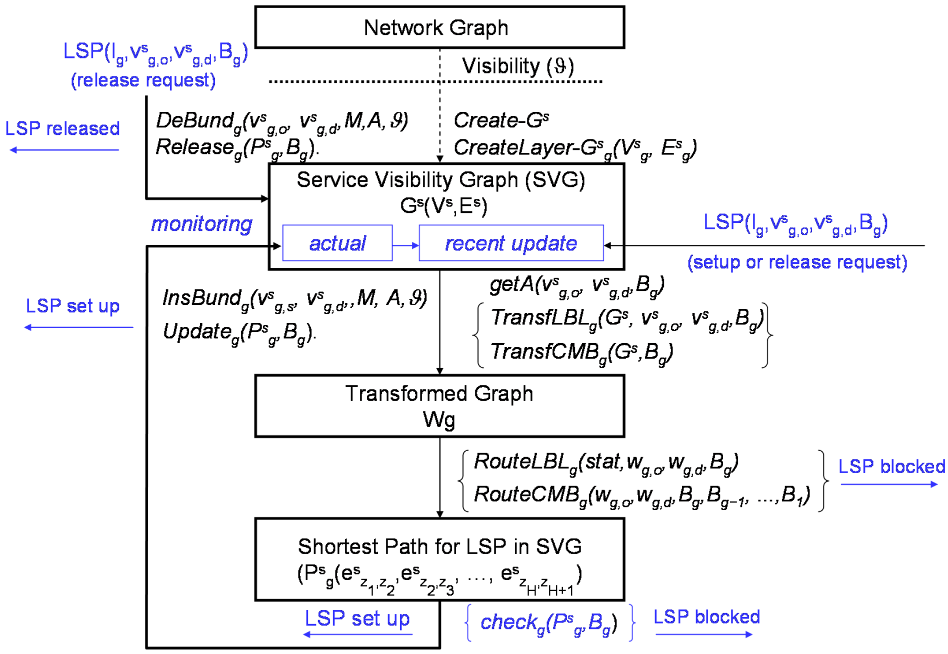


Figure 4.9: *Infrastructure service provisioning model with SVG accuracy representation.*

The "actual" SVG holds information about the actual resource usage. When a new virtual TE link is created by a service S_1, it is added to the actual SVG of S_1. This TE link is also assigned a visibility attribute depending on the attributes of resources used for its creation, and would be therefore, in a real system, periodically advertised to other services which are allowed to use it. A "recent update" SVG is produced by periodically copying the state of an actual SVG. Depending on the type of the updating process some of the changes in the actual SVG may propagate into the "recent update" SVG. For example for the third model the "recent update" is equal to the "actual" SVG. For the first and second model the "recent update" may be enhanced with either only the links terminating on the source node, or the actual topology and link state information of the access domain. When the impact of the update process is taken into account, a blocking may occur when resources cannot be allocated along the computed path. Therefore, before the procedure $Update_g(p^s_g, M)$ is performed it is first checked on the "actual" SVG, (corresponding to a distributed signalling attempt in a real network) whether the resources are available for this path. The output of the procedure $check_g(p^s_g, M)$ may therefore show that an LSP request is blocked, as depicted in Figure 4.9.

In the following sections we describe the details of the infrastructure provisioning based on the SVG creation, maintenance and transformation.

4.4.3 SVG Creation

The SVG creation is modeled with the procedure Create-G^s, described below.

<u>*Create-G^s*</u>

Step-1: for each $l_g \in L^s$ run *CreateLayer-*$G^s_g(V^s_g, E^s_g)$

Step-2: for each $v^s_{g,i} \in V^s_g$

if $(v^s_{g-1,i} \in V^s_{g-1}))$

{ insert $e^{sm}_{g,g-1,i}$ and $e^{sm}_{g-1,g,i}$ in G^s,

initialize $B()$, $R()$, $\vartheta()$ of $e^{sm}_{g,g-1,i}$ and $e^{sm}_{g-1,g,i}$ with values from the network graph,

set $A(e^{sm}_{g,g-1,i})$=1; set $A(e^{sm}_{g-1,g,i})$=1 }

Step–1 iterates through layers and invokes procedure *CreateLayer-*$G^s_g(V^s_g, E^s_g)$, which populates the layer specific graph (G^s_g) of SVG with all visible resources within all domains. In each layer l_g, the vertices representing visible resources of customer sites and provider domains within the network graph, are mapped into a set of vertices V^s_g. Correspondingly, a set of edges E^s_g represent all edges from the network graphs, visible to S_s at layer l_g.

In Step–2, for each pair of adjacent subgraphs, (G^s_g, G^s_{g-1}), we interconnect the corresponding vertices representing different switching capability at the same node, with edges modeling multiplexing capabilities. Unidirectional edges $e^{sm}_{g,g-1,i}(v^s_{g,i}, v^s_{g-1,i})$, and $e^{sm}_{g-1,g,i}(v^s_{g-1,i}, v^s_{g,i})$ are referred to as service inter-layer edges.

The total bandwidth, residual bandwidth, weight, and visibility of all edges in SVG, including inter-layer edges, are initialized with the corresponding values from the network graph (in *CreateLayer-*$G^s_g(V^s_g, E^s_g)$ and in Step–2). The edge weight of all inserted edges is set to a default value, here 1.

The overall complexity of *Create-*G^s is determined by the number of visible links and edges and number of layers. The maximum number of vertices is bounded by $|L^s|(|C|+|N||D^s|)$, where N is the max number of nodes per domain, $|D^s|$ is the number of service domains, $|C|$ is maximal number of service members, and $|L^s|$ number of granularity layers visible within a service. The number of visible links is bounded by $|L^s|(|D^s||C| + |X^s||D^s|(|D^s| - 1)/2 + |D^s|N(N-1)/2)$, where $|C||D^s|$ is a number of customer access links (CE-PE). $|X^s||D^s|(|D^s|-1)/2$ is the maximum number of interconnections at GRX, where $|X^s|$ is the number of GRX supporting the service, and $|D^s|N(N-1)/2$, is the maximum number of inter-domain edges in all domains. Hence, the complexity is bounded by $O(|L^s|(|D^s|(|C| + N) + |X^s||D^s|^2 + |D^s|N^2))$.

4.4.4 SVG Maintenance

The SVG maintenance is modeled with four procedures: $InsBund_g(v^s_{g,o}, v^s_{g,d}, M, A)$, $Update_g(p^s_g, M)$, and their counterparts $DeBund_g()$, and $Release_g()$.

Procedure $InsBund_g(v^s_{g,o}, v^s_{g,d}, M, A, \vartheta)$ is invoked when a server layer LSP, i.e., a topology extension LSP, is set up. It inserts in SVG (i.e., the layer graph G_g^s) the object for the corresponding dynamic component link $e_{g,o,d}^{scomp}(v_{g,o}^s, v_{g,d}^s)$, established between vertices $v_{g,o}^s$ (origin) and $v_{g,d}^s$ (destination), with a component weight A, bandwidth in terms of number of contained labels M, and visibility ϑ_R.

$\underline{InsBund_g(v_{g,o}^s, v_{g,d}^s, M, A, \vartheta_R)}$

Step-1: if ($e_{g,o,d}^s \in E_g^s$) {goto **Step-2**} else {insert $e_{g,o,d}^s(v_{g,o}^s, v_{g,d}^s)$ in E_g^s}

Step-2: create $e_{g,o,d}^{scomp}(v_{g,o}^s, v_{g,d}^s)$ and associate it with $e_{g,o,d}^s$

set $A(e_{g,o,d}^{scomp}) = A$, $B(e_{g,o,d}^{scomp}) = M$, $\vartheta(e_{g,o,d}^{scomp}) = \vartheta_R$

Step-3: set $B(e_{g,o,d}^s) = B(e_{g,o,d}^s) + M$; set $R(e_{g,o,d}^s) = R(e_{g,o,d}^s) + M$;

Step-4: set $R(e_{g,g-1,o}^s) = R(e_{g,g-1,o}^s)$ - M; set $R(e_{g-1,g,d}^s) = R(e_{g-1,g,d}^s)$ - M

In Step–1, a new aggregated virtual link is created between $v_{g,o}^s$ and $v_{g,d}^s$, if such does not already exist. In Step–2, an object representing a component link denoted as $e_{g,o,d}^{scomp}$ is created, with bandwidth M, weight A and visibility ϑ_R, and is associated with an aggregated virtual link, i.e., and edge $e_{g,o,d}^s$. In this way an edge in SVG is always maintained as an aggregated dynamic link to which a number of component links may be associated, each of which has its own weight, bandwidth in terms of discrete labels, and visibility. The aggregated dynamic link is further assigned new total bandwidth, by adding a capacity M to the previous value, and the residual bandwidth is also updated (Step–3). Component link $e_{g,o,d}^{scomp}(v_{g,o}^s, v_{g,d}^s)$ uses multiplexing capacity between layers l_g and l_{g-1}, therefore the bandwidth of the corresponding multiplexing edges is reduced by M at the origin and the destination nodes of the link (Step–4).

Procedure $Update_g(p_g^s, M)$, described as next, changes the state of M labels along the path p_g^s in SVG, to reflect the allocation of resources for an extension LSP (a component link). Path p_g^s is expressed as an ordered set of edges in SVG, i.e., $p_g^s(e_{g,z^1,z^2}^s$, e_{g,z^2,z^3}^s, ..., $e_{g,z^H,z^{H+1}}^s)$, and is an output from the path computation (routing) performed within a RToE method described later in Section 4.6.

$Update_g(p^s_g(e^s_{g,z^1,z^2}, e^s_{g,z^2,z^3}, ..., e^s_{g,z^H,z^{H+1}}), M)$:

Step-1: for each $e^s_{g,z^i,z^j} \in p^s_g$

{select M labels of an aggregate link e^s_{g,z^i,z^j} mark them as used

set $R(e^s_{g,z^i,z^j})= R(e^s_{g,z^i,z^j}) - M$}

Step-2: (End)

In Step-1, for each edge along the chosen path, M labels are allocated and their state changed to used. The edge residual bandwidth attribute R is reduced by M which is the bandwidth of the service traffic LSP. The complexity of the procedure is bounded with $O(H_m U_m \log U_m)$ where H_m is the maximum number of hops in the path and U_m is the maximal number of containers in one link.

4.4.5 Edge Weighting on a Routing Graph

As illustrated in Figure 4.7, when a request for a traffic LSP arrives, a transformation of the Service Visibility Graph is created, and the request is routed on it. The transformation of SVG is a routing graph, which may represent resources of one or several layers. Omitting the layer indexes, we denote the transformed graph its vertices and edges by W, w_i and $e^w_{i,j}(w_i, w_j)$, respectively. Each edge $e^w_{i,j} \in W$ is associated with an SVG edge $e^s_{i,j}$. Accordingly, a traffic LSP request $LSP(l, v^s_o, v^s_d, B)$ specified relatively to vertices in SVG, maps into a request $LSP(w_o, w_d, B)$ in W. The weight $A(e^w_{i,j})$ of an edge $e^w_{i,j} \in W$, is computed using the SVG procedure $getA(e^s_{i,j})$. This procedure offers a flexible interface which may implement various dynamic edge weighting schemes, i.e., different weighting functions operating on the edge attributes, such as the residual bandwidth, and on the weights of the component links.

As already pointed out, an SVG represents aggregated TE links as graph edges, and the component links as objects associated with them. To capture the two-layer client-server relationship between a component link and an underlying LSC LSP, the weight of a component link shall reflect the properties of the underlying LSP, in the first place its length, i.e., the number of physical hops. In our model we adopted the function proposed in [MANN04] where the cost of the dynamic component TE link at l_g is set to $A(e^{comp}) = (H-1)$ where H is the number of hops of a supporting l_{g-1} LSP. Further, the weight of an aggregated link may be a function of the weight of all its component links. For example, it may be equal to the cost of the least cost available component, or the mean value of costs of the available components. The weight of an aggregate link may also be a constant, e.g., always equal to 1, or a function of the residual bandwidth, R(e), e.g., $A(e) = 1/R(e)$, which is a common approach to traffic engineering in MPLS networks. In this way the links with higher residual capacity are preferred.

Procedure $getA(e^s_{i,j})$ which computes the weight of the bundled link as a minimum weight of all its components is described next.

$getA_g(v^s_{g,o}, v^s_{g,d}, B_g)$

Step 1: find all $\{e^{scomp}_{g,o,d}\}$ associated with $e^s_{g,o,d}$, for which $R(e^{scomp}_{g,o,d}) > B_g$.

Step 2: find $A_{min} = \min\{A(e^{scomp}_{g,o,d})\}$

set $A = A_{min}$

Step 2: (return A)

In this case the complexity of $getA_g(v^s_{g,o}, v^s_{g,d})$, is bounded by the complexity of a binary sort $O(U_m \log U_m)$ where U_m is the maximum number of component links in one FA.

It is important to note that the component link weights cannot be used for source-based computation of the route, as they are known only locally at the nodes terminating the component link. However, when a bundled link is selected to accommodate an LSP, i.e., as belonging to the selected path, then one or several labels of the associated component links become selected, and allocated for the LSP. This link-local label selection may be random or based on the component cost. For example, in the previously described procedure $Update_g(p^s_g, M)$, M least-cost labels are selected.

By applying the Dijkstra algorithm in the appropriately weighted transformed graph W, the least-cost path P can be found expressed as an ordered set of edges $P^w(e^w_{z^1,z^2}, e^w_{z^2,z^3}, ..., e^w_{z^H,z^{(H+1)}})$ where H is the hop count of P. Thereby, for a path P in W, a corresponding path, or a layer specific segment P^s_g in SVG is denoted as $P^s_g(P^w)$. The path P^s_g is used to dynamically update SVG upon the successful routing of the LSP, in procedure $Update_g(P^s_g, B_g)$.

Preference Extension for Visibility Attribute

In a scenario in which multiple carrier domains are involved in service provisioning, the procedure $getA_g(v^s_{g,o}, v^s_{g,d})$ also takes into account the domain-specific weighting scheme. For this purpose we extended the resource visibility attribute with the so called penalty, i.e., the weight multiplying factor, defined for each service S_s for which the resource is visible [Tom03a]. For example, for all resources from the domain D_k the penalty for service S_s can be expressed as $p_{s,k} \in \Re$, $k = 1, 2, .., D$ and $p_{s,k} > 1$. When the least-cost route is computed for an LSP request of S_s, the original cost of a particular visible edge of D_k is multiplied by $p_{s,k}$. If a service S_1 is defined as an intra-domain service of a domain D_m the penalty $p_{1,m}$ is set to 1. If this service is to be accommodated only domain-locally, then for any domain D_n other than D_m the penalty $p_{s,n}$ must have sufficiently large value. If domain-neutral network-global accommodation is preferred, the resources of all domains should be taken into account for routing, and each $p_{1,n}$ is set to 1 or have some variable value. In general the penalty can be specified resource-specifically instead domain-specifically. That means that each resource can be configured with different penalties for different services; this extended configuration is included in the service visibility attribute.

4.5 Infrastructure Service Topology Engineering (ToE)

The infrastructure service model assumes that the topology maintenance is based on a per-service traffic and link monitoring process, which decides whether to insert new virtual links, and whether to bundle or de-bundle component links to/from the existing links.

We assume that a per-service monitoring and the corresponding topology engineering (ToE) methods are implemented either within the service CPI, or the provider CPI. We consider two groups of ToE methods as a constituent part of the infrastructure service model. To the first group belong the Routing and Topology Engineering (RToE) methods, which integrate topology engineering actions for the virtual topology extension within the traffic routing process. These are invoked if a requested new traffic LSP cannot be accommodated on the existing virtual resources. Two complementary RToE methods proposed for infrastructure service provisioning are described later in Section 4.6. To the second group belong the stand-alone methods based on the periodic link state monitoring. In general, for each dynamic virtual link in the service topology (SVG) the utilization threshold may be defined, the violation of which can trigger either an extension or a reduction in virtual link bandwidth. The infrastructure service model includes stand-alone ToE actions only for *releasing resources* from the virtual topology. With this simple approach, it is guaranteed that new resources are added to the virtual topology only for the actual traffic need, within RToE action, and released as a result of link monitoring.

We define three link release policies, namely, *never release*, *release when idle*, and *conditional release*, which apply to any dynamically established *component* link. As already described a component link in a client layer is a representation of an LSP in a server layer and can be bundled to or de-bundled from a virtual link.

Never release (NREL) ToE policy keeps all dynamically configured component links configured for the whole service duration, even when their labels are not used.

Release when Idle (RELI) policy releases a component link if it is not used, either immediately, or during a specific measurement interval $T_{rel} \geq 0$. This can be supported by inspecting all component links (i.e., server layer LSPs) periodically, or after each configuration action. The monitoring of the link usage can be performed at the origin, or at both the origin and destination node of the link in order to synchronize the release particularly in the case of the bi-directional component links (bi-directional LSPs). Further, in our performance study, we consider only the case of the immediate release. Study on the impact of different measurement intervals, is out of the scope and could be the subject of further work.

Conditional Release (CREL) policy releases an idle virtual link according to its length in terms of the physical layer hops. When all the labels of one component link are free only the links with the length equal and larger than some configured value L will be released. In the performance study we set $L = 3$.

When designing a ToE resource release actions, it may be important to consider that the time needed to set up or to release an LSP is not negligible, and sometimes may even be comparable with the duration of the service. Therefore, the set-up time of an LSP in a particular layer should be known, e.g., as a function of the link configuration times T_c, proportional to the number of hops. When the virtual links are released when idle, it can happen that requests for LSPs are penalized with the long waiting time for a service. The study on how the decision to release a link can take the link configuration time into account is out of the scope of this thesis, and is subject of further work.

4.6 Routing and Topology Engineering (RToE)

Complementary to the ToE actions which reduce service bandwidth, the infrastructure service use Routing and Topology Engineering (RToE) methods to accommodate new traffic requests, and to extend topology if needed.

We assume that a service user, who may also be an application agent, can detect that a new traffic stream towards another user, e.g., another application agent, should be established. This need will be translated into a *request for a traffic LSP* , denoted as $LSP_g(l_g, v^s_{g,s}, v^s_{g,d}, B_g)$, where l_g is the granularity layer at which the traffic should be mapped (e.g., TDM OC-3), B_g is the requested capacity at the layer l_g, and $v^s_{g,s}$ and $v^s_{g,d}$ are LSP source and destination, respectively. From the service user perspective, a traffic LSP can have a known duration (τ), in which case for an LSP successfully set-up in t_s, the release LSP request will be scheduled at $t_e = t_s + \tau$. Alternatively, a user can also detect the need to de-allocate an LSP, which duration was not known in advance, and issue an LSP release request. When accommodating traffic LSP requests, the infrastructure service has to take care of constraints and capabilities of the transport plane, the most prominent of which are bi-directionality of links, and availability of virtual concatenation. In general, in GMPLS a connection may be a uni-directional or a bi-directional LSP. Unidirectional LSPs offer more flexibility as the capacity of each direction may be extended separately. However, some LSPs may be inherently bi-directional, e.g., due to the bi-directional transport plane. Setting uni-directional LSPs over bi-directional dynamically created links may incur ineffective usage of resources. The *virtual concatenation* capability makes traffic routing far more flexible than in the case of contiguous concatenation. With contiguous concatenation the service traffic must use contiguous blocks routed along the same route. With virtual concatenation a traffic LSP between two service customers can be mapped into a number of virtually concatenated labels, routed on a number of service layer LSPs either along the same path or along different paths, which may reduce the probability of request blocking [Kir05a]. Due to the fact that the differential delay between the virtually concatenated labels must be compensated, an appropriate admission control is needed [Alic05]. In addition, virtual concatenation may be deployed only at some interfaces in the network, which rises a need to discover them, and include this information in the routing algorithm. The infrastructure service model includes both virtual and contiguous concatenation options, however, we do not consider the issues of discovering concatenation capabilities.

To accommodate a traffic LSP request, an infrastructure service model uses a graph routing approach: first, Service Visibility Graph is transformed into a routing graph W_g, and then the requested $LSP_g(l_g, w^s_{g,o}, w^s_{g,d}, B_g)$ is routed in this graph. At the granularity layer of the traffic LSP the routing graph may be a *single layer graph* in which case it includes only representation of the *configured* resources in this layer. Alternatively, it may be a *multi-layer graph* that includes both the representation of the *configured* resources in the traffic granularity layer, and the representation of *available* resources in one or multiple server layers, i.e., bandwidth extension layers. An attempt to route a traffic LSP_g may result in either (1) finding a suitable path over existing resources in l_g, (2) finding no path, which further triggers a sequence of topology engineering actions in server layers, and a new request routing attempt, or (3) definitive blocking when extension is not possible.

If for a requested traffic LSP, a path cannot be found on the existing resources, i.e., case (2), the topology engineering (ToE) process will decide which new virtual links should be set up in order to accommodate the request. Generally, to accommodate an LSP in a granularity layer l_g of a multi-layer network, a topology extension actions may be taken in a sequence of server layers, l_{g-1}, l_{g-2}, ..., l_1. A multiplexing factor U_g defined as the maximum number of l_g labels within an l_{g-1} labels, is used to calculate the bandwidth of the extension links. For example, a request for B_g labels in l_g layer is translated into a request for $int(B_g/U_g)$ labels in l_{g-1} layer. If an optical channel carrying four STM64 labels is mapped into one wavelength, then $U_{TDM,LSC} = 4$ and a request for an LSP with a bandwidth equal to five STM64 maps into a request for two wavelengths.

The infrastructure service model defines two complementary RToE methods: the layer-by-layer method (LBL) and the combined method (CMB). The layer-by-layer method considerers each layer separately when routing a traffic LSP request, and performing a ToE action, i.e., routing extension LSPs. It combines a single-hop traffic maximizing approach with multi-hop routing: an $LSP_g(v_o, v_d, B_g)$ in layer l_g is first routed across all available single-, or multi-hop paths in this layer; if this attempt fails a new *direct virtual link* will be added (if possible) between v_o and v_d in a next server layer l_{g-1}. Further important aspect of the LBL method in a two-layer network is that it uses only TDM grooming resources visible at the CE nodes. On the other hand, the combined RToE method uses all available (visible) grooming resources, including CE, P and PE nodes, and establishes an integrated view of multi-layer resources, including virtual and transit links, as previously introduced. Hence, a path for a traffic LSP computed in such combined routing graph includes sub-paths for all necessary extension LSPs. The CMB RToE aims at creating short virtual links and is in this respect delay minimizing virtual topology approach. Complemented with the ToE actions which release low utilized virtual links, LBL and CMB RToE produce virtual topology which adapts to the traffic requirements.

4.6.1 Layer-by-Layer (LBL) RToE Algorithm

The Layer by Layer (LBL) routing and topology engineering (RToE) algorithm starts with creating a routing transformation of SVG, i.e., a graph W_g in the traffic LSP layer l_g. This is done within the procedure $TransfLBL_g(G^s)$.

A routing graph W_g replicates the vertices and intra-layer edges of the SVG subgraph G^s_g, together with the inter-layer edges between G^s_g and G^s_{g-1}. Included are only those edges that have sufficient bandwidth to accommodate the requested LSP. The inter-layer edges included in W_g only indicate that a ToE action in a server layer t_{g-1} can be started when necessary. The procedure $TransfLBL_g(G^s)$ has the complexity equal to that of the graph search and is comparable to the complexity of the procedure *CreateLayer*-$G^s_g(V^s_g, E^s_g)$.

LBL RToE proceeds with the invocation of the procedure $RouteLBL_g()$, which searches for the shortest path for an $LSP(l_g, v^{sc}_{g,o}, v^{sc}_{g,d}, B_g)$ in W_g. The result of this search may either be a path for LSP accommodation, or a blocking event in l_g. As a reaction to the blocking event in l_g, a ToE action for topology extension in the next granularity layer $(l_h, h = g - 1)$ is invoked.

Within the recursive procedure $RouteLBL()$ formally described below, the layer of the requested *traffic* LSP is captured in r (Step–1). If LSP routing in any layer l_h fails, a ToE request for an *extension* LSP in layer l_{h-1} is generated between the source and the destination node of the original LSP. In general, the multiplexing factor U_h is used to calculate the bandwidth for an extension LSP in l_{h-1}, i.e., $B_{h-1} = int(B_h/U_h)$.

$\underline{RouteLBL(g, r, w_{g,o}, w_{g,d}, B_g)}$

Step-1: if (r = "") {set $r = g$; set $h = g$;}

Step-2: find the shortest path $P_h(e(w_{h,o}, w_{h,z^2}), ..., e(w_{h,z^H}, w_{h,d}))$

if P_h exists {$Update_h(P_h, B_h)$; goto STEP-6} else {goto STEP-3}

Step-3: set $h = h - 1$;

if ($h = 0$) {return "*blocked*"};

if ($e^w_{h+1,h,o}$, $e^w_{h,h+1,d} \in W_g$) {goto STEP-4} else {return "*blocked*"};

Step-4: $TransfLBL_h(G^s)$; set B_h=$int(B_{h+1}/U_{h+1})$;

Step-5: if($RouteLBL_h(stat, w_{h,o}, w_{h,d}, B_h)$="success") {goto STEP-6} else {STEP-3};

Step-6: if ($h \neq r$) {goto STEP-7} else {return "*sucess*"}

Step-7: set $h = h + 1$

insert $e(w_{h,o}, w_{h,d})$ in W_h; set $\mathrm{A}(e(w_{h,o}, w_{h,d})) = (H - 1)$;

set $\vartheta(e(w_{h,o}, w_{h,d})) = \cap_{(e \in P_h)} \vartheta(e)$;

$InsBund_h(v^s_{h,o}, v^s_{h,d}, B_h, A(e(w_{h,o}, w_{h,d})), \vartheta(e(w_{h,o}, w_{h,d})))$

Step-8: goto STEP-5

The extension action starts in Step–4, where a graph W_{h-1} is created with $TransLBL_{h-1}()$. In Step–5 $RouteLBL_{h-1}()$ is called, where $r \neq$ "" indicates routing of an extension LSP. In the case of a route search failure, the next server layer is entered, or if such does not exist, the request is blocked (Step–3). In the case of successful routing, the procedure makes a distinction between the layer in which the original traffic LSP should be routed (l_r), and all other server layers in which extension LSPs are routed, $l_h \neq l_r$, (Step–6). After successfully routing in layer $l_h \neq l_r$ the transformed graph of the next client layer W_{h+1} is updated with a new edge representing a dynamically created bandwidth (Step–7). This edge is assigned a routing weight, e.g., equal to $(H - 1)$, where H is the number of hops of the underlying LSP. The visibility of a new link is calculated as the common subset of visibility sets at all edges in P_h^s. Next is an update of an SVG at the layer G_h^s. Procedure $InsBund_h()$ either creates a new edge (FA) or bundles a new capacity to an existing edge. Eventually, the initial request for a traffic LSP is finally routed in l_r (Step–6), which completes the RToE. The procedure can iterate through all layers visible within the service, so the maximum number of iterations is $|L^s|$.

The complexity of this RToE procedure is equal to the complexity of the shortest path search, which may be executed maximally $2|L^s|$ times, hence the complexity is bounded with $O(|D^s||C| + |X^s||D^s|^2 + |D^s|N^2))$, if the Dijkstra shortest path algorithm is used [Dijk59]. Procedures $TransfLBL_h(G^s)$, and $Update_g(P_g^s, B_g)$ are executed $|L^s|$ times at most. Thus the overall complexity is bounded by $O(|L^s|(H_m U_m \log U_m + |D^s|(|C| + N) + |X^s||D^s|^2 + |D^s|N^2))$, where N is the max number of nodes in all domains, $|D^s|$ is the number of service domains, $|C|$ is maximal number of service members, $|L^s|$ number of granularity layers visible within service, U_m maximal number of component links in one aggregated FA, H_m is the maximal number of path hops, and $|X^s|$ number of GRXs supporting the service.

