D3.3: Final report on the Adaptive Network Manager and GMPLS control plane

Status and Version:	Version 0.13	
Date of issue:	20151113	
Distribution:	Public	
Author(s):	Name	Partner
	Ramon Casellas	CTTC
	Ricardo Martínez	CTTC
	Raül Muñoz	CTTC
	Ricard Vilalta	CTTC
	Felix Wissel	DT
	Matthias Gunkel	DT
	Filippo Cugini	CNIT
	Francesco Paolucci	CNIT
	Antonio D'Errico	TEI
	Roberto Morro	TI
	Sergio Lopez Buedo	NAUDIT
	Victor López	Telefónica
	Oscar González de Dios	Telefónica
	Juan Pedro Fernández-Palacios	Telefónica
Checked by:	Adrian Farrel / Daniel King	ODC

Table of Contents

Table of Figures

1 Introduction

1.1 Executive Summary

This document constitutes IDEALIST Deliverable D3.3, *"Final report on the Adaptive Network Manager and GMPLS control plane*".

This deliverable reports; first, the conclusions and final design for the GMPLS control plane for the control of flexi-grid optical networks, covering the requirements of dynamic provisioning and recovery; second, the design and evaluation of a hierarchical PCE for interoperable end-to-end path computation, provisioning and restoration; third, the final design covering the Adaptive Network Manager, including the assessment of selected use cases and the interfacing with the GMPLS control plane and PCE for global concurrent optimization.

It is not a goal of this deliverable to re-state what has been reported and documented in previous ones [D3.1] and [D3.2] but to present a summary and main highlights, insisting on the final solution and wrapping up and concluding.

Where appropriate, new studies that were not previously reported and focusing, e.g. on the use of Segment routing or the use of a controller for multi-layer service provisioning combining packet and optical circuit are detailed. This is also in line with Year 3 objective of considering multi-layer (e.g. signal/media and packet/optical) integration.

In particular, The IDEALIST control plane architecture encompasses a transport core network, where underlying network covers mainly the multi-domain optical network, with limited consideration of the IP/MPLS routers located at the edges for the purposes of multi-layer service provisioning. While emphasis has been given for such nodes to be IP/MPLS routers, nothing precludes other interfaces (e.g., Ethernet) to be used, provided that the adaptation functions are defined.

The network is segmented into domains, typically for scalability reasons. It also reflects the potential deployments in which each domain corresponds to a given (hardware) vendor. By design, each domain has full visibility of the details on the network elements within each domain and limited (e.g. reachability) visibility of endpoints not in its domain. From a networking perspective, a flexible grid network is assumed to be a layered network in which the media layer is the server layer and the optical signal layer is the client layer. In the media layer, switching is based on a frequency slot, and the size of a media channel is given by the properties of the associated frequency slot. In this layered network, the media channel transports an Optical Tributary Signal (OTS). Within each domain, the architecture assumes a deployment model in which an active-stateful PCE controls the establishment of media channels (flexi-grid LSPs). Each domain PCE acts as a child PCE (cPCE) of a parent PCE (pPCE) that has ultimate control over path computation across all the domains (with full or abstracted topology visibility). Ultimate control of the pPCE and the orchestration integration with the IP/MPLS layer is the responsibility of the Adaptive Network Manager (ANM), implemented in terms of the Applications Based Network Operations (ABNO) [ABNO] architecture.

The GMPLS architecture defines a set of standard protocols constituted by three pillars: a signalling protocol (i.e. RSVP-TE [RSVP-TE]) used for setting up the end-to-end connections (Label Switched Paths - LSPs), a routing protocol (e.g. OSPF-TE [OSPF-TE]) used for topology and network resource dissemination, and a Link Management Protocol, such as LMP [LMP]. Control plane extensions affected all the protocols of the GMPLS suite together with the ones adopted as northbound interfaces between the cPCEs and the pPCE (i.e.

PCEP and BGP-LS). The main purpose of these extensions is the handling of the media layer in which switching is based on a frequency slot described as central frequency and a slot width. Most of them have already been documented in previous WP3 deliverables and are reported here only for completeness; others, covering the case in which transceivers in different domains must interwork, are detailed.

The multi-domain aspects are addressed by relying on the hierarchical PCE (H-PCE) [HPCE] architecture, suitably extended to cover flexi-grid extensions and with stateful and active capabilities. The recently proposed BGP-LS protocol has been adopted to provide TE information to the pPCE, thus enabling efficient domain sequence computation and, according to the adopted policy, also intra-domain segment definition. Indeed, in addition to inter-domain links, selected topological elements can be provided to the pPCE, either representing an abstracted or complete view of intra-domain resources.

IDEALIST main focus has been the design of the control plane of the so-called *media-layer* in which variable sized frequency slots are stablished. That said, a significant amount of work has been dedicated to multi-layer aspects, notably in terms of packet and optical integration. For example, the multi-layer provisioning use case on the ABNO was defined as automated IP link provisioning and as MPLS service provision. The automated IP link provisioning between two routers which are connected via an optical network composed of (elastic) ROADMs providing connectivity to several IP routers. Additionally, several approaches such as based on Software Defined Networking (SDN) or Segment Routing (SR) have appeared as new networking paradigm to drastically reduce the cost of operation and reduce the time-to-market of introducing a new service, by increasing the level of automation on the control and configuration of the network infrastructure.

From the point of view of integration with network management and network planning, the network capacity planning process is typically an offline activity, and is based on very long planning cycles (yearly, quarterly). Generally, this is due to the static and inflexible nature of current networks. This can be said for both the transport -optical and Ethernet- layer, as well as for the IP/MPLS layer, which should be inherently more dynamic compared to the underlying transport infrastructure. To support dynamicity in the connection deployment, current network architecture needs evolve to include the service layer and the network elements to support multi-service provisioning in multi-vendor and multi-technology scenarios. To do so, two standard interfaces are required. Firstly, the *north bound* interface that, among other functionalities, gives an abstracted view of the network, enabling a common entry point to provision multiple services and to provision the planned configuration for the network. Moreover, this interface allows coordinating network and service layer according to service requirements. Secondly, the *south bound interface* covering provisioning, monitoring, and information retrieval. Moreover, operators should require some man-machine iteration, and new configurations have to be reviewed and acknowledged before being implemented in the network.

The final architecture of the proposed monitoring scheme relies on passive monitoring probes located in the edges of the network, at the grey interfaces of the routers. These probes are in charge of obtaining network flows, which are the main input to the bandwidth estimation algorithm.

Last, WP3 in collaboration with WP4 has defined the scenario for the main IDEALIST multi-partner integrated testbed. The scenario is built by the interconnection of different components, physically distributed within labs. The testbed encompasses three Flexi-grid domains with different capabilities, one hierarchical PCE, an ABNO Controller and the PLATON planning tool.

The last part of this deliverable is dedicated to summarize the dissemination activities within WP3 and to perform an exhaustive review and assessment of IDEALIST project objectives that are applicable to WP3.

IDEALIST WP3 has had an excellent impact on both; scientific dissemination in peer reviewed conferences, journals and magazines; and contributions to standards. The main documents in the IETF CCAMP/PCE working groups related to GMPLS control plane for flexi-grid networks and ABNO are authored by IDEALIST members.

All the objectives related to WP3 have been completed, in terms of objectives per year, as well as technical (measurable) parameters, for which metrics have been obtained by either simulations, feedback or combinations of both, gained from real implementations in the scope of WP4. This has been reported in the deliverable, with details on a per objective basis.

1.2 Reference Material

1.2.1 Reference Documents

[ABNO] King, D., Farrel, A., "PCE-Based Architecture for ABNO", IETF RFC7491 March 2015

[Agu14] A. Aguado, V. López J. Marhuenda, Ó. González de Dios and J. P. Fernández-Palacios: ABNO: a feasible SDN approach for multi-vendor IP and optical networks, in Proc. Optical Fiber Conference (OFC), Th3I.5, Mar 2014

[Agu15] A. Aguado, V. López J. Marhuenda, Ó. González de Dios and J. P. Fernández-Palacios: ABNO: a feasible SDN approach for multi-vendor IP and optical networks, Journal of Optical Communications and Networking, Vol. 7, Iss. 2, pp. A356–A362, February 2015

[Agu15b] A. Aguado, S. Peng, M. V. Alvarez, V. López, T. Szyrkowiec, A. Autenrieth, R. Vilalta, R. Muñoz, R. Casellas, R.Martínez, N. Yoshikane, T. Tsuritani, R. Nejabati, D. Simeonidou: Dynamic Virtual Network Reconfiguration over SDN Orchestrated Multi-Technology Optical Transport Domains, in European Conference on Optical Communication (ECOC), Sep 2015.

[AliIETF] Z. Ali, et al. "Path Computation Element Communication Protocol (PCEP) Extensions for remote-initiated LSP Usage", http://tools.ietf.org/html/draft-ali-pce-remote-initiated-lsp-usage

[BGP-LS] Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", Work in Progress, draft-ietf-idr-ls-distribution-10, January 2015

[Clem14] A. Clemm, H. Ananthakrishnan, J. Medved, T. Tkacik, R. Varga, N. Bahadur, "A YANG Data Model for Network Topologies", IETF Internet draft draft-clemm-i2rs-yang-network-topo-00.txt, February 2014.

[Cug14] Cugini, F.; Paolucci, F.; Fresi, F.; Meloni, G.; Berrettini, G.; Sambo, N.; Giorgetti, A; Foggi, T.; Poti, L.; Castoldi, P., "Benefits of active stateful PCE for flexgrid networks," Optical Fiber Communications Conference and Exhibition (OFC), 2014, San Francisco, USA

[Cug15] Cugini, F.; Fresi, F.; Paolucci, F.; Meloni, G.; Sambo, N.; Giorgetti, A; Foggi, T.; Poti, L.; Castoldi, P., "Active Stateful PCE With Hitless LDPC Code Adaptation [Invited]", JOCN, 2015

[D1.5] IDEALIST Deliverable D1.5, "Evaluation of Flexgrid Technologies", October 2014

[D2.1] IDEALIST Deliverable D2.1, "Assessment of all potential flexgrid/flexrate data plane technologies including transmission &switching", September 2013.