4.6.2 Combined (CMB) RToE Algorithm

The combined routing and topology engineering procedure uses the integrated view of resources pertaining to all the layers that exist in the network. The graph transformation procedure $TransfCMB_g(G^s, B_g)$ for CMB RToE approach is more complex than $TransfLBL_g(G^s)$ for LBL RToE as it has to create one single graph with resources visible at all layers. This procedure is described as follows.

$\underline{TransfCMB_g(G^s, B_g)}$

Step-1: for each $v^s_{g,i} \in V^s_g$, create $w_{g,i}(v^s_{g,i})$ in W_g

for each $e^s_{g,i,j}(v^s_{g,i}, v^s_{g,j}) \in E^s_g$

{ if $(B(e^s_{g,i,j}) > B_g)$ insert $e^w_{g,i,j}(w_{g,i}, w_{g,j})$ in W_g; set $A(e^w_{g,i,j}) = A(e^s_{g,i,j})$ }

Step-2: for each $l_h \in L^s$, start with l_{g-1} end with l_1 do Step-3 - Step-4

Step-3: set B_h=$int(B_{h+1}/U_{h+1})$

Step-4: for $e^s_{h,i,j}(v^s_{h,i}, v^s_{h,j}) \in E^s_h$ for which $B(e^s_{h,i,j}) \times U_h \times ... \times U_g > B_g$

{ if $(w_{g,i} \in W_g$ and $w_{g,j} \notin W_g$ and $B(e^s_{h+1,h,i}) > B_{h+1})$ {insert $w_{g,j}(v^s_{h,j})$ in W_g}

if $(w_{g,i} \notin W_g$ and $w_{g,j} \in W_g$ and $B(e^s_{h,h+1,j}) > B_{h+1})$ {insert $w_{g,i}(v^s_{h,j})$ in W_g}

if $(w_{g,i} \notin W_g$ and $w_{g,j} \notin W_g)$ {insert $w_{g,i}(v^s_{h,j})$, $w_{g,j}(v^s_{h,j})$ in W_g}

if $(w_{g,i} \in W_g$ and $w_{g,j} \in W_g$ and $e^w_{g,i,j}(w_{g,i}, w_{g,j}) \notin W_g)$

{ insert $e^w_{g,i,j}(w_{g,i}, w_{g,j})$ in W_g; set $type(e^w_{g,i,j})$ ="$transit_h$; set $A(e^w_{g,i,j}) = 1$} }

In Step–1, akin to transformation for LBL RToE, the vertices and edges from SVG G^s_r that have sufficient bandwidth are replicated in W_g. The algorithm further iterates through each layer l_h starting with the layer $l_h = l_{g-1}$ and ending with l_h=l_1. In Step–4, for each edge $e_{h,i,j}$ in l_h, which has enough capacity to route the original LSP request, the algorithm check the type of the edge. If one of the edge's end-nodes are already in W_g, but not both, the inter-layer edge at the already included node must be checked for sufficient bandwidth to support inter-layer routing of the original LSP. If this is true, first the other end-node is included in the graph and then the edge between the nodes. If none of the end-nodes is already included this is a pure transit edge. In this case both end-nodes, and the edge are included in the graph. In all cases the newly inserted edge $e_{h,i,j}$ is marked as $transit_h$. The weight of a transit edge is set to a default value 1. Note that the sufficient bandwidth is calculated based on the multiplexing factors at different layers.

The complexity of this procedure is bounded by the complexity of the graph search, i.e., O$(|L^s|(|D^s|(|C| + N) + |X^s||D^s|^2 + |D^s|N^2))$, where N is the max number of nodes in all domains, $|D^s|$ is the number of service domains, $|C|$ is maximal number of service customers, and $|L^s|$ number of granularity layers visible within a service.

The transformed graph W_g is next used by $RouteCMB(g,\ w_{g,s},\ w_{g,d},\ B_g, B_{g-1}, ..., B_1)$, which is the combined routing procedure, where $w_{g,s}$ and $w_{g,d}$ are the source and destination of the request, and B_g is the LSP's requested bandwidth. The values of $B_{g-1}, ..., B_1$ are calculated in $TransfCMB_g(G^s, B_g)$.

$\underline{RouteCMB(g, w_{g,o}, w_{g,d}, B_g, B_{g-1}, ..., B_1)}$

Step-1: find the shortest path $P_g(e(w_{g,o}, w_{g,z^2}), ..., e(w_{g,z^H}, w_{g,d}))$

if (P_g not exist) { return "*blocked*" }

Step-2: for each $l_h \in L^s$, start with l_1 end with l_{g-1}

{ find a set of the shortest connected transit paths $Q_{r,h}$, r=1,2,...,k in P_g,

for each $Q_{r,h}(w_{g,a}, w_{g,b})$ in P_g

{ substitute $Q_{r,h}$ by $e(w_{g,a}, w_{g,b})$ in P_g; $Update_h(G^s_h, Q^s_{r,h}, B_h)$

set *type*$(e(w_{g,a}, w_{g,b}))$="$transit_{h+1}$"; set $\vartheta(e(w_{g,a}, w_{g,b})) = \cap_{e \in Q^s_{r,h}} \vartheta(e)$;

$InsBand_{h+1}(v^s_{h+1,a}, v^s_{h+1,b}, B_{h+1}, H^r - 1, \vartheta(w_{g,a}, w_{g,b}))$ } }

Step-3: $Update_g(G^s_g, P_g, B_g)$; return "success"

In Step–1, if the shortest path P_g is not found, the request must be blocked, since the available resources at all layers do not form a connected graph. If a path is found, SVG and the transformed graph are updated in each layer l_h (Step-2) starting from l_1, i.e., the physical layer, and ending with l_{g-1}. The path P_g is searched for concatenations of transit paths at different layers (Step-3), and for each transit path $Q_{r,h}$ in P_g, which comprises the transit edges of type "$transit_h$", first, this transit paths is substituted in P_g by a single edge between the ingress and egress nodes of $Q_{r,h}$, and the type of this single edge is set to "$transit_{h+1}$" edge. Also, the procedure $Update_h()$ is called to update the resource information in G^s_h which corresponds to resource allocation for the transit path

$Q_{r,h}$ mapped into SVG $Q^s_{r,h}$. After that, the procedure $InsBundle_{h+1}$ is invoked to insert a new dynamic component link in the service graph G^s_{h+1}. The weight of the inserted component link is set $(H^r - 1)$, where H^r is the number of hops in $Q_{r,h}$. The bandwidth is set to B_h (as calculated within $TransfCMB()$. The visibility of a new link is calculated as the common subset of visibility sets at all edges in $Q^s_{r,h}$. In a loop of this process, all concatenations of transit links at all layers are gradually substituted until the path P_g consists only of links in l_g layer. In Step-3, the procedure $Update_g()$ updates the resource information in G^s_g which corresponds to resource allocation for the requested traffic LSP.

The problem of finding a set of concatenated transit paths in the path P_g can have a number of solutions. In our approach a set with the shortest transit paths is selected. In this way the available grooming capability in the network is used until exhausted, in which case flow-through links will be created.

The complexity of the routing procedure is determined by the complexity of the shortest path routing, as well as by the search for concatenations of transit edges ($O(H_m \log H_m)$) and the update procedure ($O(U_m \log U_m)$). Thus the overall complexity is bounded by $O(U_m \log U_m + H_m \log H_m + |D^s|(|C| + N) + |X^s||D^s|^2 + |D^s|N^2))$, where N is the max number of nodes in all domains, $|D^s|$ is the number of service domains, $|C|$ is maximal number of service members, $|L^s|$ number of granularity layers visible within service, H_m is the maximal number of path hops (in one layer), and $|X^s|$ number of GRXs used. Optimization and reduction of complexity may include the consideration of shortest path algorithms faster than Dijkstra, and a reduction of the number considered layers and domains.

4.6.3 Infrastructure Services in 2-layers Hybrid Optical Networks

We illustrate the infrastructure service provisioning in a two-layer network with lambda switching capability (LSC), and TDM grooming and switching capability, both at the customer equipment and in the provider domain, i.e., L=$\{l^1 = LSC, l^2 = TDM\}$. This two-layer configuration is used in our performance study as well. For simplicity and without loss of generality, we consider a single service S_1, provisioned within one provider domain, with LSP traffic requests of TDM granularity. In this example, we compare two scenarios which differ in visibility of the TDM layer grooming resources: the *reduced visibility* and *full visibility* scenario. In the reduced visibility scenario, the TDM-layer grooming and switching resources in the nodes of the provider domain are not visible, although existing, and cannot be used. In the *full visibility* scenario all provider domain nodes support TDM layer switching and grooming, and the corresponding resources are made visible to the service. In both cases, we assume a full availability and visibility of LSC resources.

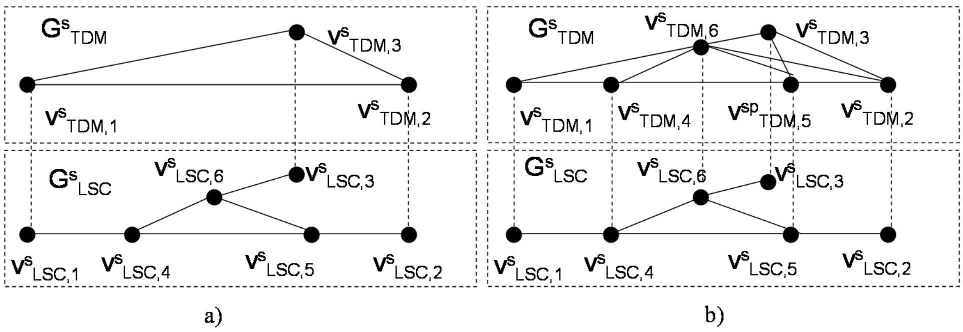


Figure 4.10: *SVG for (a) Reduced and (b) Full Grooming Visibility*

Figure 4.10 shows an SVG created for a simple network topology, for the reduced and full grooming visibility. The network topology includes three service members ($v^s_{LSC,1}, v^s_{LSC,2}, v^s_{LSC,3}$) and three LSRs in the provider domain ($v^s_{LSC,4}, v^s_{LSC,5}, v^s_{LSC,6}$). SVG is composed of two subgraphs G^s_{TDM} and G^s_{LSC} which are interconnected with unidirectional inter-layer edges, i.e., $e^{sm}_{TDM,LSC}$ and $e^{sm}_{LSC,TDM}$, depicted as dashed. For both visibility cases G^s_{LSC} includes vertices which represent the LSC switching capability at the customer and provider sites.

In the reduced visibility case, G^s_{TDM} includes only vertices representing TDM switching capability at the customer sites G^s_{TDM}, connected with three dynamic links $e^s_{TDM,1,2}$, $e^s_{TDM,3,2}$ and $e^s_{TDM,1,3}$. The edge $e^s_{TDM,1,2}$ may be for example an aggregated link composed of one component link created over LSC-LSP ($v^s_{LSC,1}, v^s_{LSC,4}, v^s_{LSC,5}, v^s_{LSC,2}$), and the other created over LSC-LSP ($v^s_{LSC,1}, v^s_{LSC,4}, v^s_{LSC,6}, v^s_{LSC,5}, v^s_{LSC,2}$).

In the full visibility scenario, TDM switching capable provider nodes are visible in G^s_{TDM} and therefore also included in G^s_{TDM}. The set E^s_{TDM} may include up to 30, i.e., 6×(6-1), dynamically created link. Figure 4.10, E^s_{TDM} includes 10 of them.

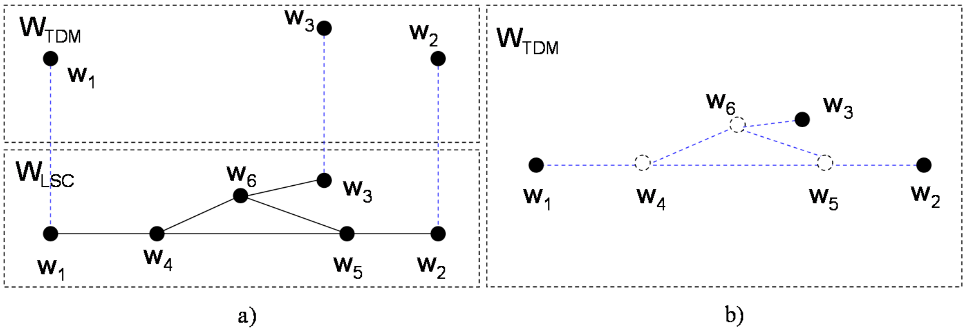


Figure 4.11: *SVG Transformation for (a) LBL RToE, and (c) CMB RToE.*

Figure 4.11 illustrates the difference between the transformed graphs for layer-by-layer (LBL) and combined (CMB) RToE approaches, under the assumption of the reduced grooming visibility, and just after the first TDM LSP is requested. The transformed graph W_{TDM} created with the LBL method, includes only the inter-layer edges (dashed), which indicate a possibility of graph connectivity extension in the LSC layer. LBL attempt to route an LSP between $w^s_{TDM,1}$ and $w^s_{TDM,2}$ results in blocking in W_{TDM}, and a topology engineering action for topology extension starts: first W_{LSC} is created and then an LSC LSP is routed in this graph. A dynamic link corresponding to a successfully routed LSC LSP is inserted between $v^s_{TDM,1}$ and $v^s_{TDM,2}$ in W_g, and is used to route the original TDM LSP traffic request. On the other hand, the routing graph W_{TDM} which is created with CMB method, (b), includes the LSC layer edges (from G^s_{LSC}) as "transit" edges, shown as dashed. These edges connect the otherwise disconnected TDM topology. The CMB routing finds a path over the transit links which are then used to set up the corresponding LSC-LSP and a dynamic link, which is then used for a TDM LSP request.

We illustrate in Figure 4.12, how the resource visibility attribute controls sharing between two infrastructure services, e.g., S_1 and S_2. Let us assume here that only the customer edge resources are TDM-switching capable. All provider switching resources, including the inter-layer resources are visible to both services S_1 and S_2. These services can use

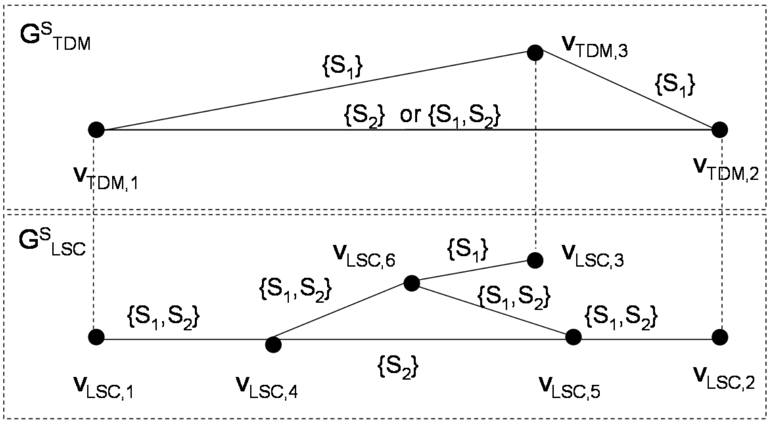


Figure 4.12: *Sharing with Resource Visibility Attribute*

network links in the LSC layer (G^s_{LSC}) according to the visibility attribute assigned to each link. For example, a link between $w_{LSC,4}$ and $w_{LSC,5}$ is exclusively visible to S_2. The edges in G^s_{TDM} denote the dynamic links. If in S_2 a dynamic link is created over $\mathrm{P}(e^s_{LSC,1,4}, e^s_{LSC,4,5}, e^s_{LSC,5,2})$, then its visibility attribute includes $\{S_2\}$. If in S_2 a dynamic link is created over $\mathrm{P}(e^s_{LSC,1,4}, e^s_{LSC,4,6}, e^s_{LSC,6,5}, e^s_{LSC,5,2})$ its visibility will include $\{S_1, S_2\}$. This means that this dynamic link will be visible to both services and the information about a new link may be exchanged between the control plane instances of the services. It should be noted that this is one possible policy for configuring visible resources. Another policies, e.g., when some logical resources are made visible to other services despite of the physical layer visibility constraint may be desirable, an subject to further studies.

4.6.4 A Case for Dynamic Visibility

The infrastructure service model introduces the visibility attribute as a means to flexibly associate resources with the services that can use them. For example in a case of a two-layer optical network, physical resources that can be flexibly configured as visible to different services are WDM interfaces, optical switching resources, and switching and multiplexing resources at TDM layer. In SVG, these resources are represented as edges and vertices pertaining to the TDM and LSC sub-graphs (G^s_{TDM} and G^s_{LSC}), as well as the inter-layer edges modeling the multiplexing and layer adaptation capabilities. Virtual links, that are dynamically added to, or deleted from SVG inherit their visibility from the physical resources used to create them. How the visible resources are used, depend on the RToE and ToE methods, as well as the configuration of the control process, which, as already discussed, may further constrain the information available in the service CPI, creating the need for tighter coordination between service CPIs and provider CPIs.

From a perspective of a single service, the resource visibility can be static or dynamic information. In the static case a set of visible resources from the common infrastructure will not change during the service life-time. In the dynamic visibility case, however, it is possible to use visibility as a means to dynamically change the access to resources. For example, let us assume that in time T_0, the grooming visibility is reduced, meaning as previously introduced, that only LSC layer resources are visible. The service can only set up end-to-end lightpaths - i.e., LSC LSP - between the CE nodes. These lightpaths are used to route service traffic flows, i.e., TDM LSPs. Each lightpath is represented as a virtual edge in SVG and its utilization can be monitored, either within the service CPI, or both within the service CPI, and the provider CPI. Assuming appropriate level of information exchange between different CPIs, we can assume that a monitoring process can detect how many flow-through lightpaths are routed through each hybrid node, and in addition the utilization of the corresponding virtual links at the TDM layer can be measured. The node with the highest number of flow through links, or the lowest mean utilization of the flow-through links, can be chosen to change its visibility. Low utilized virtual links will eventually get released, and shorter lightpaths will be created using new visibility. In a scenario with multiple services the flow-through node can be opened for one set of services, and another for another set. This mechanism can be also seen as "resource optimizer" of the unused links in the carrier domain, which may be even relatively non-transparent for the service user. In our simulation study we tackle this scenario by measuring the number of flow-through LSPs at each transit node, and the utilization of these LSPs in a network with the reduced visibility, to determine potential candidates for visibility adaptation. We have also illustrated how the visibility adaptation based on this measurement can improve service performance.

4.7 Conclusion

In this Chapter we described the proposed Infrastructure Service model, including the visibility concept and resource sharing approach, and presented the proposed methods for traffic routing and topology engineering. The revisited theoretical work relevant for the infrastructure service RToE, albeit brief, shows the complexity of the virtual topology design problem, which motivated many heuristic approaches. We described in detail virtual topology design heuristics that inspired our Routing and Topology Engineering (RToE) methods. The demonstrated performance of some heuristics for virtual topology creation provide important guidance on how the dynamic topology can be created.

Central to the infrastructure service model is the resource visibility concept, which facilitates selective usage and sharing of resources. The visibility concept enables resource providers to dynamically select physical resources that can be dynamically allocated and time-shared by different services. The bandwidth of the dynamically configured virtual links can also be jointly used by different services. We described how visibility is accounted for in the infrastructure service provisioning concept based on multi-layer Service Visibility Graphs (SVG), its routing transformation and continuous updates. We show how the proposed service model combines Routing and Topology Engineering (RToE) methods, and the stand-alone bandwidth release ToE strategies, in an approach, in which each new component link added to a service topology becomes directly associated with a specific traffic LSP, which was a direct cause of its creation. Two complementary RToE methods are proposed and discussed. We have also identified the potential benefits of the dynamic resource visibility, in a scenario when changing visibility of grooming resources, based on the measurements of virtual link utilization, can improve the service performance.

Some aspects related to the infrastructure service provisioning, such as restorability or survivability of the service, are not directly addressed here. However, we assume that the specific service restorability can be processed within the service admission control and mapped into a set of requests for traffic LSPs. For example, in case of the 1+1 protection two path-diverse LSPs have to be computed and provisioned, as a working and protecting path. In this case the notion of LSP path diversity has to be included in the traffic routing and extension algorithms. We also assume that a virtual concatenation is a available and can be used to include protection methods based on traffic spreading. Specific service restorability approach is a possible topic for future work.

5 Performance Study

5.1 Introduction

This Chapter presents the results of a simulation study conducted to quantify the impact that the visibility concept and the GXP-based federation architecture have on the performance of infrastructure service provisioning. The Chapter is organized as follows. Section 5.2 describes general assumptions used in the simulation study, as well as the implementation constraints and capabilities of the simulation environment. In this section, we also define the configuration of network nodes and the traffic load model. The performance of the proposed traffic routing and topology engineering (RToE) methods is presented in Sections 5.3 and 5.4 in single- and multi-domain scenarios, respectively. Section 5.5 summarizes and concludes this chapter.

5.2 General Assumptions

The simulation study was conducted in an object-oriented, event-based simulator developed at the Institute of Broadband Communications, which we extended with functions relative to network resource management and infrastructure service provisioning. Input to the simulation tool was the textual description of the physical topology of the network, including the CE, PE, P, B, and GRX nodes and links, along with the description of infrastructure services, including the required service connectivity and traffic requirements. In addition, for each service, this file defines the selected ToE link-release method and the RToE method. Based on this description, the simulation tool creates the network graph, the corresponding service visibility graphs, and the traffic sources that generate the unidirectional customers' traffic originating from different services. During a simulation run, each traffic source, which is associated with one service, generates the LSP requests, which are processed by the RToE method selected for this particular service.

Regarding equipment capabilities, we simulate the infrastructure service provisioning in a two-layer optical network of label-switched routers (LSR), with lambda switching (LSC) and TDM switching and grooming capabilities. Customer equipment, the CE nodes, can multiplex a variable number of TDM containers - i.e., labels - of one basic granularity into wavelengths and also switch in both TDM and LSC layers. A customer can initiate the setup of the TDM granular LSPs and wavelength switched paths (LSC LSPs). While TDM granular LSPs are the traffic LSPs, LSC LSPs are set up to extend the service's virtual topology and are referred to as extension LSPs. Provider domain nodes (PE, P, B, or GRX) can also switch in the LSC and TDM layer. By means of the visibility attribute, for each hybrid LSR, a set of services can be configured that can use its TDM switching and grooming capability. The architectures of the CE and PE nodes are illustrated in Figure 5.1.

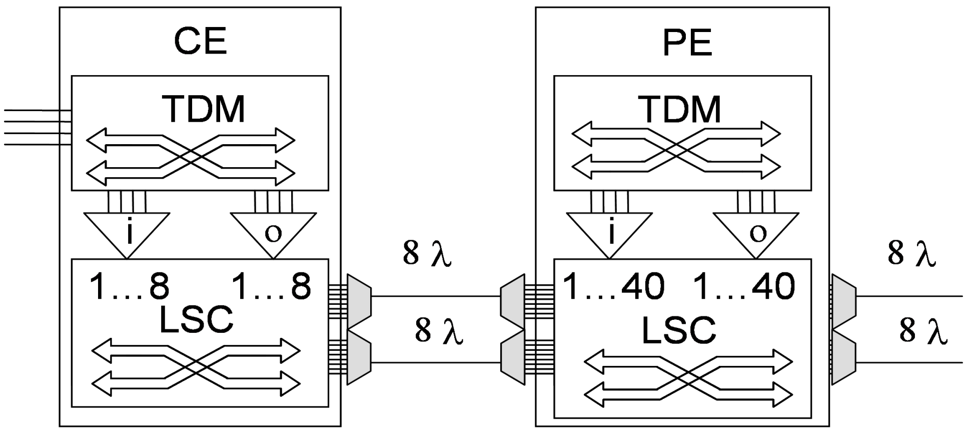


Figure 5.1: *Network Node Architecture for CE and PE nodes*

The textual node description includes, therefore, supported switching capability, the number of transceivers and switching ports, and the visibility for different services. In a multi-domain network, each node is also assigned to one specific domain. The link description includes the number of wavelengths and the multiplexing factor, i.e., the number of TDM labels within one wavelength.

The connectivity and bandwidth requirements for each infrastructure service are given as a set of tupples (c_s, c_d, $B_{s,d}$), where c_s and c_d are CE nodes identifiers, and $B_{s,d}$ is a requested number of TDM labels. For each (c_s, c_d) pair, a traffic source will generate LSP requests with the inter-arrival rate and holding times following a negative exponential distribution. Each service S_i is associated with a specific mean inter-arrival rate λ_i, and a mean holding time t_i^h, and all traffic sources associated with the service generate requests for LSP setup and release, according to these distributions. The mean load generated by one source is $\lambda_i \times t_i^h$, and the total service load is given by $\sum_{(c_s,c_d) in S_i} \lambda_i t_i^h$.

During a simulation run, each successfully created TDM LSP, i.e., a traffic LSP, is inserted into a linked list and is deleted from it after the release request. Each successfully created LSC LSP, i.e., an extension LSP, is inserted into a service SVG as a component link and is deleted from it as a result of a ToE link release action. Active traffic and extension LSPs are linked together such that, when a traffic LSP is released, all the resources that it was using are released as well. Extension LSPs that should be removed when idle are checked for idleness after each traffic LSP release request.

Each simulation run includes the transient phase of 60 T_{sim} periods (during which the network is filling up) and the active phase of 30 T_{sim} periods. With an appropriate T_{sim} value, the active periods can be considered as uncorrelated experiments, and a certain required level of confidence for the results can be obtained [Jain91]. The traffic sources are started within the first 15 periods of the transient phase, i.e., their starting time is uniformly distributed within this period. Simulation time is calculated so that the service with the lowest load can attempt some specified minimum number N_{LSP} of LSPs during each active simulation period. In our simulations N_{LSP} is set to 100. At the start of the simulation, all network capacities are free, and during the transient phase the network is filling up. In the active phase, the values of relevant performance indicators are collected and averaged at the end of each active period. At the end of simulation, the performance values obtained are used to calculate the mean, the deviation, and the confidence interval.

The results are shown with the confidence level of 95%.

5.2.1 Performance Metrics and Result Representation

During each active period in a simulation run, the incremental changes of several performance counters are logged and integrated in time, and the corresponding performance metrics are calculated at the end of each period. Based on these counters, we are able to characterize the network with the following performance metrics:

- **Average WDM Link Usage ($\overline{W}_{i,j}$):** For each WDM link $e_{i,j}$ in a network graph, where $B_{i,j}$ is a total number of wavelengths, the capacity allocated within the logical topology (in wavelengths) is a step function of time $W_{i,j}(t_k)$, which changes each time new LSC LSP is added to, or released from, the virtual topology, i.e. in t_k, $k = 1, 2, ...$ The weighted average in time (Equation 5.1) is calculated at the end of each active period, hence $\sum_k t_k = T_{sim}$.

$$\overline{W}_{i,j} = \frac{\sum_i t_k W_{i,j}(t_k)}{B_{i,j} \sum_k t_k}, \quad 0 \leq \overline{W}_{i,j)} \leq 1 \tag{5.1}$$

  For different types of links, the results are summarized within histograms that show the number of links with $\overline{W}_{i,j}$ in intervals 0%, (0%-3%], (3%-5%], (5%-7%], (7%-10%], (10%-15%], (15%-20%], ...,(90%-100%). If we denote access links CE-PE), inter-domain links (B-B), intra-domain links (PE-PE), and GXP links, as M^{cp}, M^{pp}, M^{bb}, and M^{x}, respectively, then the histograms correspond to the values of $M^t_{0\%}$, ..., $M^t_{10\%}$, ... , and $M^t_{100\%}$. The superscript t shows the type of the link, $t \in \{"cp","pp","bb",x\}$, for access, intra-domain, inter-domain, and GXP links, respectively. The final result is the normalized usage, i.e., the ratio $M^t_{k\%}/M^t$.

- **Average Virtual Link Usage ($\overline{U}_{o,d}$)**: For a virtual link $e_{o,d}$, a ratio of the used capacity $U_{o,d}(t)$ (in TDM labels) to the total capacity $K_{o,d}(t)$ is averaged in time by monitoring changes in both the total and used capacity, in a series of intervals

t_k, $k = 1, 2, ...$ The total capacity changes each time a new extension LSP, i.e., a component link, is bundled to, or de-bundled from, a virtual link. The used capacity changes each time a traffic LSP is routed or released. Virtual Link Usage is calculated as:

$$\overline{U}^s_{o,d} = \frac{\sum_k t_k U_{o,d}(t_k)/K_{o,d}(t_k)}{\sum_k t_k}, \quad 0 \leq \overline{U}^s_{o,d} \leq 1 \tag{5.2}$$

We summarize the results for the whole topology in a histogram that shows the number of virtual edges with $\overline{U}_{o,d}$ for values and in intervals 0%, (0%-3%], (3%-5%], (5%-7%], (7%-10%], (10%-15%], (15%-20%], ...,(90%-100%).

Service performance is expressed in terms of blocking probability of service traffic requests and extension requests. We also observe per-service bandwidth allocated for virtual links or used for traffic flow routing. The resources allocated for the creation of virtual links are the WDM layer resources, and, for routing of traffic flows, the TDM layer resources. Due to the shared visibility, services may establish virtual links that are then used by other services. For example a service S^s allocates R^s resources in the WDM layer to establish a virtual topology and out of these uses $R^{s,s}$ resources to route a traffic flow. It also uses $R^{s,z}$ resources, out of R^z established by a service S^z. The ratio $R_{s,z}/R_s$, calculated for each (s, z) pair, is an important indicator of resource sharing.

For each service S^s, we calculate the following bandwidth-related statistics:

- **Traffic bandwidth blocking probability (PB^s_t):** A fraction of the service traffic LSPs blocked, weighted by their capacity requirements. PB^s_t is given in Equation 5.3, where the capacity of the i^{th} blocked traffic flow of the service S_s is denoted as $c^s_t(i)$, and the total capacity of all flows of S^s, which routing was attempted as C^s_t.

$$PB^s_t = \frac{\sum_i c^s_t(i)}{C^s_t}, \quad 0 \leq PB^s_t \leq 1 \tag{5.3}$$

- **Extension bandwidth blocking probability (PB^s_e):** A fraction of the extension LSPs requested by an RToE method, which were blocked, weighted by the capacity requirements of LSPs. PB^{se} is given in Equation 5.4, where the capacity of the i^{th} blocked LSP request of the service S_s is denoted as $c_e^s(i)$, and the total capacity of all LSC LSPs requested by S_s is denoted as C_e^s.

$$PB_e^s = \frac{\sum_i c_e^s(i)}{C_e^s}, \quad 0 \leq PB_e^s \leq 1 \tag{5.4}$$

- **LSC layer resource usage ($\overline{R}_e^s$):** A time average of LSC layer resources allocated in the topology of service S_s, normalized to a total bandwidth of the network $C = \sum_e B(e)$, $e \in E_{LSC}$. E_{LSC} denote a set of network edges in the physical topology (LSC layer), and $B(e)$ denote the number of wavelengths. Each i^{th} unidirectional extension LSP set up by a service S^s is characterized with the resources it uses on the physical topology, i.e., $r_e^s(i) = c_e^s(i) \times h_e^s(i)$, where $c_e^s(i)$ is the number of wavelengths, and $h_e^s(i)$ is the length in physical hops. Resources used are a step function of time $R_e^s(t)$, $t = t_1, ..., t_k, ...$, which is increased by $r_e^s(i)$ when an LSP is set up and decreased by the same value when it is released. The time average is calculated as follows:
$$\overline{R}_e^s = \frac{\sum_k R_e^s(t_k)t_k}{C\sum_k t_k}, \quad 0 \leq \overline{R}_e^s \leq 1 \tag{5.5}$$

- **TDM layer resource usage $\overline{Q}^s$, and $\overline{Q}^{s,z}$, for each $S^z \neq S^s$** A time average of the bandwidth of all traffic flows (TDM LSPs) of service S^s, routed over virtual resources established by S^s or by any other service S^z. Resources are expressed in terms of TDM containers, as a function of time. Resources of an i^{th} uni-directional traffic LSP, characterized with the capacity $c_t^s(i)$ and the length (in virtual hops) $h_t^s(i)$, routed over virtual links established by service S^s, are denoted as $p^{s,s}$, and over virtual links established by any other service S_z, $p^{s,z}$, so that it holds $\sum_{z=1}^{z=|S|} p^{s,z} = c_t^s(i)h_t^s(i)$. For each service the total resource usage is a function of time $P^{s,s}(t)$ and $P^{s,z}(t)$ and is increased or decreased by $p^{s,s}$, $p^{s,z}$ for each activated or terminated traffic flow. The total bandwidth used by service S^s flows is then given by $P^s = \sum_{z=1}^{z=|S|} P^{s,z}$. We calculate the resource usage values averaged in time and normalized to the total used bandwidth as:

$$\overline{Q}^{s,z} = \frac{\sum_k P^{s,z}(t_k)/P^s(t_k)t_k}{\sum_k t_k}, \quad z = 1, 2, ...|S|, \quad 0 \leq \overline{Q}^{s,z} \leq 1 \tag{5.6}$$

- **Average Number of Active Component Links ($\overline{Z}_v^s$)**: A time average obtained by monitoring step-wise changes in number of active extension LSPs (component links) in intervals t_k, $k = 1, 2, ...$ It is calculated as:

$$\overline{Z}_v^s = \frac{\sum_k t_k Z_v^s(t_k)}{\sum_k t_k} \tag{5.7}$$

- **Average Virtual Link Capacity** ($\overline{V}_{o,d}$)**:** For each virtual link $e_{o,d}^s$, which is established by a service S^s and is represented as an edge in SVG graph (G_{TDM}^s), the total capacity of the link is a function of time $K_{o,d}(t)$, as it changes each time a new bandwidth (a new extension LSC LSP between v_o, v_d) is added to a link. Averaged in time and normalized to K^{max}, the capacity of a virtual link with the maximal capacity is given by:

$$\overline{V}_{o,d}^s = \frac{\sum_k t_k K_{o,d}^s(t_k)}{K^{max} \sum_k t_k}, \quad 0 \leq \overline{V}_{o,d}^s \leq 1 \tag{5.8}$$

For the whole service topology, the final result summarizes within histograms the number of virtual links with $\overline{V}(e_{o,d}^s)$ for values and in intervals 0%, (0%-3%], (3%-5%], (5%-7%], (7%-10%], (10%-15%], (15%-20%], ...,(90%-100%).

- **Mean Weighted Transceiver Ports Usage ($\overline{X}_t^s$)**: When a uni-directional LSC LSP is de-multiplexed at the TDM layer at the access or intermediate nodes, 1 or 2 transceiver ports are used, respectively. $\overline{X_t^s}$ is calculated in Equation 5.9, where $X_t^s(t)$, $t = t_k, k = 1, 2, ...$ is the total number of grooming ports used by S^s, and X_t is the total number of transceivers in the network.

$$\overline{X}_t^s = \frac{\sum_k t_k X_t^s(t_k)}{X_t \sum_k t_k}, \quad 0 \leq \overline{X}_t^s \leq 1 \tag{5.9}$$

- **Mean Weighted TDM Switching Ports Usage ($\overline{X}_s^s$)**: When a uni-directional TDM LSP with m multiplexed connections is switched at the TDM layer of an access or intermediate node, m or $2m$ switching ports are used, respectively. X_s^s for a service S^s is calculated averaged in time as the ratio of switching ports used by service S^s , $X_s^s(t_k)$, to the total number of switching ports X_s provisioned in the network (we assume $X_s = X_t \times U$).

$$\overline{X}_s^s = \frac{\sum_k t_k X_s^s(t_k)}{X_s \sum_k t_k}, \quad 0 \leq \overline{X}_s^s \leq 1 \tag{5.10}$$

- **Mean Service Load ($\overline{Z}_t^s$)**: A time average obtained by monitoring step-wise changes of the total number of traffic TDM LSPs $Z_t^s(t)$, $t = t_k$, $k = 1, 2, ...$ Service load is calculated as:[1]

$$\overline{Z}_t^s = \frac{\sum_k t_k Z_t^s(t_k)}{\sum_k t_k} \tag{5.11}$$

- **Mean Number and Utilization of Flow-through Links ($\overline{F}_o^{s,z}$)**: For each node v_o with a switching capability in a WDM layer, and for each service S^s, the number of flow-through LSC-LSPs of the service S^s ($K_o^s(t)$) is observed, together with utilization of LSPs. For i^{th} flow-through LSP of S^s at v_o, the total bandwidth is denoted as $B^s(i,o)(t)$, and used bandwidth is denoted as $b^s(i,o)(t)$. For each service S^s, the total bandwidth used by service S^z is given as $b_o^{s,z}(t) = \sum_i^{K_o^z(t)} b^s(i,o)(t)$, $z = 1, 2, ...|S|$ and the total LSP bandwidth as $B_o^s(t) = \sum_i^{K_o^s(t)} B^s(i,o)$. Let the node degree at the LSC layer in δ_l, which is equal to the number of unidirectional links multiplied by the number of wavelengths at each link. The time average is given by:

$$\overline{F}^{s,z}(o) = \frac{\sum_k t_k b_o^{s,z}(t_k)/B_o^z(t_k)}{\sum_k t_k}, \; z = 1, 2, ..., |S|, \; 0 \leq \overline{F}^{s,z}(o) \leq 1 \tag{5.12}$$

[1]Service load is specified as the input for each scenario, by calculating $\overline{Z}_t^s$ we assure the correctness of the traffic sources.

The weighted number of flow-through LSPs is given by:

$$\overline{K}_o^s = \frac{2\sum_k t_k K_o^s(t_k)}{\delta_l \sum_k t_k} \tag{5.13}$$

This result can be summarized for the whole topology within the histogram, which shows for each service the ratio of the number of nodes (classified into CE, PE, P, B, and GXP) that have value $\overline{K}_o^s$ within a specific interval, e.g., $\overline{K}_o^s = 0\%$, $0\% < \overline{K}_o^s \leq 10\%$, ... , and $\overline{K}_o^s \geq 90\%$, to the total number of nodes in the network.

5.3 Performance of Infrastructure Services in Single-Domain Scenarios

In this Section our attention is on the performance of traffic-driven strategies for routing and topology engineering (RToE) combined with the link release ToE strategies, in different resource visibility scenarios. In the first place, we are interested in blocking probability of service traffic LSPs and on the utilization of the transceivers. To understand the difference between the strategies, we observed also the utilization of the WDM links and virtual links. The results presented were obtained in a single-domain topology of the NSF network, depicted in Figure 5.2, which has 28 nodes (14 CE and 14 PE nodes), 24 intra-domain links, and 14 access links. The access and inter-domain links have eight wavelengths. Both customer (CE) and provider (PE) nodes have lambda and TDM switching capability. Each wavelength has four TDM labels, e.g., four ODU3 (2.5 Gb/s) mapped within one ODU2 (10 Gb/s). The customer nodes generate TDM LSP requests.

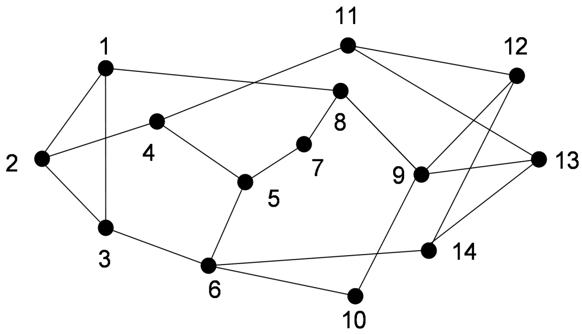


Figure 5.2: *Test Network Topology for Single-Domain Experiments.*

5.3.1 Impact of RToE strategies in a Single Service Scenario

In the first scenario, we assume that a provider domain offers visibility of both lambda switching and TDM switching capability. All CE nodes belong to one service S, and for each (c_o, c_d) pair of customers, a traffic source associated with a c_o generates traffic requests with the same mean inter-arrival and holding times t_a, t_h, and the capacity of 1 label. In this scenario, we compare the performance of the layer-by-layer (LBL) RToE method and the combined (CMB) RToE method in the configuration with the full visibility of grooming resources. With LBL RtoE, a traffic LSP between (c_o, c_d) is either routed over existing virtual links with at least one available TDM label, or if such a path does not exist, a new LSC LSP is initiated to add bandwidth to a direct virtual link between c_o and c_d. Hence, for a request $LSP(c_o, c_d)$, LBL does not use grooming capability of the domain's P/PE nodes. On the other hand, CMB RToE routes traffic LSPs on a combined two-layer graph, using all available grooming capability of CE/PE/P nodes and finding at the same time a path for a traffic LSP and the paths for extension LSPs if such should be added to the topology. We observed the combined impact of RToE and ToE link release strategies, namely, the release-when-idle (RELI) strategy the never-release (NREL)strategy, and the conditional-release (CREL) strategy. The RELI strategy releases bandwidth (in terms of component links bundled to a virtual link) when detected as unused. With the NREL strategy, each established component link is kept in the virtual topology for the whole service life time. The third strategy (CREL) is the extension of RELI in which links are released depending on their length in hops. Here, idle LSC LSPs of hop-lengths 1 and 2 are never released. The parameters of study cases are summarized in Table 5.1.

Study Case	Routing Approach	ToE Strategy
T1	LBL	RELI
T2	LBL	NREL
T3	CMB	RELI
T4	CMB	NREL
T5	CMB	CREL

Table 5.1: *Single-Domain Single-Service Scenario, Experiment 5.3.1, Study Cases.*

Figure 5.3 shows the mean blocking probability of traffic and extension LSPs in five study cases and for a number of different loads.

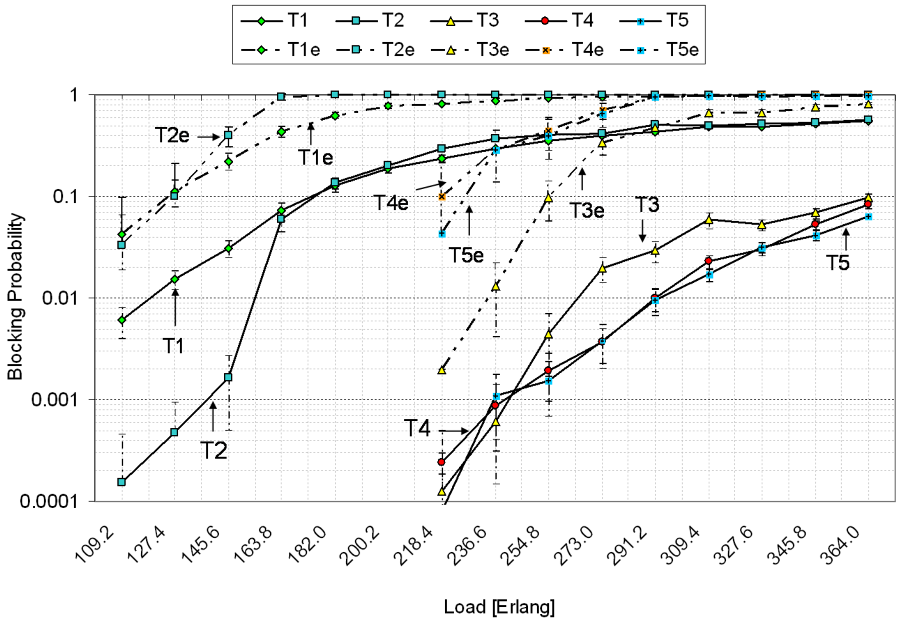


Figure 5.3: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Blocking Probability of Traffic and Extension LSPs*

The significant difference between the blocking probability experienced in T1 and T3 confirms the expected gain that is achieved by multi-hop grooming as compared to end-to-end grooming when traffic LSPs have bandwidths which are only a fraction of the total wavelength capacity, here 1/4. For lower loads and provisioning with LBL, we can further observe that the blocking probability obtained in a study case in which virtual links are not released (NREL), i.e., T2, is lower than in the one with link release (RELI), T1. This results from the fact that without releasing links, the virtual topology gradually grows as the component links (LSC LSP) are added to the virtual links between the service customers in a traffic-driven way, i.e., randomly. As the topology is gradually building up, new TDM LSPs use available single or multi-hop paths. Direct links are added until the maximum number of virtual links is established, whereupon the virtual topology becomes stable, and classic traffic engineering methods can be used to route the traffic, e.g., based on the residual bandwidth. On the other hand, when the virtual links are released when idle, the virtual topology is not stable, which affects the efficiency of multi-hop routing. The difference between the results obtained with the RELI strategy

(T1) and the NREL strategy (T2) indicates a need for topology stabilization, which can be achieved by using the results of link load monitoring to release links with the load under some variable threshold. Comparing results obtained by using all grooming capability (CMB) in combination with link-release (RELI), T3, without-release (NREL), T4, and with conditional release (CREL), T5, the conclusion is similar: a stable topology has a positive impact on the blocking probability. Particularly efficient is the CREL method, which we proposed for topology stabilization.

Concerning blocking of extension requests in scenarios T2 (LBL), T4, and T5 (CMB), in which unused bandwidth is not released, the extension success rate is lower, however the traffic blocking is also lower. However, it should be noted that the benefit of long-lived links exists only when the traffic is stable. In cases of dynamically changing traffic requirements, the release-when-idle strategy might give better performance.

Table 5.2 summarizes the normalized average number of active TDM LSPs. The values correspond to blocking probability.

Normalized Average Number of Active Traffic (TDM) LSPs					
Load	T1	T2	T3	T4	T5
109.20	1.00	1.00695	1.00778	1.00715	1.00709
127.40	1.00	1.01554	1.01591	1.01573	1.01544
145.60	1.00	1.02415	1.03400	1.03295	1.03429
163.80	1.00	1.02018	1.06615	1.06486	1.06449
182.00	1.00	1.01048	1.12376	1.12562	1.12326
200.20	1.00	0.97623	1.24977	1.25116	1.25052
218.40	1.00	0.88456	1.32594	1.32568	1.32388
236.60	1.00	0.90477	1.47471	1.47547	1.47304
254.80	1.00	0.91291	1.53108	1.53109	1.52998
273.00	1.00	0.96027	1.66360	1.68569	1.68656
291.20	1.00	0.95210	1.73007	1.78195	1.78491
309.40	1.00	0.96440	1.80778	1.86206	1.86138
327.60	1.00	0.94501	1.86733	1.93165	1.94516
345.80	1.00	0.92856	1.92770	1.97168	1.98442
364.00	1.00	0.96714	2.03402	2.09599	2.09745

Table 5.2: *Single-Domain Single-Service Scenario, Experiment 5.3.1, Normalized Average Number of Active Traffic LSPs.*

The LSC layer resource usage is given in Figure 5.4. In study cases T2, T4, and T5, in which a created LSC LSP is never released (NREL), the capacity allocated within the service virtual topology is higher than in the corresponding RELI study cases, T1 and T3, respectively. It is interesting to notice that the ratio between T3 and T4/T5 is approximately 1.7 for the load range in which T3, T4, and T5 all have comparable

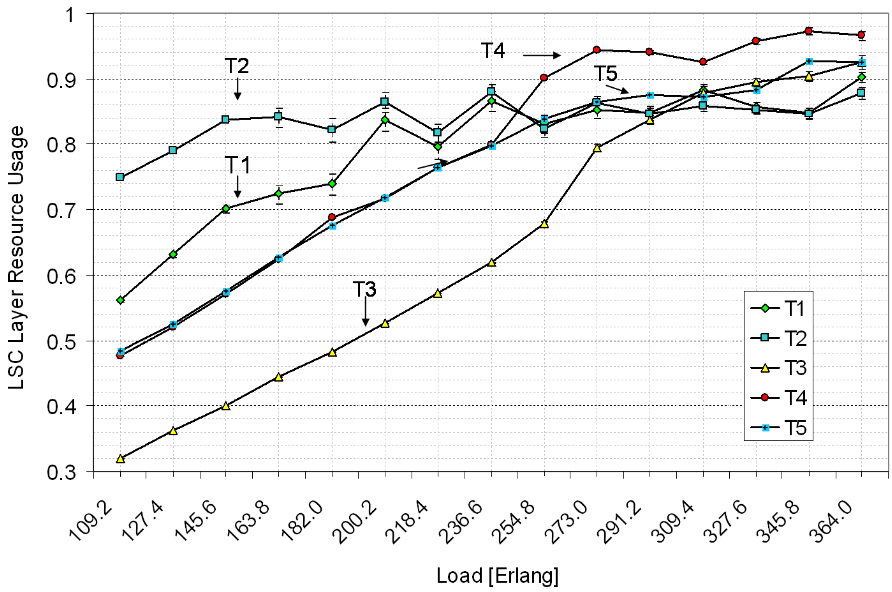


Figure 5.4: *Single-Domain, Single-Service Scenario, Experiment 5.3.1,LSC Layer Resource Usage.*

blocking probability. For the load range where blocking probabilities in T4 and T5 are still lower than 3% but higher in T3, the resource usages in T3 and T5 are approximately the same, which confirms the efficiency of the CREL release method.

This result is also confirmed in Table 5.3, which shows the average number of component links (LSC LSPs) allocated within the virtual topology, normalized to the value gained in T1. This number also corresponds to the normalized average number of transceiver ports used in different scenarios.

Normalized Average Number of Active Component Links					
Load	T1	T2	T3	T4	T5
109.20	1.00	1.30926	2.28465	3.40781	3.46184
127.40	1.00	1.24373	2.32987	3.34978	3.38321
145.60	1.00	1.16371	2.38539	3.39990	3.42925
163.80	1.00	1.08969	2.46669	3.46312	3.48302
182.00	1.00	1.05793	2.57956	3.67077	3.60516
200.20	1.00	1.02674	2.73033	3.71383	3.72456
218.40	1.00	0.96245	2.92976	3.92344	3.92356
236.60	1.00	0.98810	3.11505	4.05318	4.04768
254.80	1.00	0.97447	3.30275	4.38679	4.23181
273.00	1.00	1.00989	3.52467	4.47712	4.36295
291.20	1.00	0.98848	3.57812	4.57322	4.40046
309.40	1.00	1.00481	3.72953	4.50755	4.37858
327.60	1.00	0.99704	3.87012	4.60576	4.43730
345.80	1.00	0.97731	3.96722	4.71358	4.65389
364.00	1.00	1.00617	4.08156	4.67049	4.65354

Table 5.3: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Normalized Average Number of Active Component Links.*

The utilization of WDM intra-domain and access links provides some insight into the state of the network. Figure 5.5 shows the utilization of intra-domain WDM links for different loads in all five study cases.

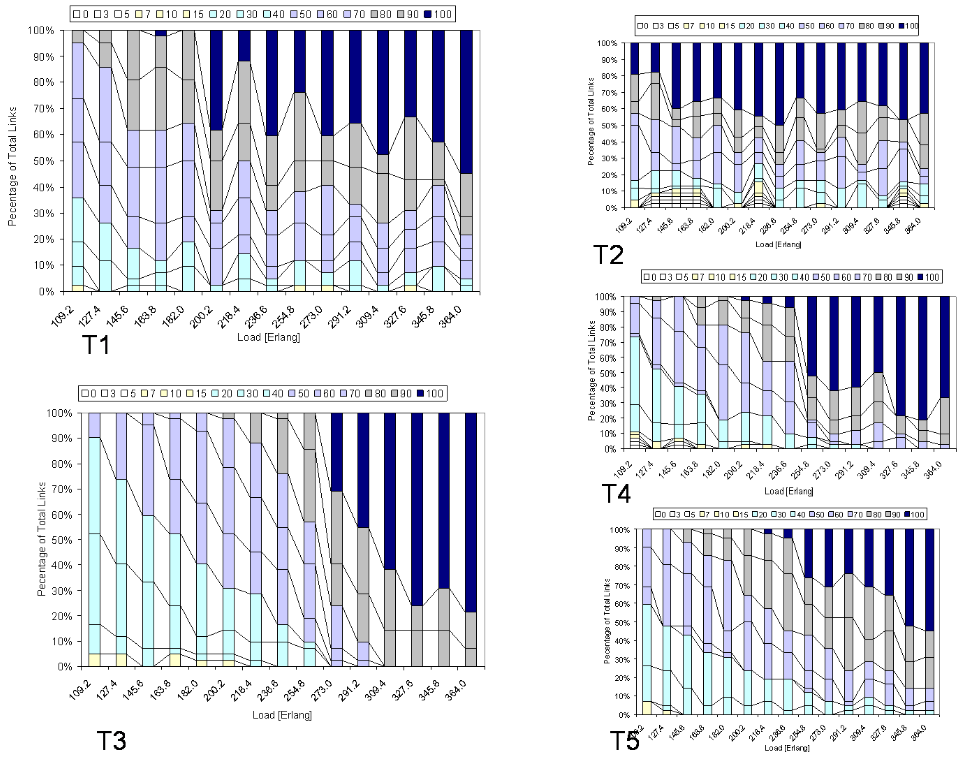


Figure 5.5: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Intra-domain WDM Link Utilization.*

In each diagram, the percentage of the total inter-domain network links, which have utilization within different intervals, e.g., equal to 0%, less then 3%, and so on, is shown. Lighter colors are used for lower loads. The effects of link release strategy (T1, T3) are obvious. It can be seen that as load increases, the average utilization of links also increases, particularly in T3. In the NREL ToE strategy (T2, T4, T5), some links may have zero utilization for higher load because they cannot be used within any component link (LSC LSP).

Figure 5.6 compares the utilization of access WDM links for the five study cases.