[D3.1] IDEALIST Deliverable D3.1 "Reference GMPLS control plane, PCE and Adaptive Network Manager architectures" July 2013

[D3.2] IDEALIST Deliverable D3.2, "Design and Evaluation of the Adaptive Network Manager and Functional Protocol Extensions", November 2014

[D4.3] IDEALIST Deliverable D4.3, "Node prototypes and adaptive network manager prototype implementation", June 2015

[D4.4] IDEALIST Deliverable D4.4, "Experimental validation of IDEALIST network prototypes: results and analysis", October 2015

[D6.3] IDEALIST Deliverable D6.3, "Standardization and dissemination activities in the second half of the project", October 2015

[Eucnc14] R. Casellas, R. Muñoz, R. Martínez, R. Vilalta, F. Cugini, F. Paolucci, O. González, V. López, J.P. Fernández-Palacios, R. Morro, A. Di Giglio, D. King, A. Farrel, IDEALIST Control Plane Architecture for Multi-domain Flexi-Grid Optical Networks , in Proceedings of European Conference on Networks and Communications (EUCNC), 23-26 June 2014, Bologna (Italy).

[Eucnc15] R. Casellas, F. Cugini, R. Martínez, F. Paolucci, R. Muñoz, P. Castoldi, R. Vilalta, Control Plane Solutions for Sliceable Bandwidth Variable Transceivers in Elastic Optical Networks , in proceedings of European Conference on Networks and Communications (EuCNC 2015), 19-2 July 2015, Paris (France).

[FlexiFw] O. Gonzalez de Dios, R. Casellas, editors, “Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks’, IETF draft-ietf-ccamp-flexi-grid-fwk, to be published as RFC.

[FlexiLabel] A. Farrel, D. King, Y. Li, F. Zhang, "Generalized Labels for the Flexi-Grid in Lambda Switch Capable (LSC) Label Switching Routers", IETF draft-ietf-ccamp-flexigrid-lambda-label, work in progress.

[FlexiOSPFTE] X. Zhang, et. al., “GMPLS OSPF-TE Extensions in Support of Flexible Grid”, draft IETF, draft-ietf-ccamp-flexible-grid-ospf-ext, June 2014.

[FlexiRSVPTE] Zhang, F., Zhang, X., Farrel, A., González de Dios, O., Ceccarelli, D., "RSVP-TE Signaling Extensions in support of Flexible Grid, IETF draft-ietf-ccamp-flexible-grid-rsvp-te-ext-02.txt, work in progress

[Funems13] R. Muñoz, V. López, R. Casellas, O. González de Dios, F. Cugini, N. Sambo, O. Gerstel, A. Denrrico, D. King, S. López-Bueno, P. Layec, A. Cimmino, R. Martínez, R. Morro, IDEALIST control and service management solutions for dynamic and adaptive flexi-grid DWDM networks , in Proceedings of Future Network & Mobile Summit, 3-5 July 2013, Lisbon (Portugal).

[G.694.1] International Telecommunications Union, "ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, draft v1.6 2011/12", 2011.

[G.709] International Telecommunications Union, "ITU-T Recommendation G.709, "Interface for the optical transport network", February 2012

[G.800] International Telecommunications Union, "ITU-T Recommendation G.800: Unified functional architecture of transport networks.", February 2012

[G.872] International Telecommunications Union, "ITU-T Recommendation G.872, Architecture of optical transport networks, draft v0.16 2012/09 (for discussion)", 2012.

[Gar08] García-Dorado, J. L. and Aracil, J. and Hernández, J. A. and López de Vergara, J. E., “A queueing equivalent thresholding method for thinning traffic capture”, IEEE/IFIP Network Operations and Management Symposium, 2008, pp176-183

[Gif14a] Ll. Gifre, F. Paolucci, A. Aguado, R. Casellas, A. Castro, F. Cugini, P. Castoldi, L. Velasco, and V. Lopez: Experimental Assessment of In-Operation Spectrum Defragmentation, in Springer Photonic Network Communications, June 2014, Volume 27, Issue 3, pp 128-140

[Gon15] O. González de Dios, R. Casellas, R. Morro, F. Paolucci, V. López, R. Martínez, R. Muñoz, R. Vilalta, P. Castoldi, "First Multi-partner Demonstration of BGP-LS enabled Inter-domain EON control with H-PCE", in Proceedings of Optical Fiber Communication Conference (OFC2015), 22-26 March 2015, Los Angeles (USA).

[Gon-JOCN15] O. González de Dios, R. Casellas, R. Morro, F. Paolucci, V. López, R. Martínez, R. Muñoz, R. Vilalta, P. Castoldi, "Multi-partner Demonstration of BGP-LS enabled multi-domain EON control and instantiation with H-PCE domain EON control with H-PCE", JOCN 2015

[HPCE] King, D., Ed., and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012, <http://www.rfc-editor.org/info/rfc6805>

[LMP] Lang, J., Ed., "Link Management Protocol (LMP)", RFC4202, October 2005, <http://www.rfc-editor.org/info/rfc4204>

[Mar14] R. Martínez, Ll. Gifre, R. Casellas, L. Velasco, R. Muñoz and R, Vilalta: Experimental Validation of Active Frontend – Backend Stateful PCE Operations in Flexgrid Optical Network Re-Optimization, accepted in European Conference on Optical Communications (ECOC) 2014.