Due to the fact that grooming is used only at CE nodes in case studies T1 and T2, each unidirectional, multi-hop traffic LSP uses resources of both directions at each CE-PE link

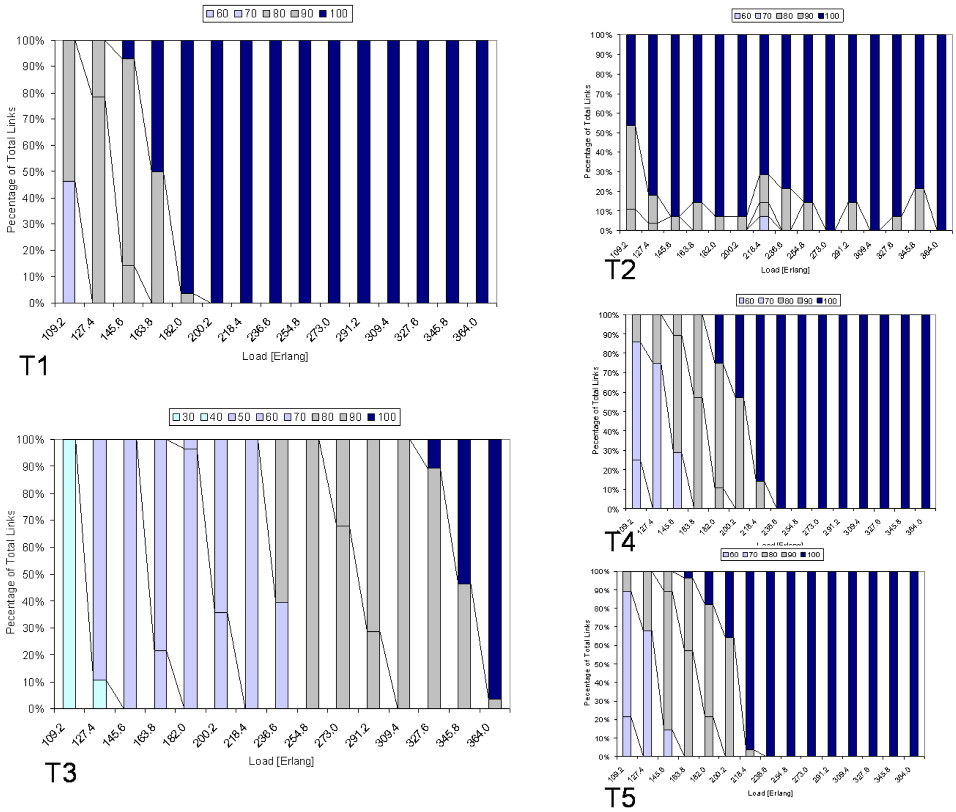


Figure 5.6: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Access WDM Link Utilization.*

of any intermediate CE node. The access link utilization is therefore much higher when LBL RToE is used (e.g., T1) than with CMB, e.g., T3, where grooming at PE nodes is used as well, and only the traffic LSPs that are directed to a particular CE node are routed over its CE-PE link. With NRELI (T4 and T5), CE-PE links are never released, and therefore their full capacity is added to the virtual topology for higher loads. For the higher traffic load, the TDM LSP blocking is mostly caused due to the blocking of access links.

How the TDM layer resources are used is shown first in Figure 5.7, where the usage ratio to the total network bandwidth at the TDM layer is presented. Test cases in which full grooming capability is used (T3, T4 and T5) use approximately the same amount of resources at the TDM layer, as compared to the 1.7 ratio observed in the LSC layer. In the reduced visibility scenario, the amount of the unused resources is much higher, due to the blocked access links, and there is higher unused residual bandwidth within the network.

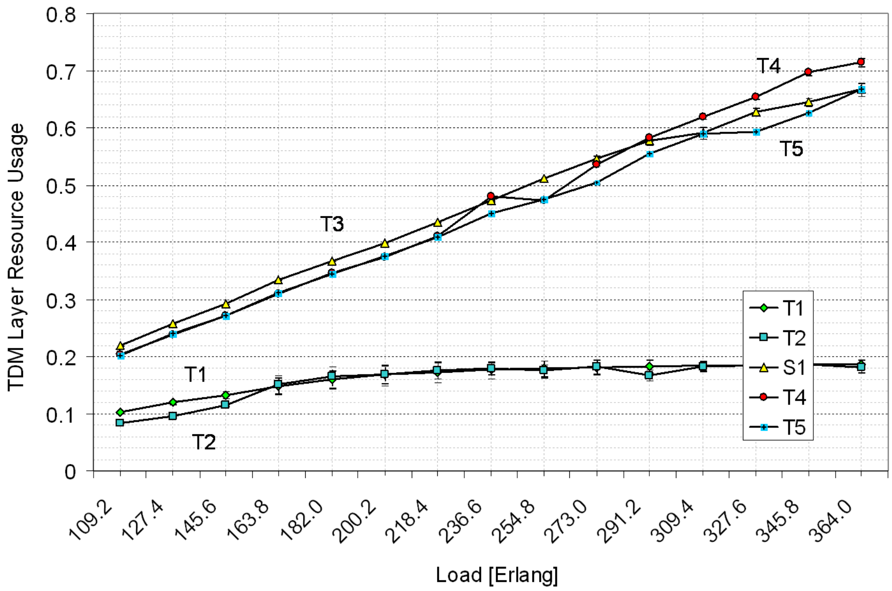


Figure 5.7: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, TDM Layer Resource Usage.*

When the full grooming capability is used (CMB), the logical topology replicates the physical topology. Therefore, high numbers of traffic LSPs are routed on the multi-hop paths over the virtual topology. The mean lengths of traffic LSPs in the five scenarios are given in Figure 5.8. It can be observed that in the case in which only CE grooming is used, the number of virtual topology hops is up to three. With full grooming capability used, the maximum hop distance corresponds to the hop distance between different CE nodes which is either 3, 4, or 5 for different (CE,CE) pairs in this topology.

We further compare the bandwidth utilization of virtual links, expressed as the ratio of the virtual link with the maximum average bandwidth to the utilization. Table 5.4 compares the maximum virtual link bandwidth (in TDM labels) for different study cases. In our configuration, four TDM labels are multiplexed within one wavelength, and, therefore, the virtual links that bundle only labels along a single path can have up to 4×8 labels. We see that in study case T4 some virtual links are bundled over different paths, resulting in maximal load over 32 labels. This is also possible in other scenarios, but it cannot be

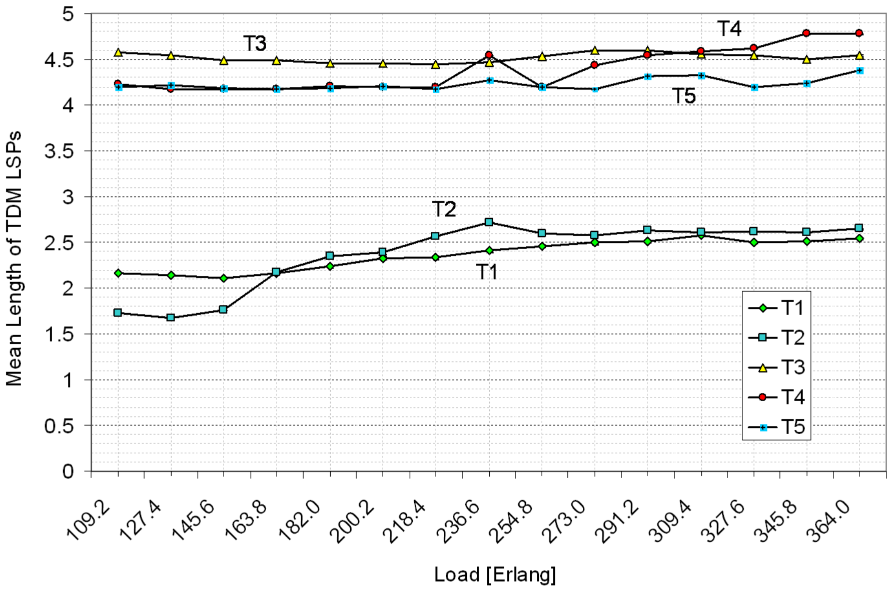


Figure 5.8: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Mean length of TDM LSP.*

deduced from the results.

Maximum Virtual Link Bandwidth					
Load	T1	T2	T3	T4	T5
109.20	5.37755	16	13.6817	24	24
127.40	5.57505	16	15.5554	28	24
145.60	5.91296	12	18.0802	28	27.1862
163.80	6.41022	12	20.2765	28	31.2083
182.00	8.7101	12	21.5612	32	32
200.20	9.99566	16	23.1256	32	32
218.40	8.57667	16	25.847	32	32
236.60	9.34431	16	26.9407	32	32
254.80	10.3834	16	27.8623	36	32.1992
273.00	11.6647	16	30.3058	36	32
291.20	15.9538	20	27.5661	36	32.0792
309.40	15.9361	16	28.199	36	32.1601
327.60	15.8883	20	31.2055	36	32
345.80	11.9848	16	30.9583	36	32
364.00	15.8957	20	33.0906	40	32

Table 5.4: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Maximum Virtual Link Bandwidth.*

In grooming networks, virtual links use WDM resources more efficiently, achieving higher maximum bandwidth. This is also demonstrated in the results shown in Figure 5.9. In case study T1, approximately 70% of virtual links has utilization under 40% of the maximal link bandwidth (3-4 wavelengths). A similar situation exists with T2 that has maximal virtual link bandwidth of 4-5 wavelengths.

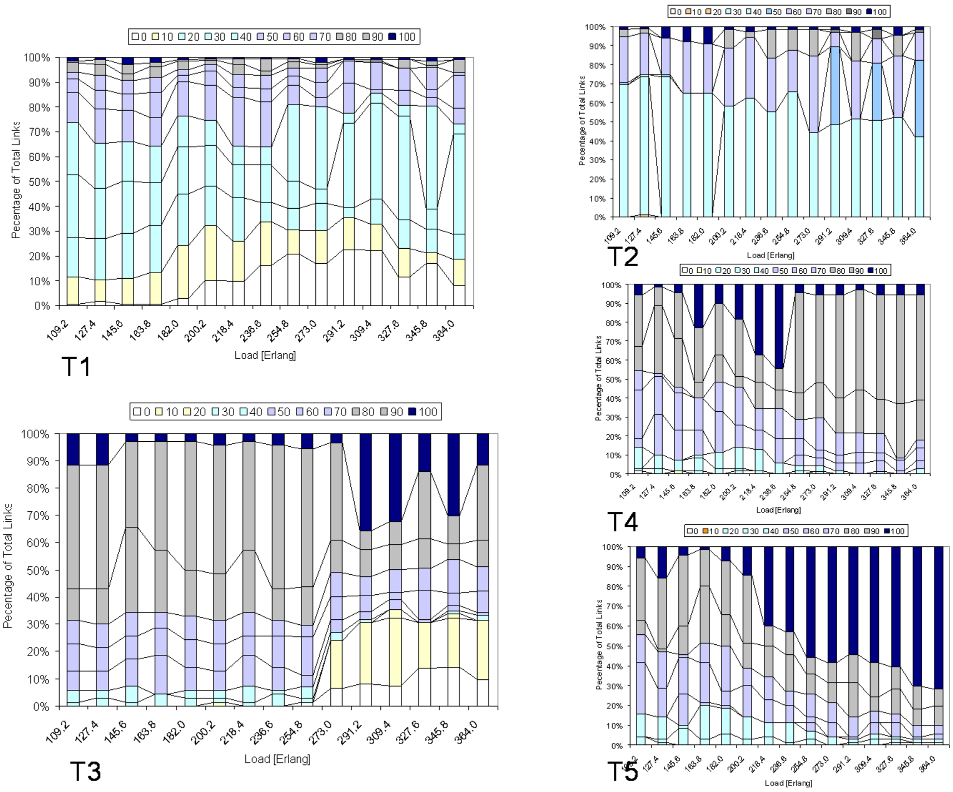


Figure 5.9: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Utilization of virtual links.*

In T3, with the higher load, some of the virtual links dominate the topology and have good utilization. They bundle more capacity, and that brings some other virtual links to "starvation." This effect can be observed for T1 as well, because in the test cases with RELI ToE, component links are de-bundled from virtual links when unused, and some 0-bandwidth virtual links can emerge. Our statistics are done based on the model in which a virtual link (edge) once included in SVG is not deleted but is monitored. Component links (extension LSPs or LSC LSPs) provide bandwidth that is either added to, or deleted from, a corresponding SVG edge (virtual link). Therefore, some virtual links can have an average total bandwidth 0. The maximum number of virtual links in SVG is equal to N^2, where N is a total number of nodes that can terminate virtual links. In study cases with the reduced grooming used (T1, T2), the number of nodes is $N = 14$, and, in those with the full grooming used (T3, T4, T5), it is $N = 28$. In T4 with the higher load, approximately 80% of links have utilizations of 70% to 90% of the highest bandwidth link.

This ratio is even better in T5, which books less bandwidth than T3. -Here, 60% of links are fully utilized maximum bandwidth links.

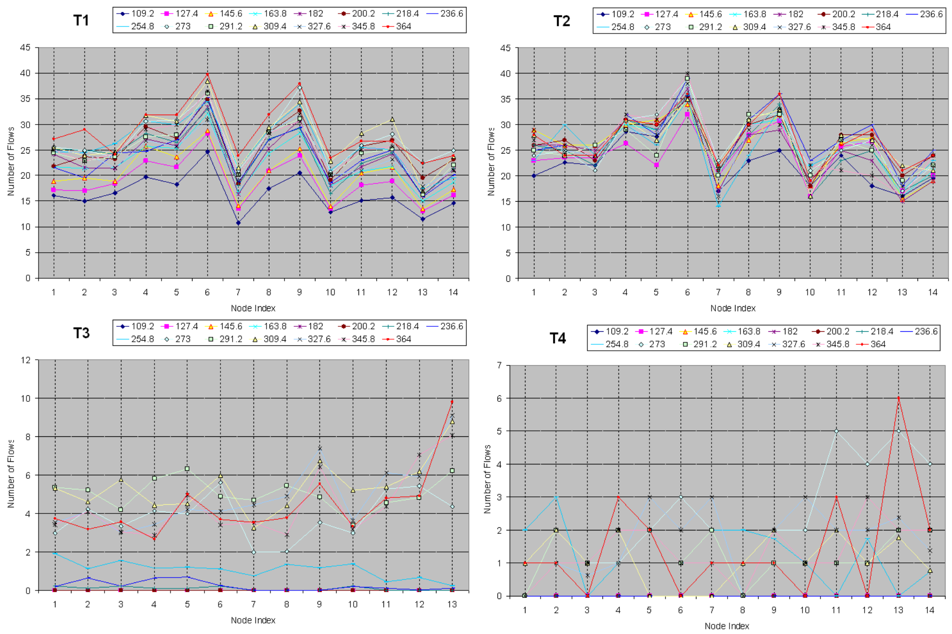


Figure 5.10: *Single-Domain, Single-Service Scenario, Experiment 5.3.1, Number of Flow-through LSPs.*

One interesting result in the LSC layer is the distribution of "flow-through" LSC LSPs over the nodes in the network domain. LSC LSPs are switched (routed) in the LSC layer, at the nodes that have no visible grooming capability (reduced visibility scenarios) or that have exhausted the available transceiver ports (full visibility scenarios). These LSC LSPs can be terminated at hybrid nodes. In the observed load range, and for configuration in which each node can terminate all the wavelengths of all its WDM links (for a node with a physical layer degree 3, this makes 24 transceivers in both directions), we measured the number of flow-through LSPs at each node.

The diagrams showing number of flows for all loads, all nodes, and all test cases are summarized in Figure 5.10. Fig 5.10 shows that two nodes (node 6 and node 9) have the highest number of flows in T2 and T1 scenarios. In T3 and T4 scenarios, at the higher loads, fewer flow-through LSPs were established, i.e., some LSPs of the length higher than one physical hop emerged. A high number of flow-through LSPs, particularly at the lower loads, indicates that the node belongs to a high number of shortest paths in the physical topology. Such nodes can be good candidates for introducing grooming capability.

This experiment has shown the differences between the two RToE strategies and the impact of releasing virtual links for constant and uniform traffic. The CMB strategy shows significantly better results. An important conclusion is that immediate release of an idle link is not useful for some medium loads. The link-release strategy should, therefore, discriminate between idle links and release only a sub-set of them. A policy for releasing idle links in a real world situation should also consider that the configuration of a virtual link takes some time and may be prohibitive for some short-duration services.

5.3.2 Impact of Different Bundling Strategies in a Single Service Scenario

As presented, the infrastructure service model uses GMPLS resource management that is based on link bundling. We have discussed the relationship between the aggregated TE link and the component links, with emphasis on the weighting of the aggregated link. In the previous Experiment 5.3.1, we used one particular strategy to assign the weight to an aggregated link based on the weights of its components and to select the labels of the links assigned the routing of a traffic LSP. This was a MEAN-cost strategy. In this experiment, the impacts of five different strategies were evaluated. The strategies included: MEAN-cost bundling, in which a bundle is assigned a mean cost of all bundled labels (SVG procedure $getA()$), and the labels are selected randomly (in the SVG procedure $Update()$); MIN-cost (and MAX-cost) bundling, where the min (max) value is assigned and the min (max) cost labels are selected; ONE-cost bundling, where the cost is always equal to one and the selection is random; and residual-cost bundling, where a cost is calculated proportional to $1/R$, where R is the residual bandwidth of the aggregated link, and the selection is random. We compare the impact of these strategies in test cases defined in Experiment 5.3.1 (Table 5.1), for the same topology and load range. The blocking probability is compared in Figure 5.11. We see that the differences are relatively insignificant and that component links (LSC LSPs) have greatly differing hop counts and are rarely established and bundled. This also shows that in all other experiments in this topology and with the same traffic, we can use the MEAN-cost weighting strategy and assume very similar results for other weighting strategies. This is an important result because it shows that the link bundling approach that guarantees scalability can be adopted without substantial performance concerns.

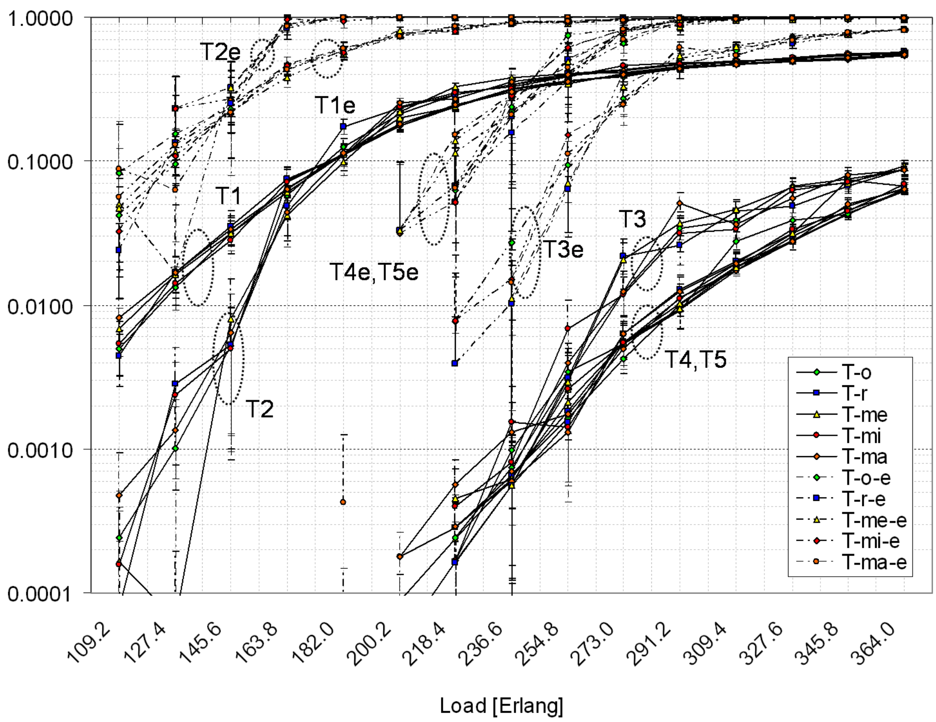


Figure 5.11: *Single-Domain, Single-Service Scenario, Experiment 5.3.2, Blocking Probability for Different Bundling Strategies.*

5.3.3 Impact of Grooming Visibility in a Single Service Scenario

In this experiment, we illustrate the benefit of adapting visibility. We show how the observed number of flow-through LSPs can be used to change grooming visibility and achieve better performance. We use results for the number of flow-through LSPs per node, obtained in study cases T1 and T2 of the previous experiment, in which no grooming PE/P capability is used (Figure 5.10). These results correspond to a study case in which all PE nodes have grooming capability that is, however, not visible to a service. In this experiment, the nodes with a peak value of flow-through LSPs are taken as candidates for being assigned grooming visibility. We further examine the performance gain achieved when either node 6, node 9, or both become grooming visibility. Compared test cases are summarized in Table 5.5. All simulations are repeated with CMB RToE and RELI and NREL ToE link release strategies. Figure 5.12 and Figure 5.13 compare the blocking probability obtained in the test cases with the release-when-idle (RELI) strategy and with the no-release (NREL) strategy. It can be observed that by just changing the visibility of a single node with RELI, we can achieve a blocking probability of 3% instead of 10%. With NREL and two visible nodes, we can achieve a blocking probability less than 1% in comparison to 10%.

Test Case	Visibility	RToE Strategy
T1	reduced	LBL
T2	reduced + n_9+ n_6	CMB
T3	full	CMB
T4	reduced + n_9	CMB
T5	reduced + n_6	CMB

Table 5.5: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Test Cases.*

The number of flow-through LSPs also decreased as shown in Figure 5.14. In a study case in which nodes n9 and n6 both gained grooming visibility, the positive impacts on the utilization of inter-domain, access, and virtual links, are shown in Figure 5.15.

The maximum bandwidth of virtual links also changed, approaching the results of the study case from the previous experiment in which full grooming capability is used. They

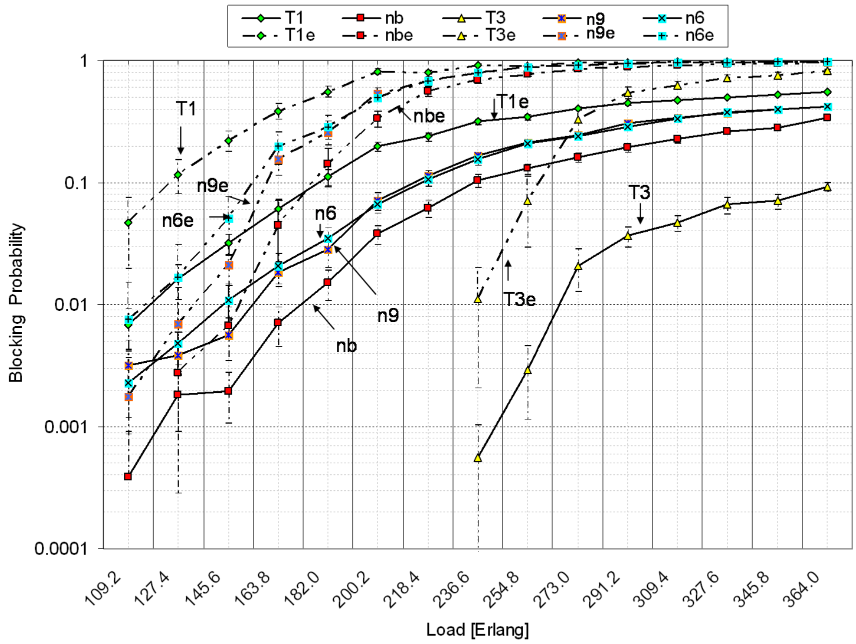


Figure 5.12: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Blocking Probability with Release-when-Idle (RELI).*

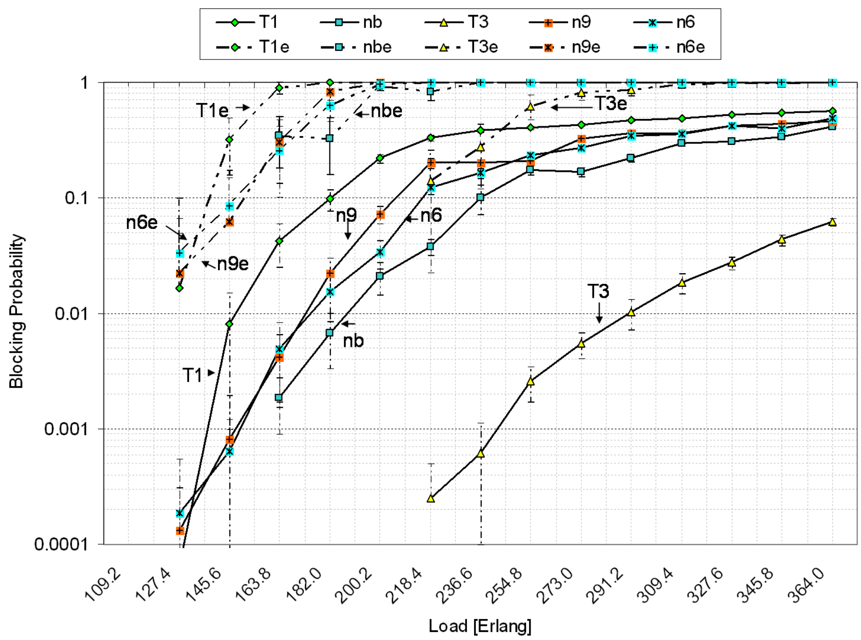


Figure 5.13: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Blocking Probability with Never-Release (NREL).*

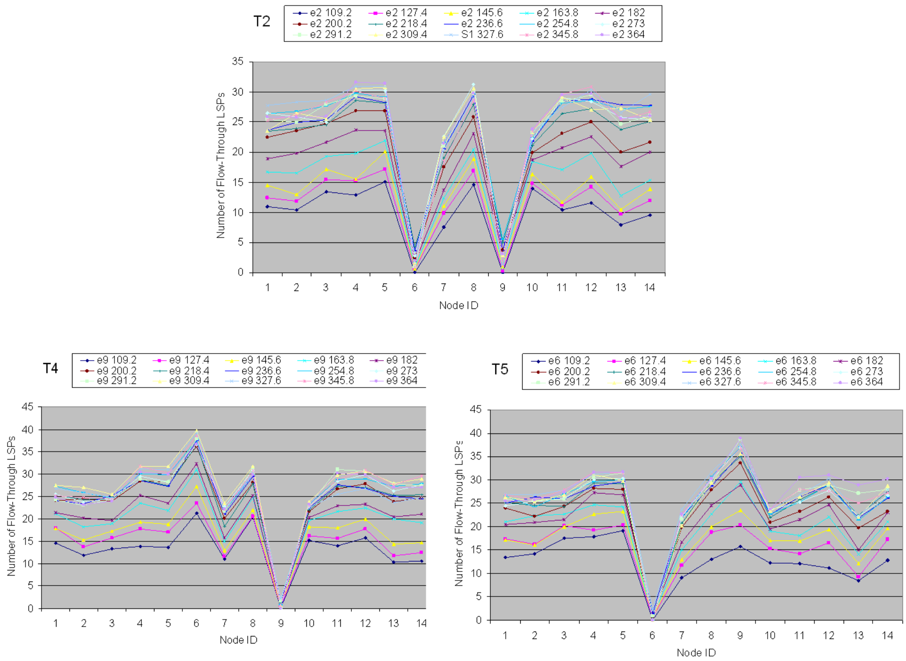


Figure 5.14: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Number of Flow-through LSPs.*

are shown in Table 5.6.