[May15] A. Mayoral, R. Vilalta, R. Casellas, R. Muñoz, R. Martínez, Traffic Engineering enforcement in multi-domain SDN orchestration of Multi-Layer (packet/optical) networks , in Proceedings of the 41st European Conference on Optical Communication (ECOC 2015), 27 September-1 October 2015, Valencia (Spain).

[Mun14] R. Muñoz, R. Casellas, R. Vilalta, R. Martínez: “Dynamic and Adaptive Control Plane Solutions for Flexi-grid Optical Networks based on Stateful PCE”, Journal of Lightwave Technology, Vol. 32, No. 16, pp. 2703 – 2715, August 2014.

[NETCONF] R. Enns, M. Bjorklund, J. Schönwälder, A. Bierman, Eds. “Network Configuration Protocol (NETCONF)”, IETF Request for Comments 6241, June 2011

[OpenFlow] Open Networking Foundation, "OpenFlow Switch Specification Version 1.4.0 (Wire Protocol 0x05)", October 2013.

[OSPFInter] M. Chen, R. Zhang, X. Duan, “OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering”, IETF RFC5392 [online: http://tools.ietf.org/html/rfc5392]

[OSPF-TE] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003, <http://www.rfc-editor.org/info/rfc3630>.

[Pao14] Paolucci, F.; Fresi, F.; Castro, A; Velasco, L.; Cugini, F.; Sambo, N.; Giorgetti, A; Poti, L.; Castoldi, P., "Filter optimization in SDN-based flexgrid networks," Optical Fiber Communications Conference and Exhibition (OFC), 2014 San Francisco, USA

[Pao15-PN] F. Paolucci, A. Castro, F. Fresi, M. Imran, A. Giorgetti, B. B. Bhownik, G. Berrettini, G. Meloni, F. Cugini, L. Velasco, L. Potì, P. Castoldi, "Active PCE demonstration performing elastic operations and hitless defragmentation in flexible grid optical networks", Journal of Photonic Network Communications (PNET), 2014

[PAP03] Papagiannaki, K. and Taft, N. and Zhang, Z.-L. and Diot, C, “Long-term forecasting of Internet backbone traffic: observations and initial models”, IEEE Infocom 2003, Vol 2, pp 1178-1188

[PCEArch] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006, <http://www.rfc-editor.org/info/rfc4655>.

[PCEFAQ] Farrel, A., King, D., "Unanswered Questions in the Path Computation Element Architecture", IETF RFC 7399, October 2014

[PCEP] Vasseur, JP., Ed., and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009, <http://www.rfc-editor.org/info/rfc5440>.

[PCEPGMPLS] C. Margaria, O. Gonzalez de Dios, and F. Zhang, "PCEP extensions for GMPLS," IETF draft-ietf-pce-gmpls-pcep-extensions (work in progress).

[PCEPHier] F. Zhang, Q. Zhao, O. Gonzalez de Dios, R. Casellas, D. King, "Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)", draft-ietf-pce-hierarchy-extensions (work in progresss).

[PCEPInit] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", Work in Progress, draft-ietf-pce-pce-initiated-lsp-03, March 2015.

[PCEState] Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", Work in Progress, draft-ietf-pce-stateful-pce-10, October 2014.

[RFC6205] Otani, T., Ed., Li, D., Ed. "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC6205, March 2011, <http://www.rfc-editor.org/info/rfc6205>

[RSVP-TE] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003, <http://www.rfc-editor.org/info/rfc3473>.

[Rui15] M. Ruiz, G. Sutter, S. Lopez-Buedo, J. Ramos, J.E. Lopez de Vergara, J. Aracil, "Leveraging Open Source Platforms and High-Level Synthesis for the Design of FPGA-Based 10 GbE Active Network Probes", 2015 International Conference on Reconfigurable Computing and FPGAs (ReConFig 2015) - In press

[Sam14] Sambo, N.; Meloni, G.; Paolucci, F.; Cugini, F.; Secondini, M.; Fresi, F.; Poti, L.; Castoldi, P., "Programmable Transponder, Code and Differentiated Filter Configuration in Elastic Optical Networks," Lightwave Technology, Journal of , vol.32, no.11, pp.2079,2086, June1, 1 2014

[Sga-NOC15] A Sgambelluri, F. Paolucci, A Giorgetti, F Cugini, P Castoldi, "SDN and PCE Implementations for Segment Routing", NOC 2015

[Sga-OFC15] A Sgambelluri, A Giorgetti, F Cugini, G Bruno, F Lazzeri, P Castoldi, "First Demonstration of SDN-based Segment Routing in Multi-layer Networks", OFC 2015

[SR] C. Filsfils et al., "Segment routing architecture", draft-filsfils-spring-segment-routing-04, July 2015

[Vel14a] L. Velasco, D. King, O. Gerstel, R. Casellas, A. Castro, and V. López, "In-Operation Network Planning," IEEE Communications Magazine, vol. 52, pp. 52-60, 2014.

[YANG] M. Bjorklund, Ed., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)," IETF Request for Comments 6020, October 2010.

[YANGFlexi] YANG data model for Flexi-Grid Optical Networks https://datatracker.ietf.org/doc/draft-vergara-ccamp-flexigrid-yang (work in progress)

[YANGLSPs] A generic YANG Data Model for Label Switch Path (LSP), https://tools.ietf.org/html/draft-zhang-mpls-lspdb-yang (work in progress)

1.2.2 Acronyms

ANM	Adaptive Network Manager
API	Application Programming Interface

ASO	Application Service Orchestrator
ABNO	Application-Based Network Operations
ASON	Automatically Switched Optical Networks
BV-OXC	bandwidth variable optical cross connect
BVT	Bitrate Variable Transponder / Transceiver
BGP	Border Gateway Protocol
BGP-LS	Border Gateway Protocol – Link State
cPCE	Child Path Computation Element
CoS	Class of Service
CLI	Command Line Interface
CCAMP	Common Control and Measurement Plane (IETF working group)
CC	Connection Controller
CDF	Cumulative Distribution Function
DCN	Data Communications Network
DWDM	Dense Wavelength Domain Multiplexing
DP	Dual Polarisation
EON	Elastic Optical Network
E2E	End-to-End
ECMP	Equal Cost Multi Path
ELC	Explicit Label Control
ERO	Explicit Route Object
ETC	Explicit Transponder Control
FEC	Forward Error Correction
FS	Frequency Slot
GMPLS	Generalized Multi-Protocol Label Switching
GbE	Gigabit Ethernet
Gbps	Gigabit per second
GCO	Global Concurrent Optimization
GSC	GMPLS enabled SDN controllers
H-PCE	Hierarchical Path Computation Element
HTTP	Hyper Text Transfer Protocol
IT	Information Technology
IaaS	Infrastructure as a Service
ISIS-TE	Intermediate System to Intermediate System – Traffic Engineering
IGP	Internal Gateway Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol

IPCC	IP Control Channels
JSON	JavaScript Object Notation
LSR	Label Switch Router
LSP	Label Switched Path
LMP	Link Management Protocol
LSA	Link State Advertisement
LSPDB	LSP Database
MAN	Metropolitan Area Network
MF-OTP	Multi Frequency - Optical TransPonder
MPLS	Multi-Protocol Label Switching
MPLS-TE	Multi-Protocol Label Switching – Traffic Engineering
NMS	Network Management System
NCF	Nominal Central Frequencies
NBI	North Bound Interface
ONF	Open Networking Foundation
OSPF-TE	Open Shortest Path First – Traffic Engineering
OSI	Open Systems Interconnection (model)
OF	OpenFlow
OFP	OpenFlow Protocol
OAM	Operation, Administration and Maintenance
OCh-P	Optical Channel Payload
OXC	Optical Cross Connect
OFDM	Optical Frequency Division Multiplexing
OLA	Optical Line Amplifier
OSNR	Optical Signal to Noise Ration
OTN	Optical Transport Network
OTS	Optical Tributary Signal
pPCE	Parent Path Computation Element
PCC	Path Computation Client
PCE	Path Computation Element
PCEP	Path Computation Element Protocol
PLI	Physical Layer Impairments
PM-16QAM	Polarization Multiplexed 16 Quadrature Amplitude Modulation
PM-QPSK	Polarization Multiplexed Quadrature Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
QoS	Quality of Service

QoT	Quality of Transmission
ROADM	Reconfigurable Optical Add Drop Multiplexer
REST	Representational State Transfer
RSVP-TE	ReSerVation Protocol – Traffic Engineering
RSA	Routing and Spectrum Assignment
RWA	Routing and Wavelength Assignment
RC	Routing Controller
SID	Segment (Routing) Identifier
SR	Segment Routing
SNMP	Simple Network Management Protocol
SBVT	Sliceable Bitrate Variable Transponder
SWG	Slot Width Granularity
SDN	Software-Defined Networking
SA	Spectrum Assignment
SSON	Spectrum Switched Optical Network
SDO	Standards Defining Organizations
SF-PD	StateFul-H-PCE with Per-Domain instantiation and stitching
SL-E2E	StateLess-H-PCE with E2E signaling and instantiation
TE	Traffic Engineering
TE-LSA	Traffic Engineering - Link State Advertisement
TED	Traffic Engineering Database
TCP	Transmission Control Protocol
TLV	Type, Length, Value
UDP	User Datagram Protocol
UNI	User-Network Interface
VNTM	Virtual Network Topology Manager
WSS	Wavelength Selective Switch
WSON	Wavelength Switched Optical Network

1.3 Document History

Version	Date	Authors	Comment
0.01	11/06/2015	WP3/CTTC	Table of contents proposal
0.02	27/07/2015	WP3/CTTC	Reworked ToC. Initial contributions
0.03	26/08/2015	WP3/DT	Integrated DT Section
0.04	01/09/2015	WP3/CNIT/TEI	Integrated CNIT and TEI sections
0.05	10/09/2015	WP3/TI	Integrated CTTC and TI contributions