Maximum Virtual Link Bandwidth					
Load	T1	nb	T3	n9	n6
109.2	5.37755	13.1447	13.6817	12.4922	12.1403
127.4	5.57505	16.4701	15.5554	13.6948	13.9127
145.6	5.91296	16.1677	18.0802	14.9722	15.8276
163.8	6.41022	17.2881	20.2765	16.4152	17.386
182	8.7101	18.7095	21.5612	18.397	17.8731
200.2	9.99566	21.0299	23.1256	18.1284	19.633
218.4	8.57667	20.6518	25.847	19.5912	19.8408
236.6	9.34431	22.1709	26.9407	20.4108	20.2453
254.8	10.3834	21.3263	27.8623	22.2507	22.7895
273	11.6647	23.3105	30.3058	22.2548	22.3256
291.2	15.9538	24.6037	27.5661	23.6114	21.6608
309.4	15.9361	25.1962	28.199	21.8875	21.8285
327.6	15.8883	25.8599	31.2055	22.3953	21.5078
345.8	11.9848	26.4699	30.9583	23.7114	22.9094
364	15.8957	26.7999	33.0906	24.0023	23.4015

Table 5.6: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Average Maximum Virtual Link Bandwidth.*

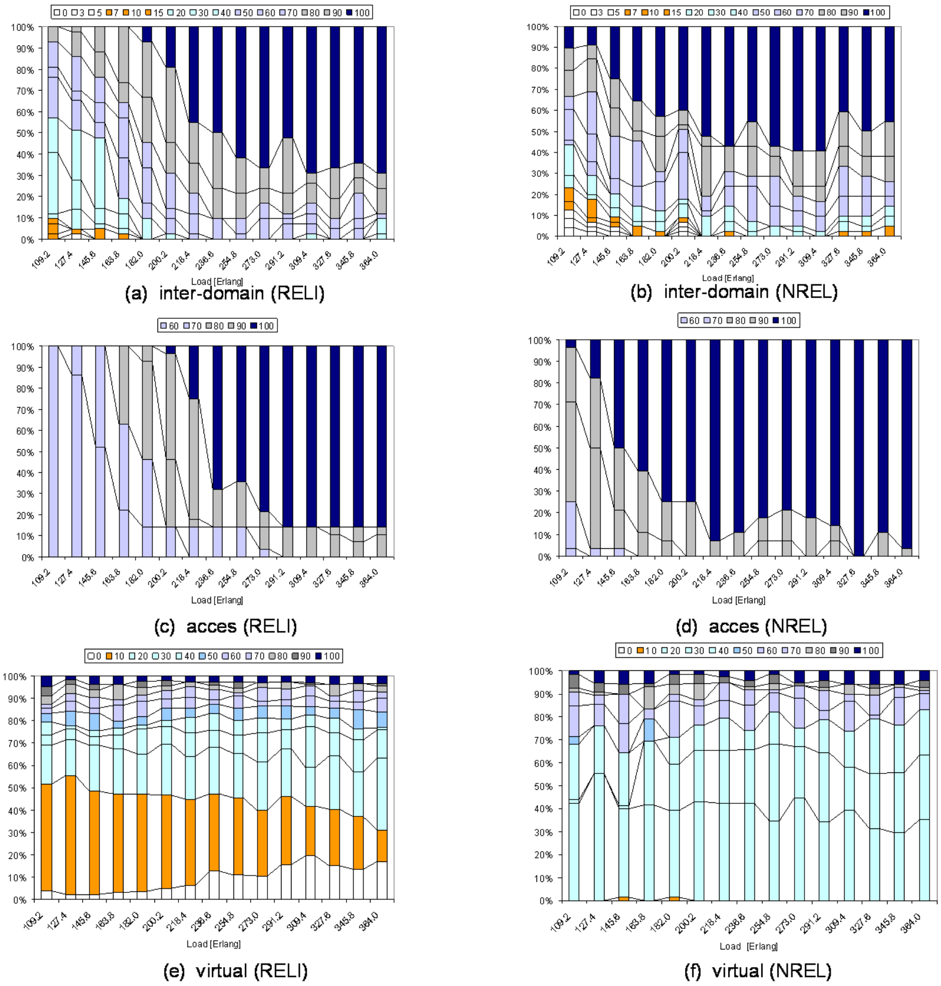


Figure 5.15: *Single-Domain, Single-Service Scenario, Experiment 5.3.3, Link Utilization with Grooming at Nodes n9 and n6.*

5.3.4 Impact of Different Traffic Granularities

In this experiment, we changed the assumption about the traffic LSP bandwidth that we used in Experiment 5.3.1, where all traffic LSPs had bandwidths equal to 1 TDM label. In this experiment, we vary the bandwidth requirement, keeping the assumption that all traffic sources generate traffic with the same characteristics. The arrival rate is adjusted, so that the same load range can be studied. The test cases from Experiment 5.3.1 (Table 5.1) are repeated for traffic LSP bandwidths of 2, 3, and 4 labels. Two approaches for traffic LSP accommodation are simulated: (1) when TDM labels of traffic LSPs are spread over different paths, and (2) when all TDM labels are routed along the same route. The motivation for this experiment is in understanding the

In Figure 5.16, we compare blocking probability in test case T1 (LBL, RELI) and T2(CMB, RELI) for different traffic LSPs' granularity. It can be seen that (LBL, RELI) and (CMB, RELI) behave similarly for the traffic with 2 labels (T1-C2, T2-C2). In this case, traffic spreading does not bring many benefits. For traffic with 3 labels the benefit of spreading is more obvious. Results comparing study cases T2 (LBL, NREL) and T5 (CMB, CREL) are similar to Figure 5.17. However, when comparing the performance of LBL with CMB RToE strategies for the traffic with mixed granularities, i.e., a traffic LSP is randomly assigned bandwidth requirement of 1, 2, 3, or 4 TDM labels, the CMB strategy again shows better results.

These results show that CMB RToE always performs as well as or better than the LBL strategy for some traffic patterns, but at the cost of using grooming resources. The positive impact of traffic spreading shows that, if separate routing of labels multiplexed within the traffic LSP is possible, it should be used.

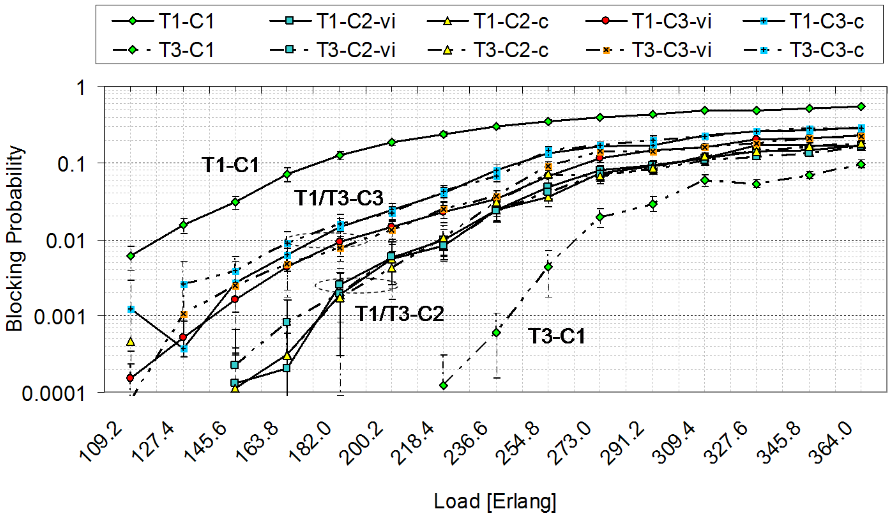


Figure 5.16: *Single-Domain, Single-Service Scenario, Experiment 5.3.4, Blocking Probability.*

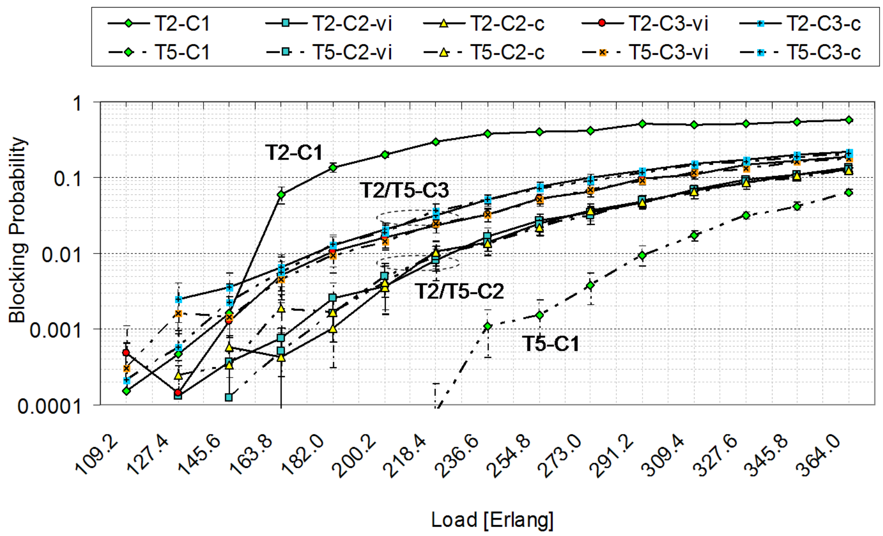


Figure 5.17: *Single-Domain, Single-Service Scenario, Experiment 5.3.4, Blocking Probability.*

5.3.5 Impact of Sharing in a Multi-Service Single-Domain Scenario

In a multi-service scenario, we look at the impact of sharing resources among services. We defined study cases with reduced, full, and constrained visibility. The constrained visibility case is particularly interesting here, since it is the result of a very simple virtual topology design heuristic. Starting from the basic assumptions made in the previous experiment, we divided the traffic into four services. Three services are shown in Figure 5.18; the fourth service S_1 is not shown, because it includes all CE nodes. In other words, on each CE, two services are active: S_1 and either S_2, S_3 or S_4.

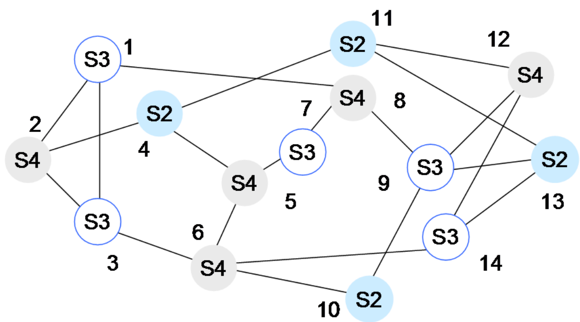


Figure 5.18: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Multi-Service Network Setup.*

All services generate traffic with the same characteristics. Services S_2, S_3 and S_4 only generate traffic among their corresponding service members, e.g., the four customers in service S_2 can initiate traffic LSPs to each other, hence 4×3 traffic sources belong to S_2 service. In addition, service S_1 includes all possible "inter-service" traffic, e.g., a traffic LSP requested between the CE node connecting at the PE node n1 (Service S_3) and the CE node connecting at the PE node n11 (S_2) belong to a service S_1.

Table 5.7 summarizes the configuration visibility in this example. In the first two cases, each service can use only the grooming capability at the CE nodes and the corresponding PE nodes. In the third and fourth cases, grooming visibility is full. For the same visibility, the study cases differ in the level of sharing. In the first and third cases, services do not share resources, and in the second and fourth cases, they support full sharing. The strategy used is CMB with RELI.

Study Case	Grooming Visibility	Sharing
T1 (nr)	only CE nodes	none
T2 (nf)	CE nodes + Service PE node	none
T3 (sr)	CE nodes + Service PE node	full
T4 (sf)	all nodes	full

Table 5.7: *Single-Domain, Single-Service Scenario, Experiment 5.3.5, Test Cases.*

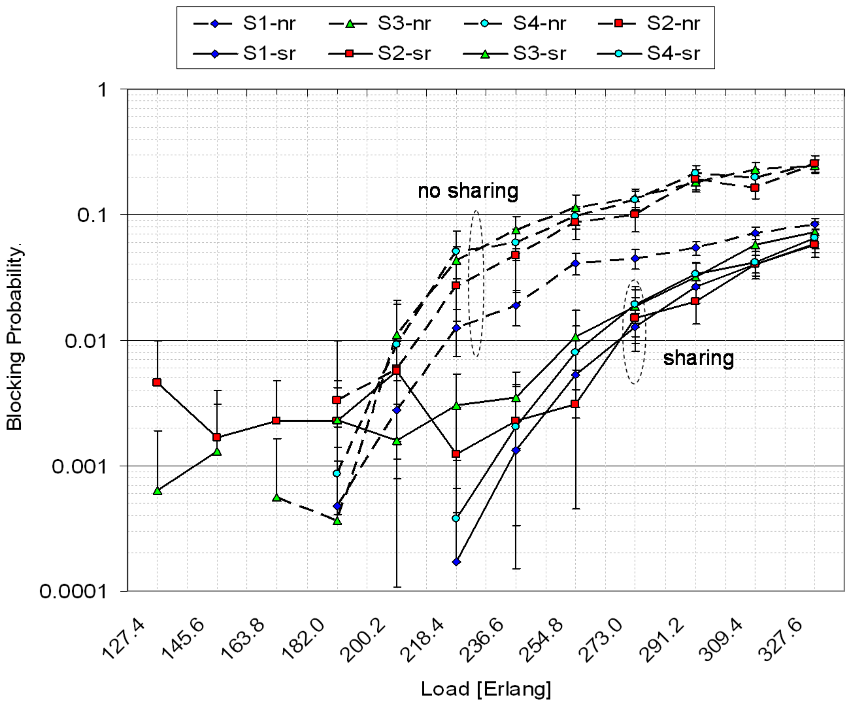


Figure 5.19: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Blocking Probability in Study Cases with Reduced Visibility.*

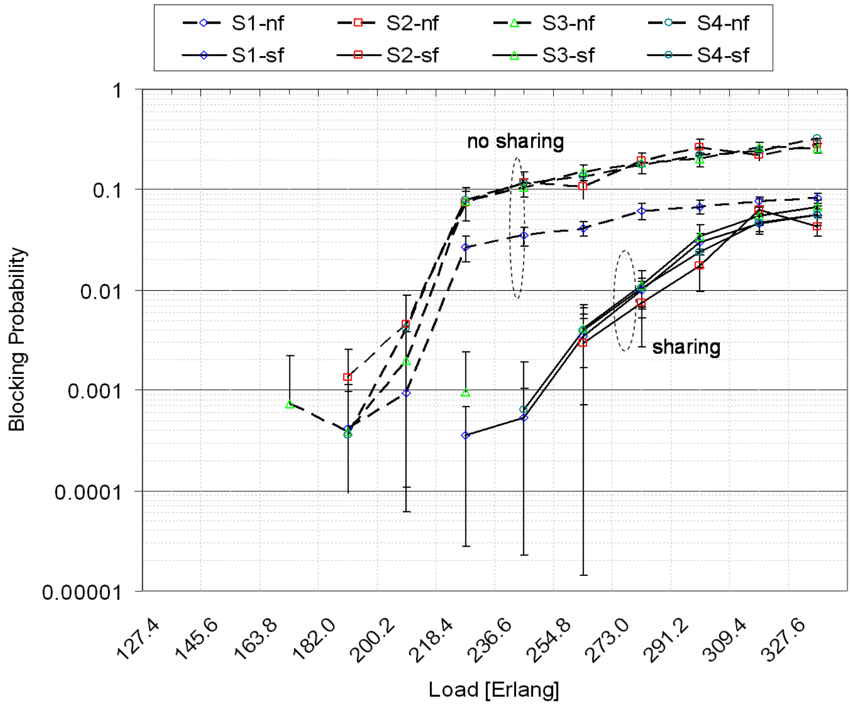


Figure 5.20: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Blocking Probability in Study Cases with Full Visibility.*

Figure 5.19 and Figure 5.20 show the blocking probability per service, in test cases T1 and T3 and in T2 and T4, respectively. In both Figures, the benefits of sharing resources between services can be seen. Here, full sharing means that any service can use unused labels from any component link (LSP LSCs) created by other services' RToE actions. In the test cases without sharing between services, we see that service S_1 experiences the lowest blocking. Service S_1 has the highest spread, and its customers can share resources service-internally. Other services depend on the smaller number of virtual links that needs to be available. When the services share resources, this difference disappears. The gain that can be achieved with sharing is significant in both reduced and full-visibility scenarios.

In Figure 5.21 we compare blocking probability in the scenarios with sharing and see that giving services full visibility (T4), as compared to the visibility of only CE and corresponding PE nodes, provides some improvement that is, however, relatively small in the range of blocking probability up to 3%.

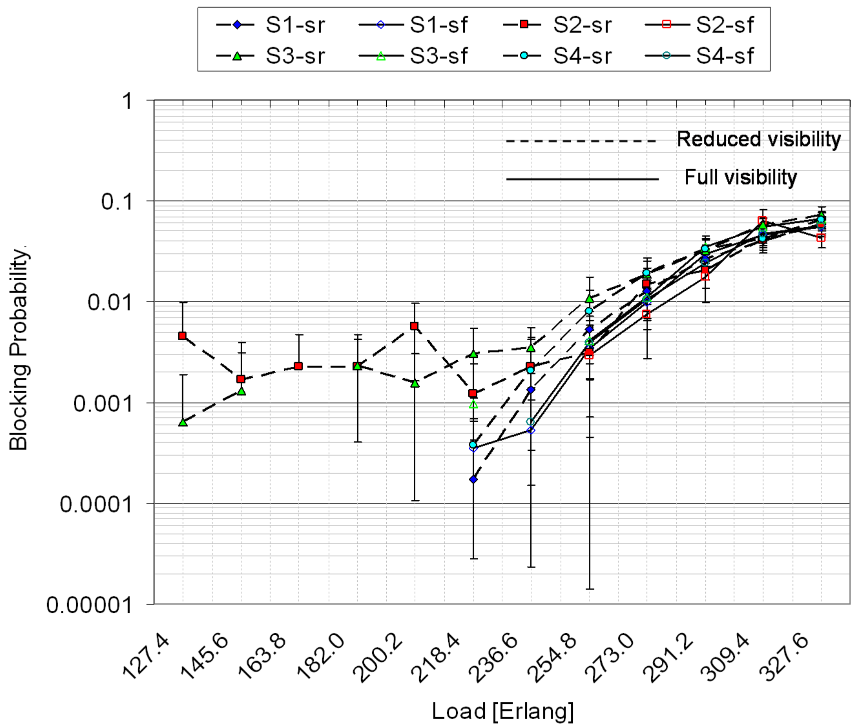


Figure 5.21: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Blocking Probability in Study Cases with Sharing.*

In Figure 5.22 and Figure 5.23, the TDM layer resource usages for all services are shown. The results obtained in test cases with different visibilities and sharing, i.e., T3 and T4, show almost no significant difference. Each of the four diagrams represent one particular service, e.g., S_1, and for different loads, the percentage of service traffic LSPs that are carried over resources of other services. For example, approximately 70% of all S_1 traffic LSPs are routed over the component links that are set up by S_1 itself. Service S_1 uses resources of services S_2, S_3, and S_4 to route the other 30% of its traffic. Services S_3 and S_4 can share approximately 9% of their resources.

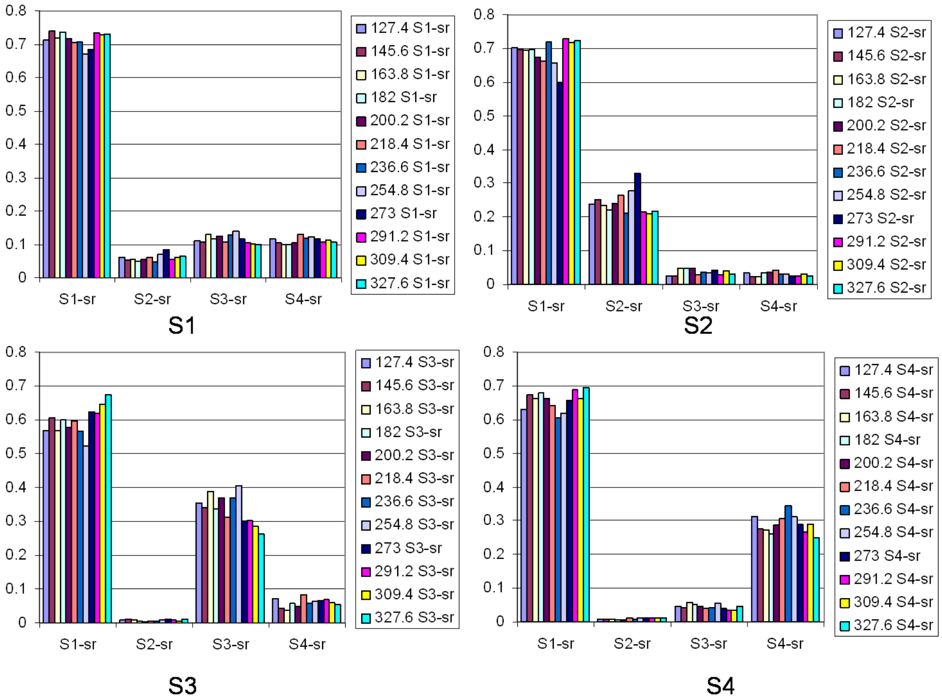


Figure 5.22: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Sharing Resources between Services, Scenario T3*

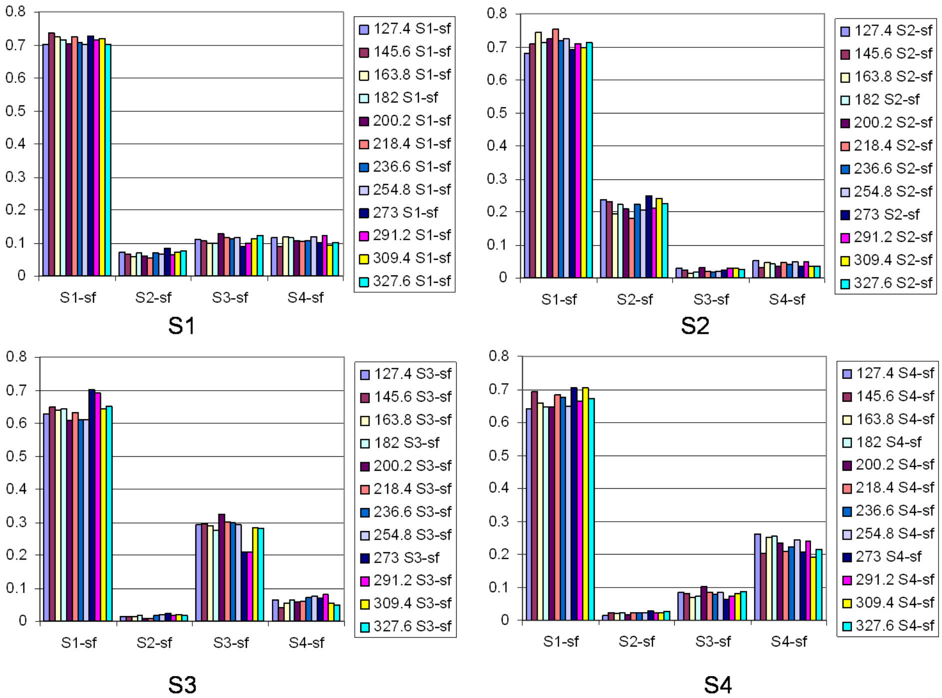


Figure 5.23: *Single-Domain, Multi-Service Scenario, Experiment 5.3.5, Sharing Resources between Services, Scenario T4*

5.3.6 Impact of Sharing in a Multi-Service Single-Domain Scenario with Different Traffic Granularity

This experiment uses the same configuration of services and resource visibility as described in Experiment 5.3.5. However, the traffic that services generate is slightly changed. The aim is to see whether the previous conclusion about the benefits of sharing are still valid. We defined three groups of test cases, summarized in Table 5.8. In the study cases of the first group, while service S_1 still generates traffic LSPs with required bandwidth equal to 1 TDM label, services S_2, S_3, and S_4 generate traffic with required bandwidth of 2 TDM labels. In the second group of scenarios, S_1 can generate traffic LSPs with either one or two TDM labels, where the probability of the lower bandwidth requirement is 2/3, and the higher bandwidth requirement is 1/3. A request with a higher bandwidth requirement comes with a lower inter-arrival rate. In the third group of test cases services, S_1, S_2, S_3, and S_4 behave as S_1 in the second group, and service S_2 generates requests for two TDM labels. Again, we compare sharing and no-sharing options with full and constrained visibility. The results confirm the conclusion from the previous experiment. By sharing virtual links, extensibility of services can be improved. In this experiment, better results are obtained with the visibility configuration that is obtained, taking the service configuration into account. This exemplifies the need for an efficient visibility configuration design algorithm.

Test Case	S_1 traffic	S_2 traffic	S_3 traffic	S_4 traffic	Grooming Visibility	Sharing
T1 (nr)	1	2	2	2	reduced (only service PE nodes)	no
T2 (nf)	1	2	2	2	full	no
T3 (sr)	1	2	2	2	reduced	yes
T4 (sf)	1	2	2	2	full	no
T1b (nr)	1 (60%) 2 (30%)	2	2	2	reduced (only service PE nodes)	no
Tb2 (nf)	1 (60%) 2 (30%)	2	2	2	full	no
T3b (sr)	1 (60%) 2 (30%)	2	2	2	reduced	yes
T4b (sf)	1 (60%) 2 (30%)	2	2	2	full	no
T1bo (nr)	1 (60%) 2 (30%)	2	as S_1	as S_1	reduced (only service PE nodes)	no
Tb2o (nf)	1 (60%) 2 (30%)	2	as S_1	as S_1	full	no
T3bo (sr)	1 (60%) 2 (30%)	2	as S_1	as S_1	reduced	yes
T4bo (sf)	1 (60%) 2 (30%)	2	as S_1	as S_1	full	no

Table 5.8: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Test Cases.*

Comparing the results achieved for reduced visibility (Figure 5.24) with the corresponding

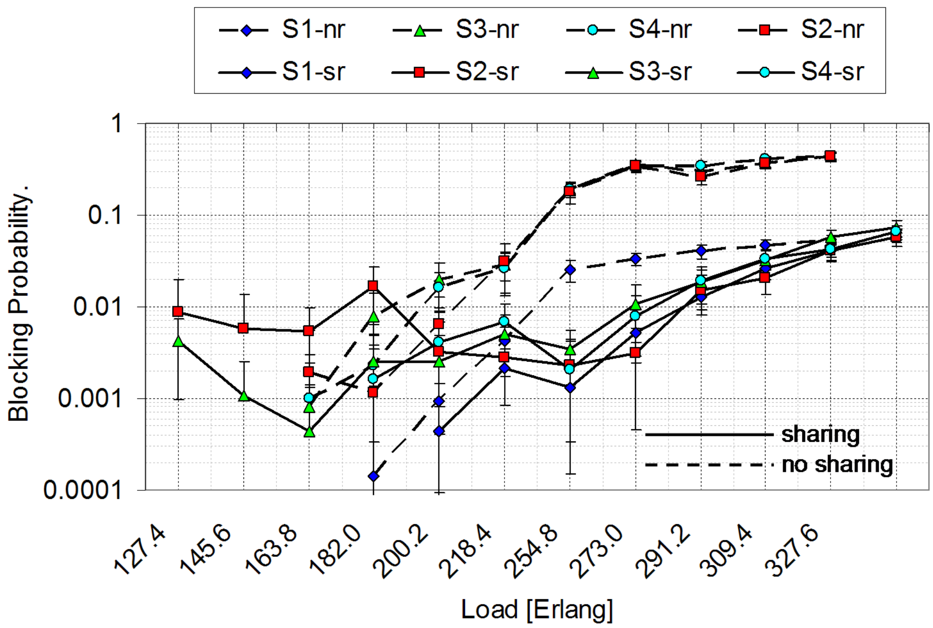


Figure 5.24: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Reduced Visibility.*

results in the previous experiment (Figure 5.19), it can be observed that, without sharing, the blocking probability of services S_2, S_3, and S_4 becomes higher due to higher bandwidth requirements. With sharing, results comparable to those of the previous experiment can be obtained, in particular in the load range where blocking is over 10% without sharing.

Figure 5.25 compares results with full visibility. Again sharing of resources reduces the blocking probability.

Comparison between results of sharing scenarios shows that controlled sharing with the constrained, i.e., service-specifically designed, visibility is better option, as all services including S_1 show better performance. With the constrained visibility, the services share only resources that are topologically most appropriate.

Similar results are obtained when service S_1 changes (increases) its bandwidth requirements slightly and generates less virtual capacity available for sharing. However, again with sharing both for constrained (Figure 5.27) and full visibility (Figure 5.28), better results can be obtained.

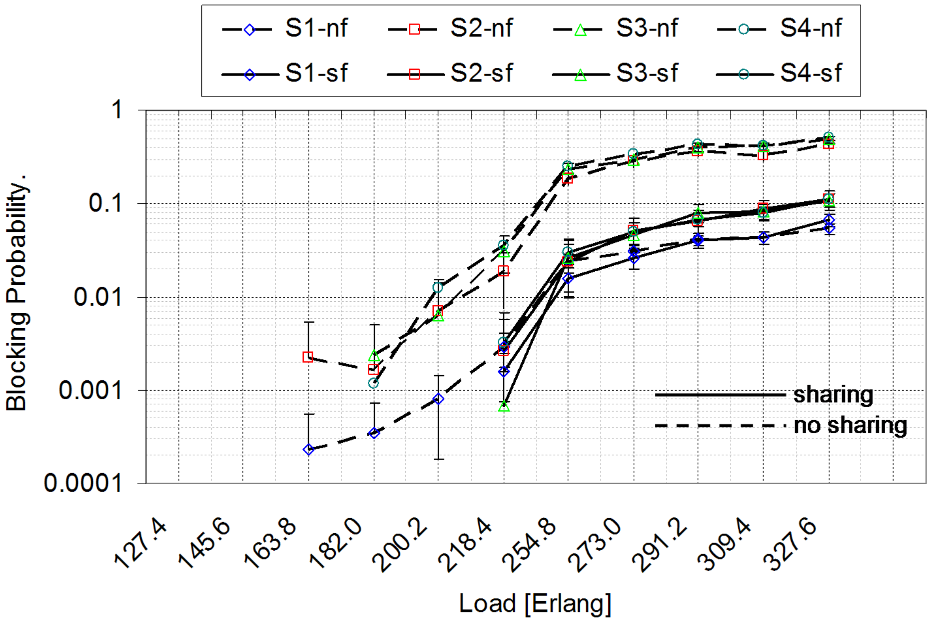


Figure 5.25: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Full Visibility.*

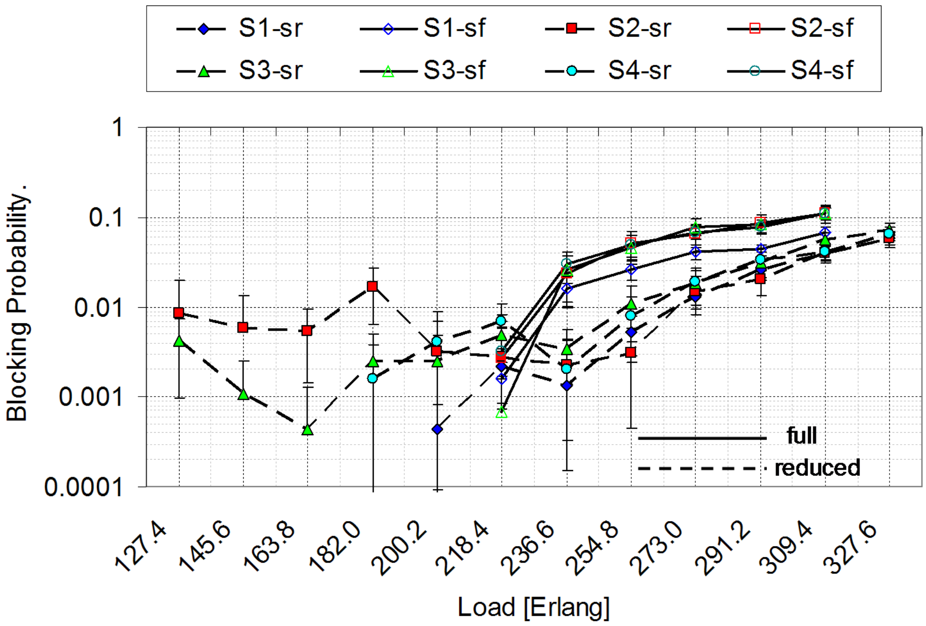


Figure 5.26: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Sharing.*

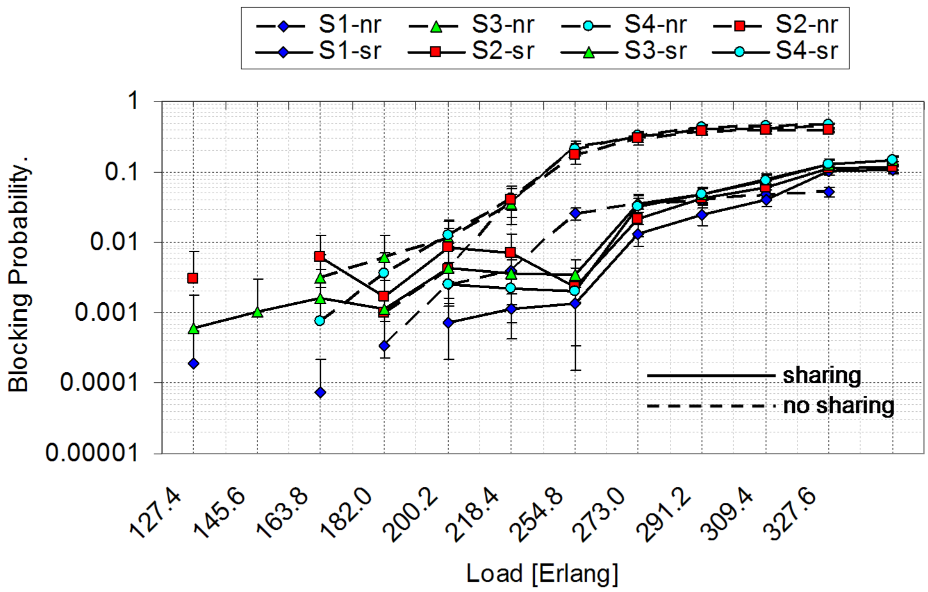


Figure 5.27: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Constrained Visibility.*

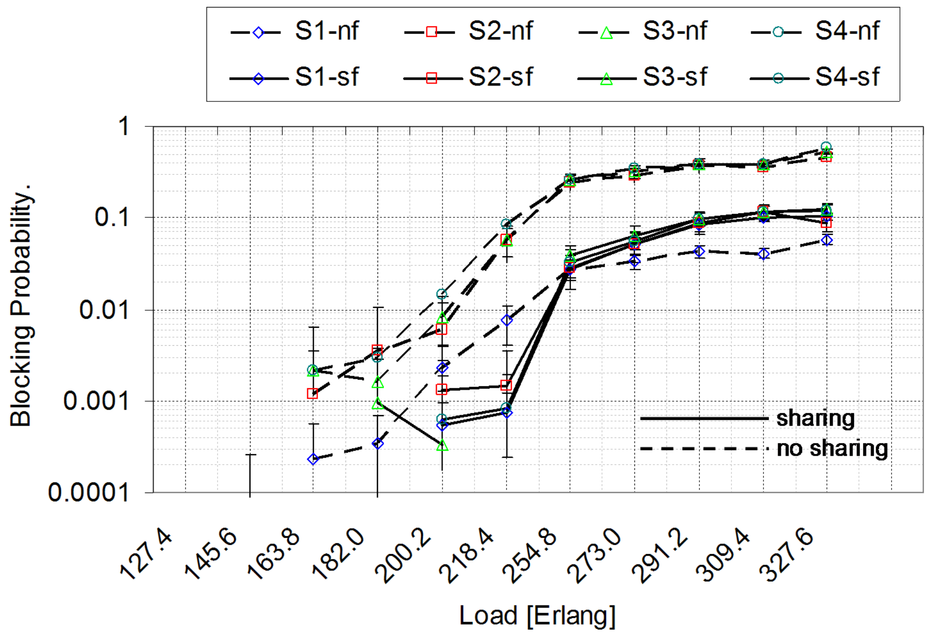


Figure 5.28: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Full Visibility.*

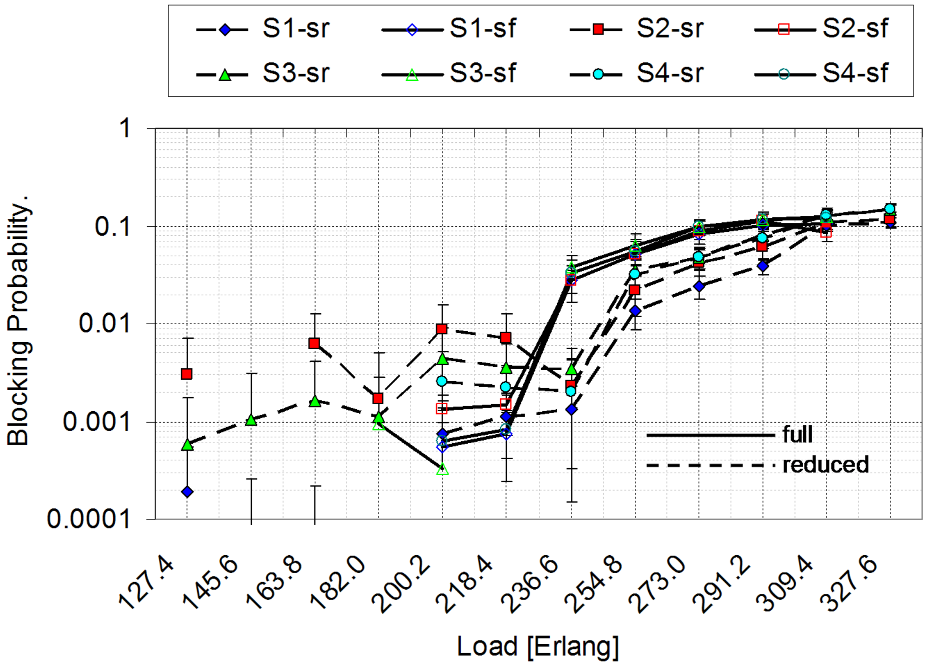


Figure 5.29: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Blocking Probability in Study Cases with Sharing.*

In cases in which all services, except service S_2, generate traffic of variable capacity requirements (either one or two TDM labels are required), no-sharing service S_1 again shows lowest blocking. Service S_2 now has the highest blocking. With sharing, the performance can be increased when both reduced (Figure 5.30) and full visibilities (Figure 5.32) are used. Here again, the best result can be obtained with sharing in the case featuring constrained visibility (Figure 5.35).

The results of these experiments show an important direction regarding sharing of virtual links between services. We have seen that sharing resources between services introduces performance gain. We also noticed that if resources are offered for sharing without constraints, lower performance may result than when the shared resources are constrained. How the visibility of grooming capabilities is configured for different services has direct impact on sharing on the logical level, hence, it can be a means to govern efficient sharing.

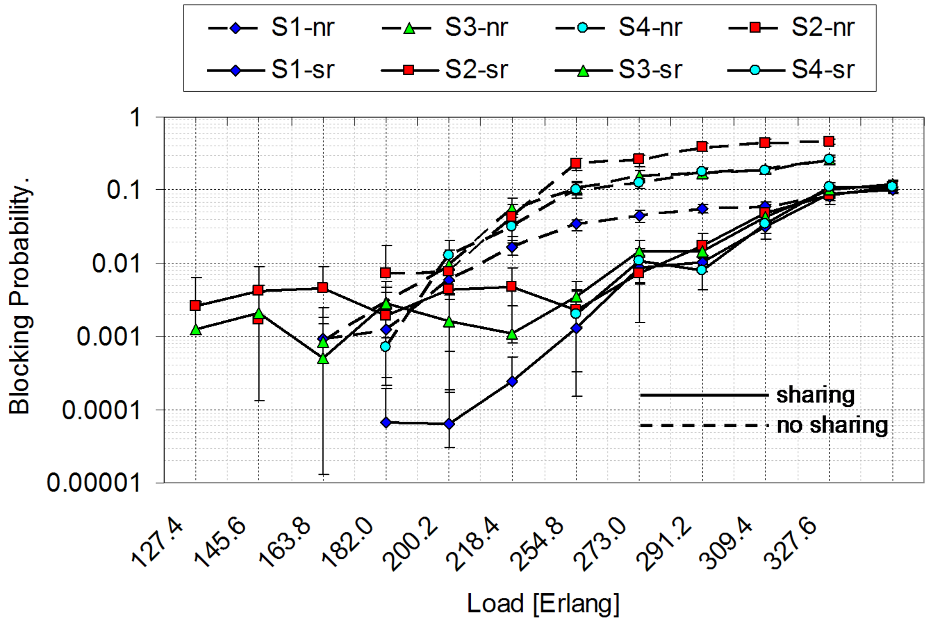


Figure 5.30: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Sharing.*

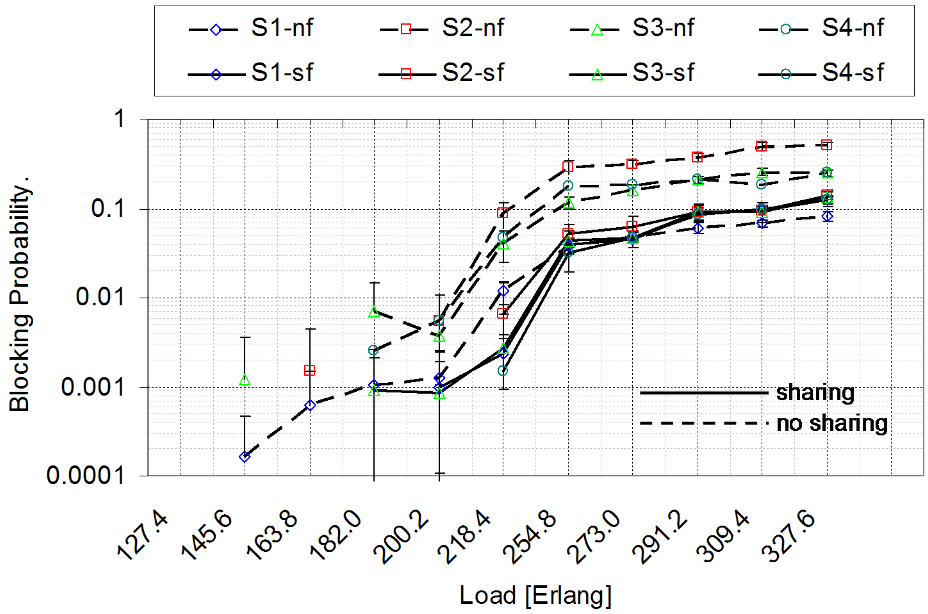


Figure 5.31: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Sharing.*

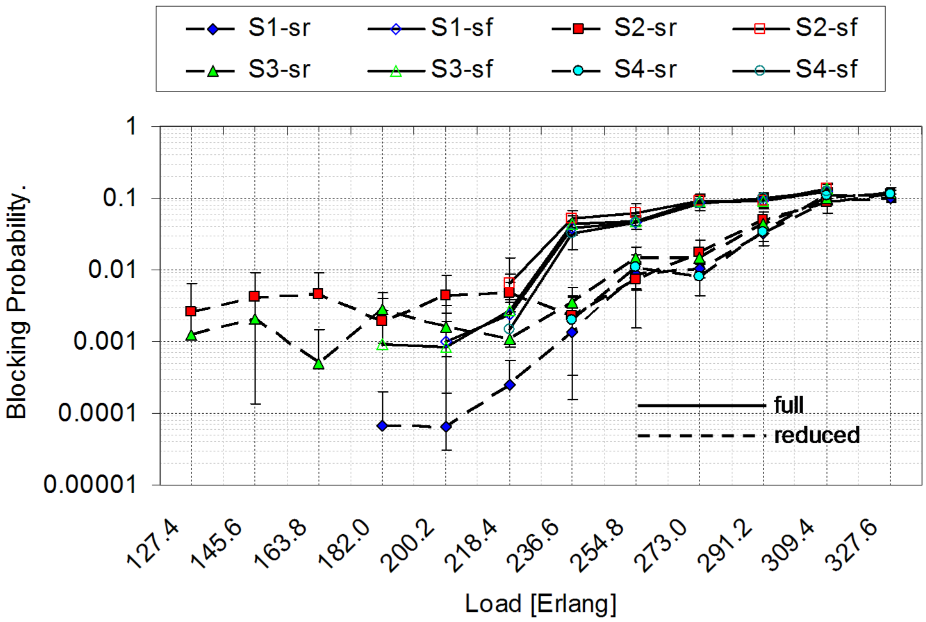


Figure 5.32: *Single-Domain, Multi-Service Scenario, Experiment 5.3.6, Sharing.*

5.4 Performance of Infrastructure Services in Multi-Domain Scenario

The simulation experiments in a multi-domain scenario are designed to quantify the impact that the architecture with GMPLS-enabled Exchange Points (GXP) has on service performance. In these experiments, we use, on the one hand, a network topology representing traditional UNI/NNI architecture with the static customer-provider (CE-PE) and carrier-carrier (B-B) connections, and, on the other hand, the architecture with GMPLS exchange points (GXP) where the dynamic inter-connectivity between the interconnected parties can be established on-demand over the deployed GXPs, and the MPE control overlay performing path computation is assumed. We simulate a network, shown in Figure 5.33, composed of two domains (D_1, D_2) with identical topologies including 15 hybrid nodes with lambda switching, TDM switching, and grooming capability. Each domain node serves a customer site.

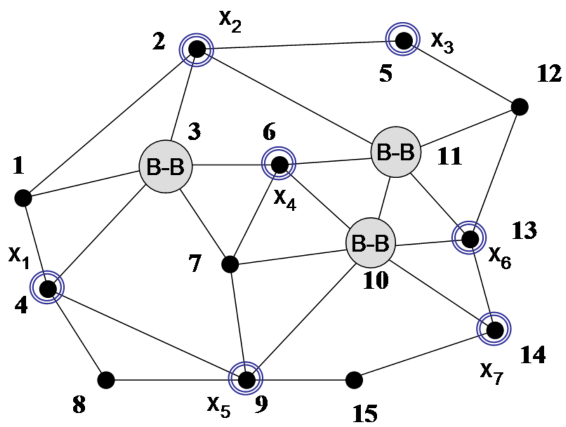


Figure 5.33: *Multi-Domain, Network Topology.*

In the UNI/NNI topology, domains establish inter-domain B-B links at three points of collocation, (B-B) in Figure 5.33. For the corresponding GXP-topology, the placement of seven GXP nodes is computed (shown as dotted circles) by using the heuristic presented in Section 3.5, starting with 15 candidate locations. In both UNI/NNI and GXP topologies, B-B, CE-PE, and GXP links have 32 wavelengths and each inter-domain link, either B-B or over GXP, has 128 wavelengths. Each GXP connects both domains and both customer sites at this collocation. In the considered test cases, we simulated different grooming visibilities and considered different sharing scenarios.

In the UNI/NNI topology, the visibility can either be reduced, where all LSRs offer only LSC switching capability, including border LSRs or partial where border LSRs offer grooming capability, or full, if all LSRs offer TDM switching and grooming. In the experiments with the GXP architecture, we again vary the visibility related to GXP and other nodes in the network. Resource sharing between services can be full or constrained.

In all study cases we evaluate and compare the performance of the $D_1's$ and $D_2's$ intra-domain services S_1 and S_4, respectively, and $D_1 - D2$ inter-domain services, S_2 and S_3. In the GXP architecture, although the customers connect directly to the node, their original assignment to the domain is used in the service definition. In all scenarios, we use CMB with RELI link release strategy. The scenarios studied are summarized in Table 5.9.

Study Case	Topology	Sharing	Visibility (D_1, D_2)	Visibility (B/GRE)	Routing
U1	UNI/NNI	inter-domain	reduced	reduced	domain constrained
U2	UNI/NNI	inter-domain	reduced	full(B)	domain constrained
U3	UNI/NNI	inter-domain	full	full	domain constrained
G1	GXP	full	reduced	reduced	prefer own
G2	GXP	full	reduced	full	prefer own
G3	GXP	full	full	full	prefer own

Table 5.9: *Multi-Domain, Multi-Service Scenario, Test Cases*

Concerning the traffic load, we assume that both intra-domain and inter-domain traffic LSPs have the same duration and that the ratio of inter-arrival times of inter-domain traffic requests to intra-domain traffic requests are 1 : 10. In the UNI/NNI topology, inter-domain requests are provisioned domain-internally, and only inter-domain requests can use inter-domain links. With GXP topology, the "prefer own" routing strategy is used in which the original domain is favored until resources are exhausted, after which the resources of other domains are used.

Figure 5.34 compares blocking probability obtained in study cases U1, U2, and G2. These are the study cases with no grooming visibility within the domains. In the U2 study case, grooming visibility is added at the B nodes. In G3, GXP nodes have grooming visibility. It can be observed that results obtained in U1 and U2 are not significantly different. The GXP architecture (G2), however, introduces significant improvement in blocking probability.

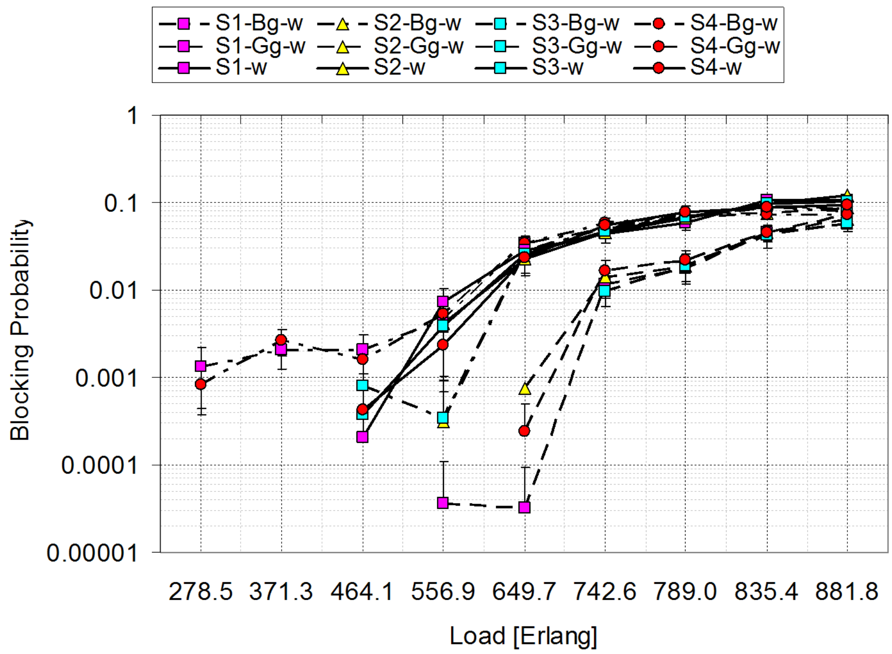


Figure 5.34: *Multi-Domain, Multi-Service Scenario, Blocking Probability in UNI/NNI, with and without grooming on B nodes compared to GXP architecture with GXPa with grooming visibility.*

Figure 5.35 compares blocking probability that can be obtained in GXP architecture with different grooming visibility at GXP nodes, i.e., test case G1 - GXP: no visibility, G2 - GXP: full visibility, and in the whole network, G3 - both GXP and domain nodes: full visibility. The results obtained show that a significant performance improvement can be obtained by introducing grooming visibility at GXP nodes.

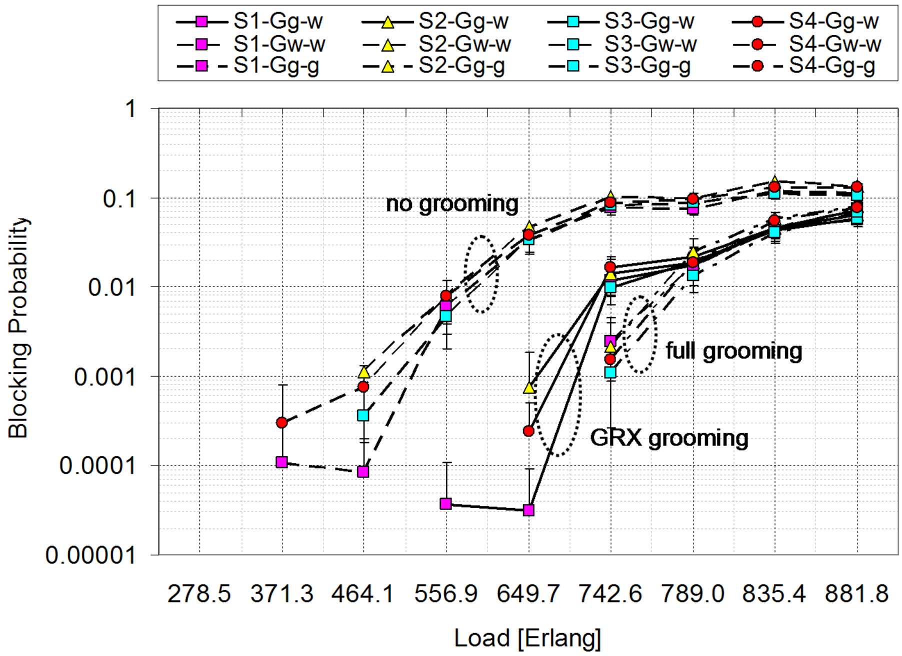


Figure 5.35: *Multi-Domain, Multi-Service Scenario, Blocking Probability.*

Differences in resource sharing between study cases with UNI/NNI (U1) and GXP architectures (G2) are depicted in Figure 5.36 and Figure 5.37. While intra-domain services S_1 and S_4 do not share resources in the UNI/NNI architecture, in GXP architecture this is possible, and it has a positive impact regarding blocking probability.

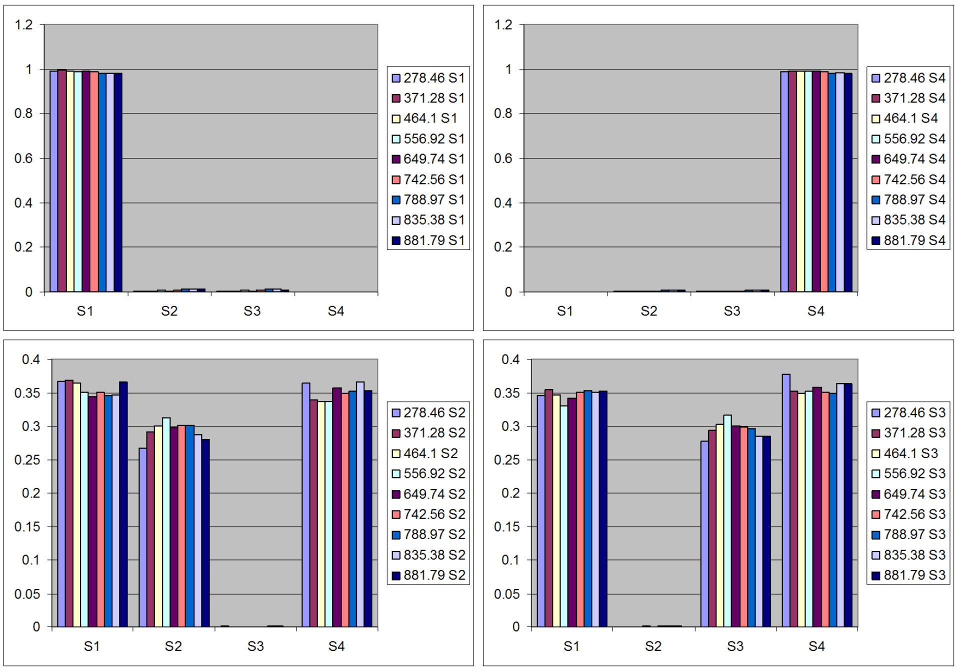


Figure 5.36: *Multi-Domain, Multi-Service Scenario, Sharing in UNI/NNI Study Case U1.*

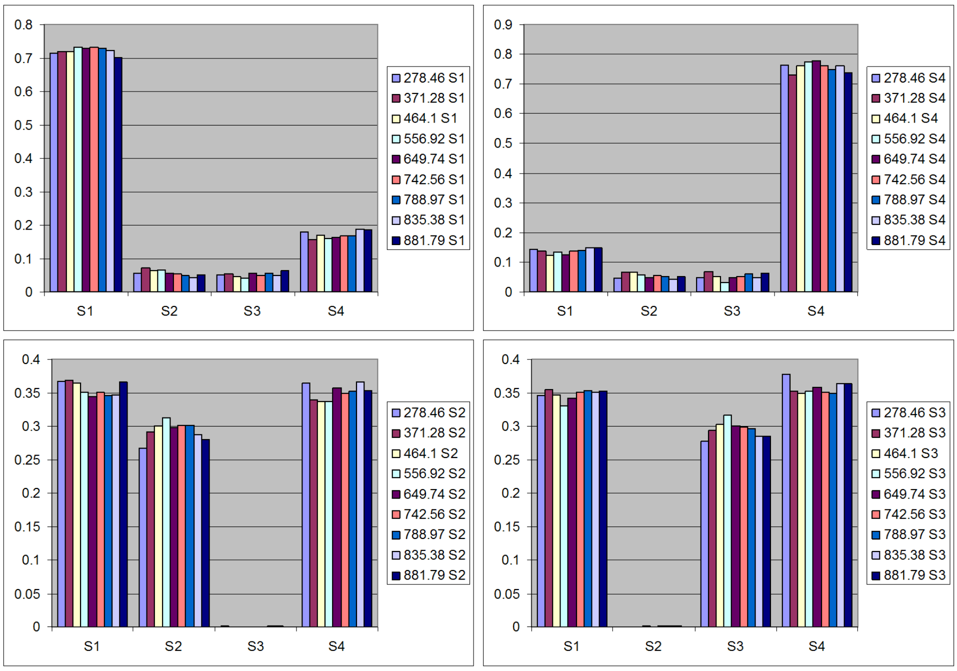


Figure 5.37: *Multi-Domain, Multi-Service Scenario, Sharing in GXP Study Case G2.*

Figure 5.38 compares the utilization of WDM links in scenarios U2 and G2. It can be observed that GXP architecture is characterized with lower utilization at all types of links (for higher throughput/lower blocking probability as previously shown).