0.05.1	13/10/2015	WP3/TID	Integrated Telefónica sections on ANM
0.05.2	14/10/2015	WP3/NAUDIT	Integrated Naudit contributions Updated sections on Exec. Summary / structure
0.06	15/10/2015	WP3/All	Updated document sections
0.07	20/10/2015	WP3/All	First integrated draft, updated sect. 8
0.08	23/10/2015	WP3/All	Updated references, further work on objective assessment and completion
0.09	28/10/2015	WP3/All	Updated measurable key performance indicators and objectives table.
0.10	09/11/2015	WP3/TID	Integrated section on signal layer multi-layer aspects and feedback from wp1
0.11-0.12	12/11/2015	WP3/All	Editorial changes, fixed spotted typos
0.13	13/11/2015	WP3/ODC	Review

1.4 Document structure

The rest of the document is structured as follows: Section 2 provides a high level overview of IDEALIST solution for the control and management of Elastic Optical Networks. Section 3 presents the details of the architecture, focusing in each of the aforementioned components. At the end of the project, there are related topics that are being developed to extend the scope of WP3, including new aspects for Multi-Layer integration; these are reported in Section 4. Section 5 is dedicated to report, in summary, the protocol extensions related to the control plane as defined in WP3 and implemented in WP4. Section 6 describes, from a high level perspective, the final IDEALIST scenario that is being conceived in cooperation with WP4, WP2 and WP1, based on the multi-domain testbed deployed in WP4. Section 7 summarizes the main feedback obtained from other work packages that has been used in WP3. Section 8 is dedicated to provide a short summary of the main dissemination activities related to WP3, in cooperation with WP6. Section 9 is devoted to the assessment of WP3 objectives, including, on the one hand the objectives committed in terms of measureable parameters and related metrics and, on the other hand, the list of yearly objectives, their completion level and how they are reported. Finally, Section 10 concludes the deliverable.

2 IDEALIST Solution for Control and Management of Elastic Optical Networks

IDEALIST WP3 main outcome is the design of a solution for the control and management of elastic optical networks, which combines the concept of Adaptive Network Manager (ANM) implemented in terms of the Applications Based Network Operations (ABNO) architecture, and a multi-domain GMPLS/PCE control plane, for the dynamic provisioning of elastic optical connections with recovery in a multi-layer (i.e. integrated with an IP/MPLS network) context.

In this sense, this section provides a high level overview of such solution, briefly mentioning each of the main components, which are later detailed in Section 3.

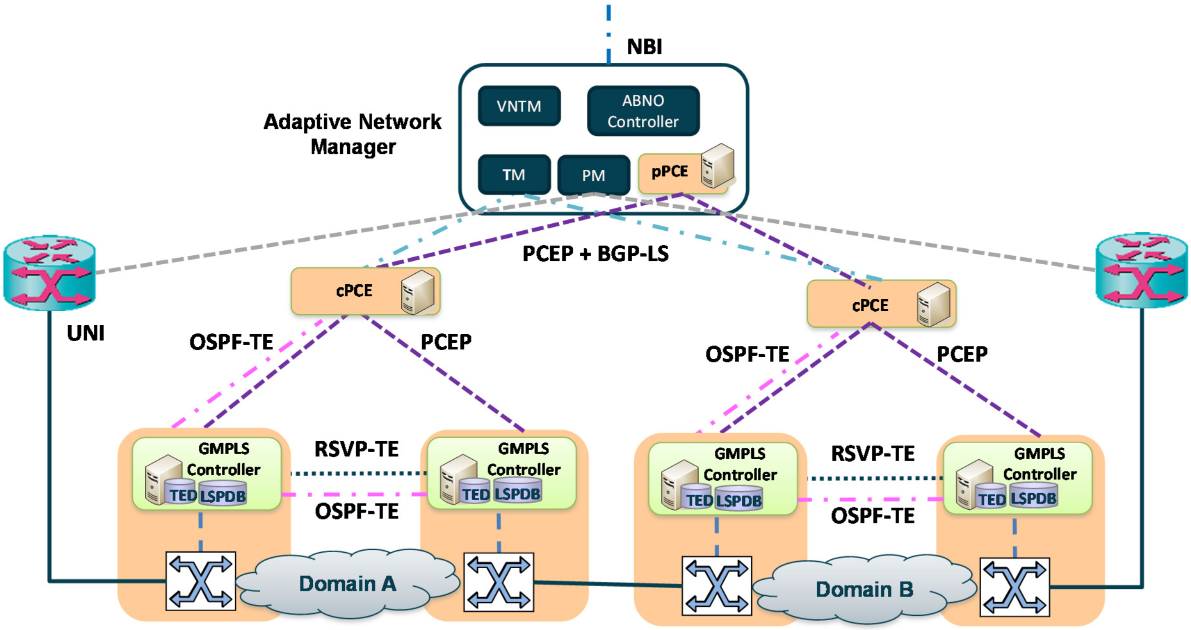


Figure 1 IDEALIST Solution for Control and Management of Elastic Optical Networks

2.1 Macroscopic Architecture

The IDEALIST control plane architecture is shown in Figure 1. The underlying network covers mainly the multi-domain optical network, with limited consideration of the IP/MPLS routers located at the edges for the purposes of multi-layer service provisioning. While emphasis has been given in such nodes being IP/MPLS routers, nothing precludes other interfaces (e.g. Ethernet) to be used, provided that the adaptation functions are defined.

The network is segmented into domains, typically for scalability reasons. It also reflects the potential deployments in which each domain corresponds to a given (hardware) vendor. By design, each domain has full visibility of the details on the network elements within each domain and limited (e.g. reachability) visibility of endpoints not in its domain. From a networking perspective, a flexible grid network is assumed to be a layered network [G.800][G.872] in which the media layer is the server layer and the optical signal layer is the client layer. In the media layer, switching is based on a frequency slot, and the size of a media channel is given by the properties of the associated frequency slot. In this layered network, the media channel transports an Optical Tributary Signal (OTS).

IDEALIST considers both the case in which domains are connected transparently or connected by means of transceivers connected e.g. back-to-back (that is, with optical / electro / optical regeneration).

Within each domain, the architecture assumes a deployment model in which an active-stateful PCE controls the establishment of media channels (flexi-grid LSPs). Each domain PCE acts as a child PCE (cPCE) of a parent PCE (pPCE) that has ultimate control over path computation across all the domains (with full or abstracted topology visibility).

Different provisioning models are considered: either by interacting directly with one (or multiple) head-end nodes, or by delegation of the provisioning to the pPCE which, in turn, may delegate to the corresponding per-domain cPCEs.

Ultimate control of the pPCE and the orchestration integration with the IP/MPLS layer is the responsibility of the Adaptive Network Manager (ANM).

2.2 ABNO based Adaptive Network Manager

The ANM is an entity in charge of deciding the optimal network configuration based on the status and monitoring information of such network. In order to decide the optimality of the network, in interacts with planning tools and on-line algorithms. The ANM does not replace the control plane, but extends and complements it, orchestrating with other systems (e.g. interacting with the client layer) and delegating specific functions (e.g. path computation) to it.

The ANM, implemented in terms of the Applications Based Network Operations (ABNO) [ABNO] architecture, stays at the top of the IDEALIST solution. The ABNO architecture is based on functional elements defined by the IETF, like the active stateful path computation element (PCE).

The ANM relies on standard and open interfaces, exporting a set of interfaces to applications (North Bound Interface or NBI) and interacting with the data plane, either directly or via the control plane.

2.3 GMPLS/PCE Control plane

The proposed control plane reference architecture for flexi-grid DWDM networks is based on the GMPLS architecture and its set of protocols, along with PCE based path computation and instantiation, considering both passive and active models. It addresses the automation of optical network functionalities such as connection provisioning and recovery (i.e. protection and restoration), traffic engineering or QoS.

The GMPLS architecture defines a set of standard protocols constituted by three pillars: a signalling protocol (i.e. RSVP-TE [RSVP-TE]) used for setting up the end-to-end connections (Label Switched Paths - LSPs), a routing protocol (i.e. OSPF-TE [OSPF-TE]) used for topology and network resource dissemination, and a Link Management Protocol, such as LMP [LMP]. A GMPLS control plane is a distributed entity composed of controllers (one per node) which execute several collaborative processes (Connection Controller -CC-, Routing Controller -RC-, path computation, etc.), and a data communication network based on IP control channels (IPCC) to allow the exchange of control messages between GMPLS controllers.

A GMPLS-enabled node (both control and hardware) is named Label Switched Router (LSR) in the GMPLS architecture. Under distributed control, each GMPLS controller manages the state of all the connections (i.e. LSPs) originated, terminated or passing-through a node, stored in the LSP Database (LSPDB), and maintains its own network state information (topology and resources), collected in a local Traffic Engineering Database (TED) repository.

2.4 GMPLS/PCE Control plane for multi domain

The multi-domain aspects are addressed by relying on the hierarchical PCE (H-PCE) [HPCE] architecture, suitably extended to cover flexi-grid extensions and with stateful and active capabilities, along with the following considerations:

- Each vendor / domain deploys its own GMPLS control plane implementation. The topology is disseminated by the instance of OSPF-TE within the domain. A set of functions such as the internal dissemination of LSAs and the status of the link

NCFS are common across all implementations. Other aspects, more research oriented scoped to a given vendor are also allowed.

- OSPF-TE [OSPFInter] extensions for the dissemination of inter-domain links are selected to enable the dissemination of inter-domain TE attributes within the domain. This covers attributes such as the local and remote autonomous system.
- Selected topological elements, such as the result of abstracting the domain (or, if applicable by policy, the whole domain) are announced to the parent PCE (pPCE) by means of BGP-LS.
- Inter-domain links are also forwarded to the pPCE; this allows the pPCE to construct a multi-domain topology in which each child PCE (cPCE) has control over the topological elements that it disseminates.
- Path computation can be done at a single domain level, by requesting the corresponding cPCE or, if the path computation involves multiple domains, the path computation is driven by the pPCE, which is responsible for the selection of domains to cross and delegates the segment expansion to the involved cPCEs.
- The provisioning is driven by the ANM, either by directly requesting the active stateful PCE using PCEP as provisioning interface [PCEPInit] or directly interacting with the head end node.