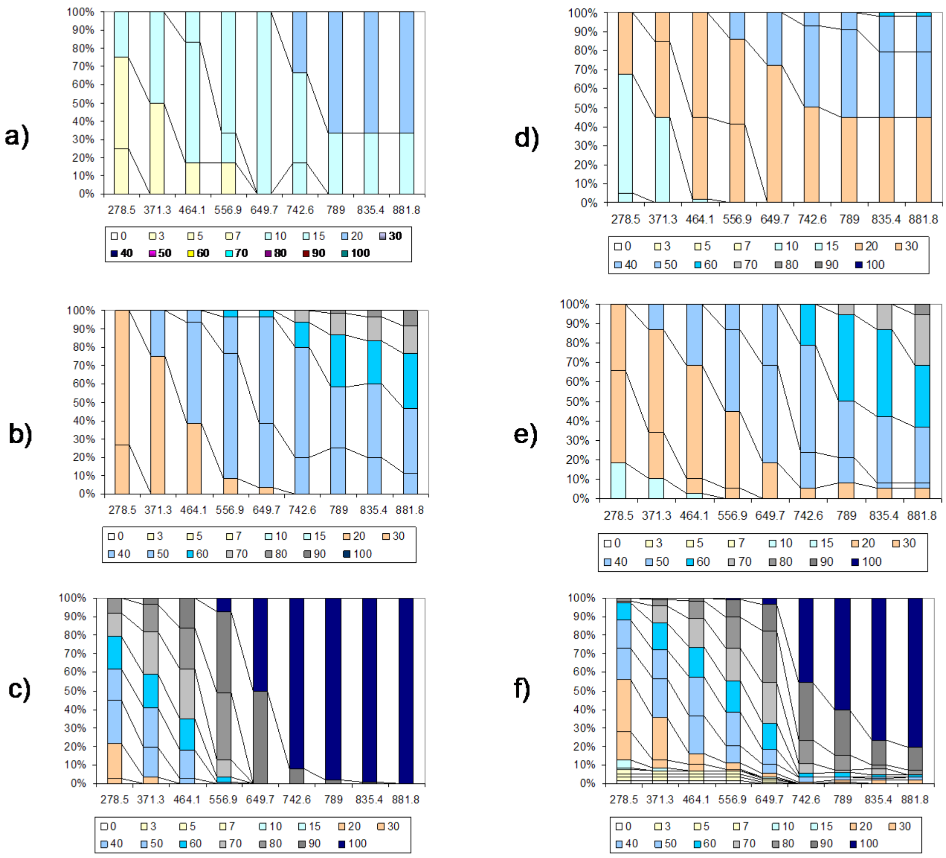


Figure 5.38: *Multi-Domain, Multi-Service Scenario, WDM Link Utilization, (a) U2 access links (b) U2 B-B links (c) U2 intra-domain links (d) G2 access links (e) G2 GXP links (f) G2 intra-domain links.*

Per-service allocation of LSC resources is shown in Figure 5.39. It can be noticed that due to the positive impact of sharing and GXP architecture, the intra-domain services allocate (together) approximately 15% less bandwidth for significantly improved blocking probability.

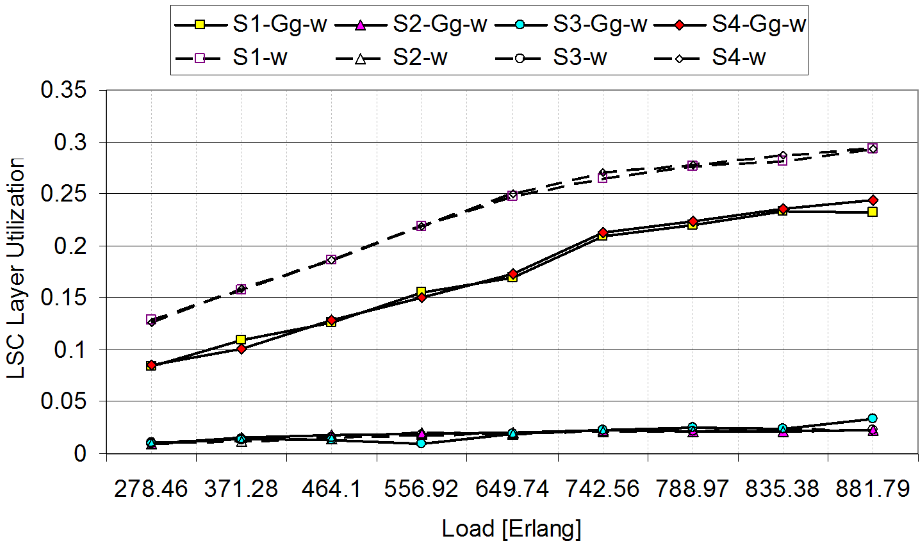


Figure 5.39: *Multi-Domain, Multi-Service Scenario, Per-service Usage of LSC Resources.*

5.5 Conclusions

This Chapter presented the illustrative results of simulation experiments that aimed at quantifying the performance of infrastructure service provisioning with the proposed methods for traffic routing and topology engineering (RToE) and at providing answers to following questions:

1. Which approach to using grooming resources gives better results under the assumption of full grooming visibility: the one that prefers direct links between CE nodes and uses only grooming at the CE nodes, or the one that uses all grooming resources, including those at P and PE nodes?
2. What is the impact of link bundling and routing based on different approaches to the assignment of weights to aggregated links?
3. Which ToE approach to releasing links/bandwidth from the virtual topology leads to better performance in combination with one of the RToEs?
4. Does full visibility of grooming resources offer the best configuration for sharing of virtual resources?
5. What is the performance gain in the GXP architecture?

To this aim, we tested the proposed RToE strategies in the single- and multi-service scenarios and single- and multi-domain topologies. The Layer-by-layer (LBL) RToE and the Combined (CMB) RToE were tested in combination with the proposed link release strategies, i.e., No-release (NREL), Release-when-idle (RELI), and Conditional release (CREL). Tests were repeated for different traffic granularities and in combination with different approaches to weighting bundled links. Results presented here are also illustrative of similar findings in other experiments.

Regarding the first question where the focus was on the configuration with full visibility of grooming resources, the CMB RToE has shown significantly better performance for traffic of very high granularity, and, for more heterogenous traffic patterns and higher bandwidth requirements, it has shown performance comparable to LBL RToE that uses less grooming resources.The CMB routing strategy compared with either RELI or CREL is therefore a strategy of choice for infrastructure service provisioning.

Regarding the impact of different approaches to assigning weights to the aggregated links, we have observed no significant differences for the topology and traffic patterns used. This means that link bundling, which is a resource management strategy adopted from the GMPLS framework, can be used without substantial performance concerns.

Regarding the approach towards link release, we have observed that keeping configured links has positive impact on the service performance. This pointed out that if the link is to be released, the decision should take into account some parameters additional to the link current state, e.g., link utilization. Furthermore, taking into account that in a real-world network the reconfiguration of virtual links should be kept to minimum because of the non-eligible setup time of virtual links, our conditional release strategy, which keeps already configured links active as long as possible, presents a promising approach. This is particularly true when the sharing of virtual links is supported.

Regarding sharing of logical links between services, we have observed that the benefits of sharing may not be the highest when full visibility of physical grooming resources is offered to all services. When services are localized within different parts of the provider domains' topology, a customized set of physical resources should be offered. The visibility concept that we introduced in this thesis is a means to adapt accordingly the set of available resources, both in the physical and logical layer.

In the multi-domain topology, we have observed significant benefits of introducing GXPs, particularly the GXP with the grooming capability.

Summarizing from the previous, the presented results illustrate the promising potential of the proposed Combined RToE, particularly when combined with visibility as a means to control the resources offered to the CMB RToE. These conclusions are also demonstrated within the GXP architecture, where the introduction of GXP nodes resulted in a significant performance gain.

6 Conclusions and Future Work

This thesis developed a network and service model for the dynamic L1 virtualization of optical networks with traffic grooming capabilities. We have proposed two novel concepts in support of service provisioning over heterogeneous technologies and multiple carriers. Firstly, we defined the GMPLS Exchange Point as an enabler of the new *federation architecture* in which different administrative GMPLS domains integrate their resources, in order to provide global services. We proposed a control model for GXP nodes, and introduced a trusted control overlay, referred to as the Multi-Provider Edge (MPE), which exploits the flexibility achieved with GXP and supports global routing. Secondly, we proposed the model of infrastructure services, which employ GMPLS protocols for topology control and traffic routing and share common resources by using a new resource visibility attribute. As a particular case study, we considered the infrastructure services provisioned over two-layer optical networks, with the virtual topology comprised of dynamically established lambda switched paths accommodating dynamically requested TDM granular connections.

One of the most promising properties of the infrastructure service model is resource sharing. To control sharing of physical and logical resources among services we extended the traffic engineering information, which describes network resources within the control plane, with the *resource visibility attribute*. The proposed model foresees that at a service creation time, the visibility attribute of the physical resources is simply updated, without the need to create any dedicated virtual topology for a service *in advance*. During its lifetime, the service can dynamically use visible resources to create its virtual topology and accommodate actual traffic.

The separation between the control of the visibility attribute, on the one hand, and the control over the visible resources by means of traffic routing and topology engineering (RToE) on the other, is one essential novelty of our approach. During an attempt to route - i.e., accommodate - a particular traffic LSP, a service can initiate a dynamic setup of virtual links and allocate physical resources tagged as visible. Established virtual links can be made visible to other services. So, in the proposed model both *physical resources* and *virtual resources* can be made selectively visible to different services, and the visibility of resources can change over time. Furthermore, sharing of the physical and logical layer resources complement each other. In the physical layer, for a resource visible to many services, it depends on the actual traffic requirements which service will allocate the resource. As long as being allocated within one service's virtual link, a physical resource, such as a wavelength, cannot be used within virtual links of other services. After the virtual link is released, all its physical layer resources become available for the ToE actions of other services. At the logical layer, the labels of a virtual link configured by one service may be visible to other services, which may concurrently use them to route their traffic. So, the labels of one service's virtual links may be used to route both service-internal and service external traffic. In this respect, the visibility attribute can reflect different objectives and different service-accommodation policies. Based on the visible resources and actual traffic requirements, the RToE methods and link release strategies transition a virtual topology of a service to a more efficient one.

For a dynamic allocation of visible resources, we proposed two traffic routing and topology engineering (RToE) methods: the Layer-by-layer (LBL) and the Combined (CMB) method. LBL maximizes the single-hop traffic by opening a new direct link each time a requested traffic LSP cannot be accommodated on the existing topology. CMB minimizes delay in the LSC layer by opening short virtual links. The main difference between these two methods is in their approach to using grooming resources: while LBL uses only grooming resources at the customer edge (CE) devices, the CMB method uses all visible grooming resources. The proposed RToE methods rely on GMPLS mechanisms and in particular on link-bundling. Therefore, the proposed service graph model makes the distinction between the aggregated (virtual) links, and the bundled or component links. The RToE methods are further complemented with the link release strategies, such as no-release, release-when-idle and conditional-release.

We implemented the proposed methods in a simulation environment and conducted the simulation-based performance studies. The results presented quantified the benefits of using traffic grooming capability of the carrier networks, and showed a potential of the CMB RToE approach to deliver good performance in link utilization and blocking. The results also showed the need to customize the use of grooming resources, which can be achieved by setting and changing the visibility attributes of the resources precisely. The CMB RToE successfully addresses the potential problem of poorly utilized links that are kept alive only by some service-external traffic. The CMB RToE creates the shortest possible virtual links, which may be efficiently used by many services. This approach complies with the requirement to reduce the amount of re-configurations in the network. Instead of creating and releasing *dedicated* virtual links to route service traffic LSP, which suffers form unused bandwidth, higher blocking, and unnecessary reconfigurations, our approach establishes high-utility shared virtual links, and leverages their yields.

In the performance study, we further observed the benefit of traffic spreading, which indicates that a flexible choice of spreading - i.e., virtual concatenation, can further improve the performance. An important result derived from this study is that routing over bundled links is insensible to the scheme for aggregating weights of component links. In our study, almost all virtual links have the shortest length both in a low load regime and in a medium load regime. In the former case, this is because there are several available shortest paths between each source-destination pair in the test topology; in the latter case, due to the uniform traffic load neither shorter nor longer paths are available. Therefore, when computing a path the aggregation of the component link weights based on the mean-cost scheme can be used; when routing the traffic over the aggregated links, the choice of labels on each of the links can be can be random. In general, this result may depend on the topology and on traffic. However it can be generalized for the topologies that have relatively short maximum diameter - here 5 - and several available alternative shortest paths for any selected $(CE - CE)$. Of course, for some extremely heterogeneous traffic patterns, there would be a need to constrain the length of the LSC LSPs that are bundled together.

In the scope of this work, which major focus was on the dynamic traffic routing and topology engineering for infrastructure services, we did not propose any particular new method for the initial configuration of the visibility attribute, i.e., we assumed that some of the existing virtual topology design heuristics could be used. Furthermore, we assumed that the dynamic adaptation of visibility could be based on measurements. We demonstrated this approach in one of the simulation experiments in which the number of flow-through LSPs at the hybrid nodes was used to select a node which visibility of grooming resources should be changed. It is important to notice that changing resource visibility does not cause any kind of re-configuration actions or traffic re-routing in a network, but only an update of the TE information.

Finally, the results presented in this thesis illustrated that GXP architecture provides manifold performance benefits. By introducing the GXP nodes into the topology, new routing paths can be created. We have shown that by introducing grooming capability in the GXP nodes a significant performance improvement can be achieved, even without any grooming capability deployed within the domain's nodes. On the other hand, we are aware that policies play a significant role on how domains would use bandwidth from other domains. For instance, we implemented a simple policy where an external capacity is used for intra-domain traffic only, in case the request would be otherwise blocked, and it resulted in improvements. While we are aware that in order to assess the full potential of the proposed architecture we need to go beyond some simplifying assumptions of the presented performance study, we are convinced of significant potential benefits of the GXP-based architecture.

Based on the concepts proposed in this thesis further work may address several open issues and promising directions. First, in the context of separation between *controlling visibility* and dynamically *using visible resources*, further research may focus on methods for configuring and adapting resource visibility based on service requirements and measurements. The modeling and classification of service requirements should be accounted for when making decisions which resources to offer to which services. Coordination between the complementary roles that the methods for the visibility control and dynamic routing play within the closed control loop could be a challenging further step towards autonomic management. An approach towards creating adaptable methods based on atomic functional blocks might also be evaluated.

The control-plane issues related to distributed path computation and signaling between different control plane instances, which were not in the scope of this work, also require attention and further study. These include the coordination of routing and signaling, which is necessary when routing is preceded or followed by several topology extension actions, particularly in the presence of incomplete advertised information. In the context of federated architecture, the issues of collaborative resource sharing among different domains, based on the visibility concept, need to be further addressed.

With the proposed federated architecture and infrastructure service model, this thesis laid foundations for virtualizations in the optical layer. As some concepts proposed here have already found their way into real-world implementations (GXP), we believe that further on the horizon, also the visibility attribute - a key to coordination and sharing of resources - can be successfully incorporated and extended within the emerging technologies, such as autonomic network control.

Appendix

Appendix A List of Abbreviations

ADM	Add/Drop Multiplexers
AMS-IX	Amsterdam Exchange Point
APN	Articulated Private Networks
AS	Autonomous System
ASON	Automatic Switched Optical Network
ASTN	Automatic Switched Transport Network
ATM	Asynchronous Transfer Mode
B	Border Node
BGP	Border Gateway Protocol
BOD	Bandwidth-On-Demand
CAC	Connection Admission Scheme
CE	Customer Edge
CLA	Control Level Agreement
CMB	Combined RToE Method
CO	Connection Oriented
CP	Control Plane
CPI	Control Plane Instance
CREL	Conditional Release when Idle ToE
CR-LDP	Constrained Routing Label Distribution Protocol
DES	Discrete-Event Simulation
DXC	Digital Cross-Connect
E-NNI	External NNI
EPL	Ethernet Private Line

ERO	Explicit Route Object
FA	Forwarding Adjacency
FCAPS	Fault, Configuration, Accounting, Performance and Security Management
FDDI	Fiber Distributed Data Interface
FSC	Fibre Switching Capable
GENI	Global Environment for Networking Innovations
GLTDA	Greedy Logical Topology Design Algorithm
GMPLS	Generalized Multiprotocol Label Switching
GPRS	General Packet Radio Service
GRWA	Grooming RWA
GRX	GPRS Roaming Exchange Points
GVPXC	Generalized Virtual Private Cross-Connect
GXP	GMPLS Exchange Point
HLDA	Heuristic Logical Topology Design Algorithm
IACD	Interface Adaptation Capability Adapter
IBX	Internet Business Exchange
IETF	Internet Engineering Task Force
I-NNI	Internal NNI
IP	Internet Protocol
ISCD	Interface Switching Capability Descriptor
ISIS	Inter-System to Inter-System
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
IX	Internet Exchange
IXP	Internet Exchange Point
L1 VPN	Layer 1 VPN
L2SC	Layer 2 Switching Capable

LAN	Local Area Networks
LBL	Layer-By-Layer RToE Method
LCAS	Link Capacity Adjustment Scheme
LD	Lightpath Demand
LMP	Link Management Protocol
LRM	Link Resource Manager
LSC	Lambda Switching Capable
LSP	Label Switched Path
LSR	Lambda Switching Router
MAN	Metropolitan Area Network
MILP	Mixed Integer Linear Programming
MLDA	Minimum Delay Logical Topology Design Algorithm
MPE	Multi-Provider Edge
MPLS	Multiprotocol Label Switching
NGI	Next Generation Internet
NNI	Network to Network Interface
NP	Non-Polynomially
NREL	Never Release ToE Method
NREN	National Research Experimental Networks
NSF	National Scientific Foundation
OADM	Optical Add
OBS	Optical Burst Switching
OCh	Optical Channel
ODU	Optical Data Unit
OIF	Optical Internet Forum
OPEX	Operational Expenditures
OPS	Optical Packet Switching
OPU	Optical Payload Unit

OSI	Open Systems Interconnection
OSPF-TE	Open Shortest Path First extended for Traffic Engineering
OTDM	Optical Time Division Multiplexing
OTN	Optical Transport Network
OTU	Optical Transport Unit
OXC	Optical Cross-Connect
P	Provider Domain
PCE	Path Computation Element
PE	Provider Edge
PSC	Packet Switching Capable
PVC	Permanent Virtual Circuits
PXC	Photonic Cross-Connect
QoS	Quality-of-Service
RB	Regional Broadcaster
RDB	Resource Database
RELI	Release When Idle ToE Method
RSVP-TE	Reservation Protocol with Traffic Engineering
RToE	Routing and Topology Engineering
RWA	Routing and Wavelength Assignment
SDH	Synchronous Digital Hierarchy
SH-TG	Single-hop traffic grooming
SLA	Service Level Agreement
SLS	Service Level Specification
SN	Subnetwork
SNP	Subnetwork Point
SNPP	Subnetwok Point Pool
SONET	Synchronous Optical Network
SRLG	Shared Risk Link Group

SVG	Service Visibility Graph
TDM	Time Division Multiplexing
TE	Traffic Engineering
TILDA	Topology Independent Logical Topology Design Algorithm
ToE	Topology Engineering
UCLP	User-Controlled Lightpath Architecture
UNI	User-to-Network Interface
VCAT	Virtual Concatenation
VPN	Virtual Private Network
WADM	Wavelength Add
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
WRN	Wavelength Routed Network
XP	Exchange Point

Bibliography

[Alan04] W. Alanqar, A. Jukan, Extending end-to-end optical service provisioning and restoration in carrier networks: opportunities, issues, and challenges, IEEE Communications Magazine, vol. 42, no. 1, Jan 2004, pp. 52–60.

[Alan06] W. Alanqar, V. Frost, A Service Profile-Aware Control Plane with a Comparison to IETF & ITU Approaches, IEEE International Conference on Communications, ICC'06, vol. 6, June 2006, pp. 2725-2730.

[Alic05] M. Alicherry, C. Phadke, V. Poosala, Delay distributed VCAT for efficient data-optical transport, Optical Fiber Communication Conference, OFC/NFOEC, vol.4, 6-11 March 2005.

[Ande05] L. Andersson, T. Madsen, Provider Provisioned Virtual Private Network (VPN) Terminology, Internet RFC 4026, March 2005.

[Ashw03] P. Ashwood-Smith, Ed., L. Berger, Ed., Generalized Multi-Protocol Label Switching (GMPLS) Signaling Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions, RFC3472, Jan. 2003.

[Awdu99] D. Awduche et al, Requirements for Traffic Engineering over MPLS, Internet RFC 2702, Sept. 1999

[Awdu98] D.O. Awduche, J. Agogbua, and J. McManus, An Approach to Optimal Peering Between Autonomous Systems in the Internet, 7th International Conference on Computer Communications and Networks, Oct. 1998, pp. 346–351.

[Awdu01] D.O. Awdusche, Y. Rekhter, Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects, IEEE Communications Magazine, vol.39, no.3, March 2001, pp. 111–116.

[Bagu04] A. B. Bagula, M. Botha, A. E. Krzesinski, Online traffic engineering: the least interference optimization algorithm, IEEE International Conference on Communications, vol. 2, 20-24 June 2004, pp. 1232–1236.

[Bana97] D. Banerjee and B. Mukherjee, Wavelength-routed optical-networks: linear formulation, resource budgeting trade-offs, and a reconfiguration study, IEEE INFOCOM '97, pp. 269-276, Apr. 1997.

[Bane03b] A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, Y. Rekhter, Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques, IEEE Communications Magazine, vol.39, no.7, July 2001, pp. 144–151.

[Bane03a] A. Banerjee, L. Drake, L. Lang, B. Turner, L. Berger, K. Kompella, Y. Rekhter, Generalized multiprotocol label switching: an overview of routing and management enhancements, IEEE Communication Magazine, vol.39, no.1, Jan. 2001, pp. 144–150.

[Baro01] S. Baroni, Routing and wavelength allocation in WDM optical networks. PhD thesis, University College London, May 1998.

[Barr96] R. A. Barry and P. A. Humblet, Models of blocking probability in all-optical networks with and without wavelength changers,IEEE Journal on Selected Areas in Communications, vol. 14, no. 5, June 1996, pp. 858–867.

[Berg03b] L. Berger, Ed., Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description, RFC3471, Jan. 2003.

[Berg03a] L. Berger, Ed., Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions, RFC3473, Jan. 2003.

[Bern05] G. Bernstein, D. Caviglia, and R. Rabbat, Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS), Oct. 2005, Internet Draft.

[Bern06] G. Bernstein, D. Caviglia, R. Rabbat, H. Van Helvoort, VCAT-LCAS in a clamshell, IEEE Comm. Magazine, vol. 44, no. 5, May 2006, pp. 34–36.

[Bern02] G. Bernstein, V. Sharma, L. Ong, Interdomain Optical Routing, Journal of Optical Networking, vol.1, no.2, Feb. 2002, pp. 80-92.

[Berr00] R. Berry, E. H. Modiano, Reducing electronic multiplexing costs in SONET/WDM rings with dynamically changing traffic, IEEE Journal on Selected Areas in Communications, vol. 18, no. 10, Oct. 2000, pp. 1961–1971.

[Bert98] D. P. Bertsekas, Network Optimization: Continuous and Discrete Models (Optimization, Computation, and Control), Athena Scientific, 1998.

[Bhan05] S. Bhandari, E. K. Park, Dynamic reconfiguration for optical network, 14th International Conference onComputer Communications and Networks, ICCCN 2005, 17–19 Oct. 2005, pp. 243–248.

[Birm96] A. Birman, Computing Approximate Blocking Probabilities for a Class of All-Optical Networks, IEEE Journal on Selected Areas in Communications, vol. 13, June 1996, pp. 852–857.

[Bisw03] S. Biswas, D. Saha, N. Mandal, Intercarrier bandwidth exchange: an engineering framework, IEEE Communications Magazine, vol. 41, no. 1, Jan. 2003, pp. 130–138.

[Blyt01] K. J. Blyth and A. R. J. Cook, Designing a GPRS Roaming Exchange, Conference on 3G Mobile Communication Technologies, 26-28 March 2001, pp. 201–205.

[Bout02] R. Boutaba, W. Szeto and Y. Iraqi, DORA: Efficient Routing for MPLS Traffic Engineering, Journal of Network and Systems Management, vol. 10, no. 3, Sept. 2002, pp. 309–325.

[Cava05] C. Cavazzoni, A. Di Giglio, G. Edwall, G. Eilenberger, G. Ferraris, H. Haunstein, S. Herbst, M. Jaeger, G. Lehmann, J. F. Lobo, A. Manzalini, S. Santoni, M. Schiano, Achievements of the EU NOBEL project, 31st European Conference on Optical Communication, ECOC'05, vol. 5, 25-29 Sept. 2005, pp. 37–40.

[Ceru05] I. Cerutti, A. Fumagalli, Traffic grooming in static wavelength division multiplexing networks, IEEE Comm. Magazine, vol. 43, no. 1, Jan. 2005, pp. 101–107.

[Ceru03] I. Cerutti, A. Fumagalli, Multi-Rate and Multi-Hop Hybrid Optical Networks with Arbitrary Topology, Workshop on High Performance Switching and Routing, 2003.

[Chan98] V. W. S. Chan, K. L. Hall, E. Modiano, K. A. Rauschenbach, Architectures and technologies for high-speed optical data networks, Journal of Lightwave Technology, vol.16, no.12, Dec. 1998, pp. 2146–2168.

[Chel00] G. Cheliotis, Bandwidth trading in the real world: Findings and implications for commodities brokerage, In 3rd Berlin Internet Economics Workshop, pp. 26–27.

[Chen05] B. Chen, G. N. Rouskas, R. Dutta, A framework for hierarchical traffic grooming in WDM networks of general topology, 2nd International Conference on Broadband Networks, 2005, 3-7 Oct. 2005, pp. 155–164.

[Che05d] B. Chen, G. N. Rouskas, R. Dutta, A framework for hierarchical traffic grooming in WDM networks of general topology, 2nd International Conference on Broadband Networks, vol. 1, 3-7 Oct. 2005, pp. 155–164.

[Che05a] B. Chen, R. Dutta, G. N. Rouskas, On the application of K-center algorithms to hierarchical traffic grooming, 2nd International Conference on Broadband Networks, vol. 2, 3-7 Oct. 2005, pp. 1218–1224.

[Chiu00] A. L. Chiu, E. H. Modiano, Traffic grooming algorithms for reducing electronic multiplexing costs in WDM ring networks, IEEE Journal of Lightwave Technology, vol. 18, no. 1, Jan. 2000,pp. 2–12.

[Chla92] I. Chlamtac, A. Ganz, and G. Karmi, Lightpath communications: an approach to high bandwidth optical WANS, IEEE Transactions on Communications, vol. 40, no. 7, 1992, pp. 1171–1182.

[Chla93] I. Chlamtac, A. Ganz, and G. Karmi, Lightnets:topologies for high-speed optical networks, IEEE Journal of Lightwave Technology, vol. 11, no. 3, May/June 1993, pp. 951–961.

[Cink00] T. Cinkler, D. Marx, C. P. Larsen, D. Fogaras, Heuristic algorithms for joint configuration of the optical and electrical layer in multi-hop wavelength routing networks, Nineteenth Annual Joint Conference of the IEEE Computer and

Communications Societies, INFOCOM'00, vol. 2, 26-30 March 2000, pp. 1000–1009.

[InvH02] Computer Letter published by Technology Parners, The Price Is Right, May 20, 2002, http://www.invisiblehand.net/index.php?type=pc&pc_id=11.

[Cplex] CPLEX: An optimizer by LOG Inc, url : www.ilog.com.

[inec06] Declaration of Open Networks, www.i-nec.com, Nov. 2006

[Dijk04] F. Dijkstra and C. de Laat, Optical Exchanges, In Proc. of GRIDNETS, 2004.

[Dijk59] E. W. Dijkstra, A Note on Two Problems in Connexion with Graphs, Numerische Mathematik, vol. 1, 1959, pp. 269-271.

[Drez95] Z. Drezner, Facility Location: A Survey of Applications and Methods, Springer-Verlag New York, Inc, 1995.

[Dutt05] R. Dutta and G. N. Rouskas, Traffic grooming in WDM networks Past and future, IEEE Network, vol. 16, no. 6, Nov/Dec 2002, pp. 46-56.

[Dutt02] R. Dutta and G. N. Rouskas, On optimal traffic grooming in WDM rings, IEEE Journal on Selected Areas in Communications, vol. 20, no. 1, Jan. 2002, pp. 110–121.

[Dutt01] R. Dutta, G. Rouskas, Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks, Optical Network Magazine, Springer, vol. 1, no. 1, Jan. 2000, pp. 73–89.

[Farr06] A. Farrel, J. P. Vasseur, J. Ash. A Path Computation Element (PCE)-Based Architecture, Internet RFC 4655, Aug 2006.

[Fawa04] W. Fawaz, B. Daheb, O. Audouin, M. Du-Pond, G. Pujolle, Service level agreement and provisioning in optical networks, IEEE Communications Magazine, vol. 42, no. 1, Jan 2004, pp. 36–43.

[Fois05] H. M. Foisel, C. Gerlach, A. Gladisch, S. Szuppa, A. Weber, Global seamless network demonstrator: a comprehensive ASON/GMPLS testbed, IEEE Communications Magazine, vol. 43, no. 11, Nov. 2005, pp. S34–S39.

[G7042] ITU-T Rec. G7042, Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenated Signals, Feb. 2004

[G7043] ITU-T Rec. G7043, Virtual Concatenation of Plesiochronous Digital Hierarchy (PDH) Signals, July 2004.

[G709] ITU-T G709, "Interfaces for the Optical Transport Network (OTN)," Mar. 2003.

[G803] ITU-T Rec. G803 (03-2000) Architecture of transport networks based on the synchronous digital hierarchy (SDH).

[G805] ITU-T Rec. G805, Generic functional architecture of transport networks, March 2000.

[G807] ITU-T Rec. G807, Requirements for Automatically Switched Transport Networks, July 2001.

[G8080] ITU-T Rec. G8080/Y. 130411 (2001) Architecture for the Automatic Switched Optical Networks (ASON).

[G872] ITU-T Rec. G872 (11-2001) Architecture of optical transport networks.

[Genc03] A. Gencata and B. Mukherjee, Virtual-topology adaptation for WDM mesh networks under dynamic traffic, IEEE/ACM Transactions on Networking, vol. 11, no. 2, April 2003, pp. 236–247.

[geni06] NSF initative. GENI, http://www.nsf.gov/cise/geni/, 2006.

[Gers99] O. Gerstel, P. Lin, G. Sasaki, Combined WDM and SONET network design, Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM'99, vol. 2, 21-25 March 1999, pp. 734–743.

[Gers00] O. Gerstel, R. Ramaswami, G. H. Sasaki, Cost-effective traffic grooming in WDM rings, IEEE/ACM Transactions on Networking, vol. 8, no. 5, Oct. 2000, pp. 618–630.

[Wang01] J. Wang, W. Cho, R. Vemuri, B. Mukherjee, Improved approaches for cost-effective traffic grooming in WDM ring networks: ILP formulations and single-hop and multihop connections, IEEE Journal of Lightwave Technology, vol. 19, no. 11, Nov. 2001, pp. 1645–1653.

[Ghan03] N. Ghani, S. Dixit, and Ti-Shiang Wang, On IP-WDM integration: a retrospective, IEEE Communications Magazine, vol. 41, no. 9, Sept. 2003, pp. 42–45.

[Gibb04] S. Gibbard, Economics of peering, Packet Clearing House/Steve Gibbard Consulting, October, 2004, www.pch.net/resources/papers/Gibbard-peering-economics.pdf

[Gira90] A. Girard, Routing and dimensioning circuit-switched networks, Addison Wesley, 1990, pp 556.

[Habi06] I. W. Habib, Q. Song, Z. Li, N. S. V. Rao, Deployment of the GMPLS control plane for grid applications in experimental high-performance networks, IEEE Communications Magazine, vol. 44, no. 3, March 2006, pp. 65–73.

[Herr05] K. Herrmann, G. Muehl, K. Geihs, Self Management: The Solution to Complexity or Just Another Problem, IEEE Distributed Systems Online, vol. 6, no. 1, Jan. 2005, pp. 17

[HuJQ04] J. Q. Hu, B. Leida, Traffic grooming, routing, and wavelength assignment in optical WDM mesh networks, Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, INFOCOM'04, vol. 1, 7-11 March 2004.

[Huan06] S. Huang, R. Dutta, G. N. Rouskas, Traffic grooming in path, star, and tree networks: complexity, bounds, and algorithms, IEEE Journal on Selected Areas in Communications, vol.24, no.4, Apr 2006.

[Jain91] R. Jain, The Art of Computer Science Performance Analysis: Techniques for Experimental Design, Measurment, Simulation and Modeling, Wiley-Interscience, New York, N.Y., April 1991.

[Jone05] J. D. Jones, L. Ong, M. Lazer, Interoperability update: dynamic ethernet services via intelligent optical networks, IEEE Communications Magazine, vol. 43, no. 11, Nov. 2005, pp. S4–S10.

[Jone04] J. D. Jones, L. Ong, M. A. Lazer, Creating an intelligent optical network worldwide interoperability demonstration, IEEE Communications Magazine, vol. 42, no. 11, Nov. 2004, pp. S4-11.

[JJue05] J. P. Jue and V. M. Vokkarane, Optical Burst Switched Networks, Springer, 2005.

[Juka00] A. Jukan, QoS-based Wavelength Routing in Multi-Service WDM Networks, PhD thesis, Vienna University of Technology, 2000.

[Kar00] K. Kar, M. Kodialam, and T. V. Lakshman, Minimum interference routing of bandwidth guaranteed tunnels with MPLS traffic engineering applications, IEEE Journal on Selected Areas in Communications: Quality of Service in the Internet, vol. 18, no. 12, Dec. 2000, pp. 921–940.

[Karm05] G. Karmous-Edwards, Global e-Science collaboration, Computing in Science and Engineering, vol. 7, no. 2, March-April 2005, pp. 67–74.

[Katz03] D. Katz, D. Yeung, and K. Kompella, Traffic engineering extensions to OSPF version 2. , December 2003, RFC3630, Proposed Standard.

[Klon05] D. Klonidis, C. T. Politi, R. Nejabati, M. J. O'Mahony, D. Simeonidou, OPSnet: design and demonstration of an asynchronous high-speed optical packet switch, Journal of Lightwave Technology, vol.23, no.10, Oct. 2005, pp. 2914–2925.

[Komh05] K. Kompella and Y. Rekhter, Label switched paths (LSP) hierarchy with generalized multi-protocol label switching (GMPLS) traffic engineering (TE), Internet RFC 4206, Oct 2005.

[Komp05b] K. Kompella, Ed., Y. Rekhter, Ed., Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS), RFC4205, Oct. 2005.

[Komp05a] K. Kompella, Ed., Y. Rekhter, Ed., OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS), RFC4203, Oct. 2005.

[Komp05e] K. Kompella, Ed., Y. Rekhter, Ed., Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS), RFC4202, Oct. 2005.

[Komp02] K. Kompella, Y. Rekhter, J. P. Vasseur, and T. W. Chung, Multi-area mpls traffic engineering, 2002, draft-kompella-mpls-multiarea-te-03. txt Expiration December 2002.

[Komb05] K. Kompella, Y. Rekhter, L. Berger, Link Bundling in MPLS Traffic Engineering (TE), RFC4201, Oct. 2005

[Komu03] K. Kompella, Y. Rekhter, Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE), RFC3477, Jan. 2003

[Kir05a] J. Kuri, M. Gagnaire, N. Puech, On the resource efficiency of virtual concatenation in next-generation SDH networks,Workshop on High Performance Switching and Routing, HPSR'05, 12-14 May 2005, pp. 515–519.

[Lang05] J. Lang, Ed., Link Management Protocol (LMP), RFC4204, Oct. 2005.

[Laza98] A. A. Lazar, N. Semret, Design, Analysis and Simulation of the Progressive Second Price Auction for Network Bandwidth Sharing", Columbia University, 1998.

[DeLe06] M. De Leenheer, P. Thysebaert, B. Volckaert, F. De Turck, B. Dhoedt, P. Demeester, D. Simeonidou, R. Nejabati, G. Zervas, D. Klonidis, M. J. O'Mahony, A view on enabling-consumer oriented grids through optical burst switching, IEEE Communications Magazine, vol.44, no.3, March 2006, pp. 124–131.

[Lehm06] T. Lehman, J. Sobieski, B. Jabbari, DRAGON: a framework for service provisioning in heterogeneous grid networks, IEEE Communications Magazine, vol. 44, no. 3, March 2006, pp. 84–90.

[Leon00] E. Leonardi, M. Mellia, M. A. Marson, Algorithms for the Topology Design in WDM All-optical Networks", Optical Network Magazine, Springer, vol. 1, no. 1, Jan. 2000, pp. 35–46.

[Lian89] S. C. Liang, J. R. Yee, A gateway allocation algorithm for interconnecting existing data networks, IEEE INFOCOM'89, vol.2, April 1989, pp. 468–473.

[Mann06] E. Mannie and D. Papadimitriou, Ed. , Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control, RFC 4606, Aug. 2006

[Mann06f] E. Mannie, D. Papadimitriou, Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control, RFC4606, Aug. 2006.

[Mann04] E. Mannie, Ed. , Generalized Multi-Protocol Label Switching (GMPLS) Architecture, RFC 3945, Oct. 2004

[Metz01] C. Metz, Interconnecting ISP networks, IEEE Internet Computing, vol. 5, no. 2, March/April 2001, pp. 74–81.

[Mukh96] B. Mukherjee, D. Banerjee, S. Ramamurthy, and A. Mukherjee, Some principles for designing a wide-area WDM optical network, IEEE/ACM Transactions on Networking, vol. 4, no. 5, 1996, pp. 684–696.

[Mukh97] B. Mukherjee, Optical Communication Networks, McGraw-Hill, New York, 1997.

[Muno05] R. Munoz, C. Pinart, R. Martinez, J. Sorribes, G. Junyent, ADRENALINE testbed: user management of lightpaths over intelligent WDM networks through GMPLS and XML, First International Conference onTestbeds and Research Infrastructures for the Development of Networks and Communities, Tridentcom'05, 23-25 Feb. 2005, pp. 252–261.

[Nada07] T. Nadeau, Ed. , A. Farrel, Ed. , Generalized Multiprotocol Label Switching (GMPLS) Traffic Engineering MIB, Internet RFC 4802, Feb. 2007

[Naka02] I. Nakagawa, H. Esaki, and K. Nagami, A design of a next generation IX using MPLS technology, Symposium on Applications and the Internet, SAINT, 28 Jan. -1 Feb. 2002, pp. 238–245.

[OMah01] M. J. O'Mahony, D. Simeonidou, D. K. Hunter, A. Tzanakaki, The application of optical packet switching in future communication networks, IEEE Communications Magazine, vol.39, no.3, March 2001, pp. 128–135.

[Papm05] D. Papadimitriou et al. , Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi- Region Networks (MRN), IETF Internet Draft, Feb. 2005

[Papa06] D. Papadimitriou, B. Berde, R. Martinez, J. G. Ordas, R. Theillaud, S. Verbrugge, Generalized Multi-Protocol Label Switching (GMPLS) Unified Control Plane Validation, IEEE International Conference on Communications, ICC'06, vol. 6, June 2006, pp. 2717–2724.

[Papa06f] D. Papadimitriou, Ed., Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 OTN Control, RFC4328, Jan. 2006.

[Papr07] D. Papadimitriou, OSPFv2 Routing Protocols Extensions for ASON Routing, Internet Draft, Mar. 2007

[Pap05a] D. Papadimitriou, Z. Lin, J. Drake, J. Ash, A. Farrel, and L. Ong, Requirements for generalized MPLS (GMPLS) signaling usage and extensions for automatically switched optical network (ASON), Jul 2005, Internet RFC 4139.

[Pasq05] S. Pasqualini, A. Kirstadter, A. Iselt, R. Chahine, S. Verbrugge, D. Colle, M. Pickavet, P. Demeester, Influence of GMPLS on network providers' operational expenditures: a quantitative study, IEEE Communications Magazine, vol. 43, no. 7, July 2005, pp. 28–38.

[Pina06] C. Pinart and G. J. Giralt, On managing optical services in future control-plane-enabled IP/WDM networks, Journal of Lightwave Technology, vol. 23, no. 10, Oct. 2005, pp. 2868–2876.

[Prev06] V. Prevelakis, A. Jukan, How to buy a network: trading of resources within the physical layer, IEEE Communications Magazine, vol. 44, no. 12, Dec. 2006, pp. 94 - 102.

[Qiao06] Chunming Qiao, Wei Wei, Xin Liu, Extending generalized multiprotocol label switching (GMPLS) for polymorphous, agile, and transparent optical networks (PATON), IEEE Communications Magazine, vol.44, no.12, Dec. 2006, pp. 104–114.

[Rama00] B. Ramamurthy and A. Ramakrishnan, Virtual topology reconfiguration of wavelength routed optical WDM networks, IEEE GLOBECOM, 2000, pp. 1269-1275.

[Rama02] B. Ramamurthy, A. Ramakrishnan, Design of Virtual Private Networks (VPNs) over Optical Wavelength Division Multiplexed (WDM) Networks, Optical Network Magazine, Springer, vol. 3, no. 1, Jan. /Feb 2002, pp. 59–67.

[Rama96] R. Ramaswami and K. N. Sivarajan, Design of logical topologies for wavelength routed optical networks, IEEE Journal on Selected Areas in Communications, vol. 14, no. 5, June 1996, pp. 840–851.

[RAMA95] R. Ramaswami, K.N. Sivarajan, Routing and wavelength assignment in all-optical networks, IEEE/ACM Transactions on Networking, vol.3, no.5, Oct. 1995, pp. 489–500.

[RAMA06] R. Ramaswami, Optical networking technologies: what worked and what didn't, IEEE Communications Magazine, vol. 44, no. 9, Sept. 2006

[REKH06] Y. Rekhter, Ed. , T. Li, Ed. , and S. Hares, Ed. A Border Gateway Protocol 4 (BGP-4), January 2006.

[ROSE06] E. Rosen, Y. Rekhter, BGP/MPLS IP Virtual Private Networks (VPNs), RFC4364, February 2006.

[ROUS01] G. Rouskas, Routing and wavelength assignment in optical WD Mnetworks. In John Proakis, editor, Wiley Encyclopedia of Telecommunications. John Wiley & Sons, 2001.

[SABE03] R. Sabella, M. Settembre, G. Oriolo, F. Razza, F. Ferlito, and G. Conte, A multilayer solution for path provisioning in new-generation optical/MPLS networks, Journal of Lightwave Technology, vol. 21, no. 5, May 2003, pp. 1141–1155.

[SAHA05] M. Saha, I. Sengupta, A Genetic Algorithm Based Approach for Static Virtual Topology Design in Optical Networks, IEEE Network, vol. 20, no. 4, July-Aug. 2006, pp. 6–15.

[SANC03] S. Sanchez-Lopez, J. Sole-Pareta, J. Comellas, J. Soldatos, G. Kylafas, M. Jaeger, PNNI-based control plane for automatically switched optical networks, Journal of Lightwave Technology, vol.21, no.11, Nov. 2003, pp. 2673–2682.

[SEMR00] N. Semret, R. R. Liao, A. T. Campbell, A. A. Lazar, Peering and Provisioning of Differentiated Internet Services, INFOCOM 2000, pp. 414–420

[SENG03] S. Sengupta, V. Kumar, D. Saha, Switched optical backbone for cost-effective scalable core IP networks, IEEE Communications Magazine, vol. 41, no. 6, June 2003, pp. 60–70.

[SIME05] D. Simeonidou, R. Nejabati, G. Zervas, D. Klonidis, A. Tzanakaki, M. J. O'Mahony, Dynamic optical-network architectures and technologies for existing and emerging grid services, Journal of Lightwave Technology, vol.23, no.10, Oct. 2005, pp. 3347–3357.

[SREE01] N. Sreenath, C. S. R. Murthy, B. H. Gurucharan, and G. Mohan, A two-stage approach for virtual topology reconfiguration of WDM optical networks, Opt. Networks Mag., pp. 58-71, May 2001.

[STAR06] B. St Arnaud and at al, CA*net User Controlled Lightpah, UCLP v2, http://grid2. canarie. ca/wiki/index. php/UCLPv2.

[STAR03] B. St Arnaud, J. Wu, B. Kalali, Customer-controlled and -managed optical networks, IEEE Journal of Lightwave Technology, vol. 21, no. 11, Nov. 2003, pp. 2804–2810.

[STAR01] B. St Arnaud, M. Weir, and J. Coulter, BGP Optical Switches and Lightpath Route Arbiter, Optical Networks Magazine, Mar/Apr 2001.

[SUBR97] S. Subramaniam, R.A. Barry, Wavelength assignment in fixed routing WDM networks. IEEE International Conference on Communications, 1997, pp. 406-410, 1997.

[SWAL05] G. Swallow, J. Drake, H. Ishimatsu, Y. Rekhter, Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model, Internet RFC 4208, October 2005.

[SZIG05] J. Szigeti, I. Ballok, T. Cinkler, Efficiency of information update strategies for automatically switched multi-domain optical networks, 7th International Conference Transparent Optical Networks, vol. 1, 3-7 July 2005, pp. 445–454.

[SZIG04] J. Szigeti, J. Tapolcai, T. Cinkler, T. Henk, G. Sallai, Stalled information based routing in multidomain multilayer networks,11th International Telecommunications Network Strategy and Planning Symposium, NETWORKS'04, 13-16 June 2004, pp. 297–302.

[T1.105] ANSI. T1. 105, Synchronous optical network (SONET): Basic description including multiplex structure, rates, and formats, Oct. 2000

[Take05] T. Takeda, D. Brungard, D. Papadimitriou, H. Ould-Brahim, Layer 1 VPNs: Driving forces and realization by GMPLS, IEEE Comm. Magazine, Vol. 43 No. 7, Jul 2005, pp. 60–67.

[Take04] T. Takeda, I. Inoue, R. Aubin, M. Carugi, Layer 1 Virtual Private Networks: Service Concepts, Architecture Requirements, and Related Advances, IEEE Communications Magazine, Vol. 42, No. 6, Jun 2004, pp. 132–138.

[Tomi07] S. Tomic and A. Jukan, Performance Analysis of Infrastructure Service Provision with GMPLS-based Traffic Engineering, Special Issue on Multi-Layer Traffic Engineering, JSAC, 2007.

[Tom02b] S. Tomic and A. Jukan, MPFI: The multi-provider network federation interface for interconnected optical networks, IEEE Global Telecommunications Conference (GLOBECOM'02), vol. 3, 17-21 Nov. 2002, pp. 2365–2369.

[Tomi05] S. Tomic and A. Jukan, GMPLS-based exchange points: architecture and functionality, Chapter VIII, in Emerging Optical Network Technologies Architectures, Protocols and Performance, Editors Krishna M. Sivalingam and Suresh Subramaniam, Springer, 2005.

[Tomi04c] S. Tomic and A. Jukan, Extensible Service Provision in Grid Networks: A Case For Resource Visibility and Inter-Domain Exchange, First International Workshop on Networks for Grid Applications, San Jose, USA, 29 October 2004.

[Tom03a] S. Tomic, A. Jukan, Dynamic Provisioning over Multi-provider Interconnected GMPLS-enabled Networks, European Conference on Optical Network Design and Modeling ONDM 2003, Feb. 2003, pp. 365-383.

[Tom03b] S. Tomic, A. Jukan, GMPLS-based Exchange Points: Architecture and Functionality, Workshop on High Performance Switching and Routing, HPSR, 24-27 June 2003, pp. 245–249.

[Tom02a] S. Tomic, A. Jukan, Policy-based lightpath provisioning over federated WDM network domains, Optical Fiber Communication Conference and Exhibit, OFC'02, 17-22 March 2002, pp. 499–501.

[Tomi02c] S. Tomic, A. Jukan, Optical Virtual Private Network Services in the Multi-Provider Network Environment"; IP over WDM international conference, IPoW'02, Paris, France; June 2002; pp. 1–19.

[Tomi04b] S. Tomic, B. Statovci-Halimi, A. Halimi, W. Mllner, J. Frhwirth, *ASON and GMPLS–Overview and Comparison*, Photonic Network Communications, vol.7, no.2, 2004, pp. 111–130.

[Tomi04] S. Tomic, Issues of resource management in two-layer GMPLS networks with virtual network services, IEEE Global Telecommunications Conference, GLOBECOM '04. vol. 3, 29 Nov. -3 Dec. 2004, pp. 1803–1807.

[Tomi03c] S. Tomic, Issues of Two-Layers Routing in GMPLS Networks with Virtual Network Services, 8th European Conference on Networks & Optical Communications 2003 (NOC'2003), Vienna, Austria, July 2003, pp. 114-121.

[Trav06] F. Travostino, J. Mambretti, and G. Karmous-Edwards (Editors), Grid Networks: Enabling Grids with Advanced Communication Technology, Wiley Publishers, Sept. 2006

[Veer06] M. Veeraraghavan, X. Zheng, Z. Huang, On the use of connection-oriented networks to support grid computing, IEEE Communications Magazine, vol. 44, no. 3, March 2006, pp. 118–123.

[Vigo05] M. Vigoureux, B. Berde, L. Andersson, T. Cinkler, L. Levrau, M. Ondata, D. Colle, J. Fernandez-Palacios, M. Jager, Multilayer traffic engineering for GMPLS-enabled networks, IEEE Communications Magazine, vol. 43, no. 7, July 2005, pp. 44–50.

[Wang02] J. Wang and B. Mukherjee, Interconnected WDM ring networks: Strategies for interconnection and traffic grooming, Optical Networks Magazine, Kluwer, vol. 3, no. 5, Sept. /Oct. 2002, pp. 1–4.

[WuJ04] J. Wu, H. Zhang, S. Campbell, M. Savoie, G. V. Bochmann, B. St. Arnaud, A grid oriented lightpath provisioning system, IEEE Global Telecommunications Conference Workshops, 29 Nov. -3 Dec. 2004, pp. 395–399.

[XuBa02] Y. Xu, A. Basu, and Y. Xue, A BGP/GMPLS solution for inter-domain optical networking, work in progress draft-xu-bgp-gmpls-02. txt, Expiration Date: Dec 2002.

[Yang03] X. Yang, B. Ramamurthy, Inter-domain dynamic routing in multi-layer optical transport networks, Global Telecommunications Conference, GLOBECOM'03, vol.5, 2003, pp. 2623–2627.

[Yao05c] W. Yao, B. Ramamurthy, A Link Bundled Auxiliary Graph Model for Constrained Dynamic Traffic Grooming in WDM Mesh Networks, IEEE Journal on Selected Areas in Communications, vol. 23, no. 8, Aug. 2005, pp. 1542–1555.

[Yate97] J. Yates, Performance Analysis of Dynamically-Reconfigurable Wavelength-Division Multiplexed Networks, Dissertation, The University of Melbourne, 1997.

[Yoo01] Myungsik Yoo, Chunming Qiao, S. Dixit, Optical burst switching for service differentiation in the next-generation optical Internet, IEEE Communications Magazine, vol.39, no.2, Feb. 2001, pp. 98–104.

[Zang00] 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, Kluwer, vol. 1, no. 1, Jan 2000, pp. 47–60.

[Zhan00] X. Zhang and C. Qiao, An effective and comprehensive approach for traffic grooming and wavelength assignment in SONET/WDM rings. IEEE/ACM Transactions on Networking, vol. 8, no. 5, Oct. 2000, pp. 608–617.

[Zhou06] L. Zhou, T. Y. Chai, C. V. Saradhi, Y. Wang, V. Foo, Q. Qiang, J. Biswas, C. Lu, WDM optical network testbed and distributed storage application, IEEE Communications Magazine, Vol. 44, No. 2, Feb. 2006, pp. 23–29.

[Zhu04] Y. Zhu, A. Jukan, M. Ammar, W. Alanqar, End-to-End service provisioning in multigranularity multidomain optical networks, IEEE International Conference on Communications, vol. 3, 20-24 June 2004, pp. 1579–1583.

[Zhu02a] K. Zhu, B. Mukherjee, Traffic grooming in an optical WDM mesh network, IEEE Journal on Selected Areas in Communications, vol. 20, no. 1, Jan. 2002, pp. 122–133.

[Zhu03e] H. Zhu, H. Zang, K. Zhu, B. Mukherjee, A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks, IEEE/ACM Transactions on Networking, vol. 11, no. 2, April 2003, pp. 285–299.

[Zhu03b] K. Zhu, H. Zhu, B. Mukherjee, Traffic engineering in multigranularity heterogeneous optical WDM mesh networks through dynamic traffic grooming, IEEE Network, vol. 17, no. 2, March-April 2003, pp. 8–15.

[Zhu03c] K. Zhu, H. Zang, B. Mukherjee, A comprehensive study on next-generation optical grooming switches, IEEE Journal onSelected Areas in Communications, vol. 21, no. 7, Sept. 2003, pp. 1173–1186.