As mentioned in the last point, different scenarios have been considered in the scope of the project. The scenarios, detailed later in the document, differ based on whether there exists a single end-to-end RSVP-TE signaling session or whether the end-to-end service is obtained by the concatenation (i.e. stitching) of different segments. The scenario also defines whether optical / electrical / optical conversion happens at defined interfacing points or, on the contrary, the whole connection is transparent, end-to-end.

2.5 Multi-Layer aspects and considerations

IDEALIST main focus has been the design of the control plane of the so-called *media-layer* in which variable sized frequency slots are stablished. On top of these network media channels, there is Optical Tributary Signal (OTS).

The Multi-Layer considerations cover several aspects:

- First, the mapping and control of the so-called *Signal Layer,* which encompasses the signals that are transported over the OTS. This involves the direct control and configuration of the transponders attached at the end of the network media channel.
- Next, the coordinated control of both the packet (e.g. IP layer) and the optical (circuit) layer, addressed in IDEALIST by means of the ABNO/ANM architecture and controller that is responsible for the provisioning of network media channels and the signal adaptation of the client layer (with special emphasis on the IP/MPLS case). This means that the provisioning manager is also responsible for configuration and the IP routers end of the optical connections.
- Finally, the integration of the packet layer may also involve interacting with new control plane elements such as SDN/OpenFlow controllers or, more recently, the use of elements such as the PCE for Segment Routing.

3 Detailed IDEALIST Control plane solutions

3.1 Adaptive Network Manager (ANM)

The Adaptive Network Manager (ANM) follows the Application Based Network Orchestration (ABNO) framework. Within the IETF, ABNO is defined to provide a solution based on standard protocols and components [ABNO]. The main component of the ABNO architecture is the Path Computation Element (PCE). The IETF ABNO architecture is based on existing standard blocks defined within the IETF (PCE, ALTO, VNTM...), which could be implemented either in a centralized or distributed way according to network operator requirements. Thanks to the modular nature of ABNO architecture, building blocks can be deployed by different vendors or third parties and even by a single provider. This modularity and the standard interfaces between the modules solve the problem of vendor lock-in for the operators. On the other hand, ABNO is specially adapted to multi-domain and multivendor networks, enabling interoperability between control plane based and OF based domains. In this paper, we validate via experimentation the ABNO architecture for a multilayer use case in two scenarios with control plane and with OF in the optical layer. Next figure shows the modules that are composed in the ABNO framework.

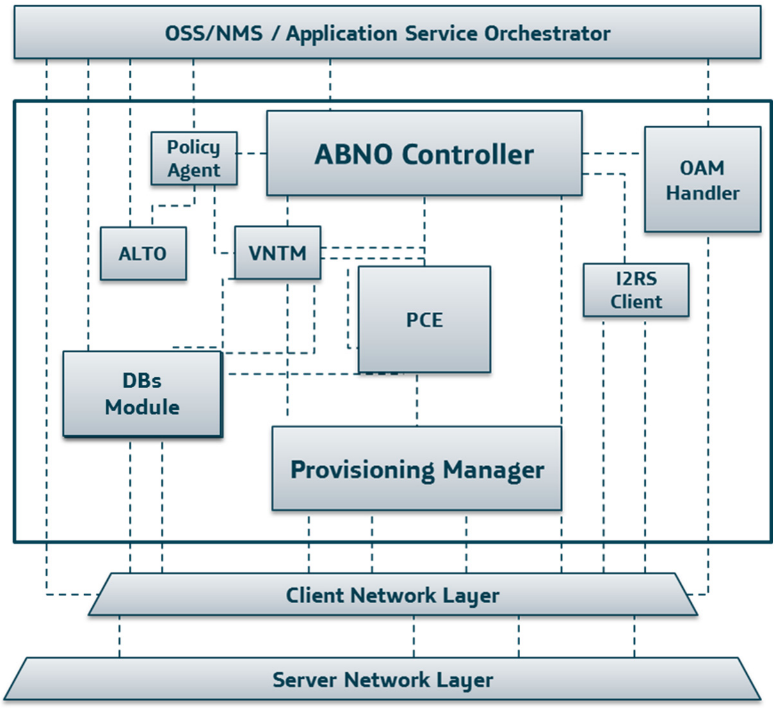


Figure 2 Block diagram of the ANM, implemented in terms of the ABNO architecture

The extensions described in [AliIETF] propose that the PCE protocol can be used to set-up GMPLS LSPs remotely. The idea behind is to enable an active PCE to set-up GMPLS LSPs. However, these extensions can be used to request a path setup from any element in the network. This feature within ABNO framework facilitates the interaction between multiple of the modules in the architecture. This allows several options in the workflow to provision a L0 connection.

The basic idea under this draft is that it allows to send a PCEP message with the ERO and it can be redirected by the GMPLS nodes using RSVP as shown in the figure bellow.

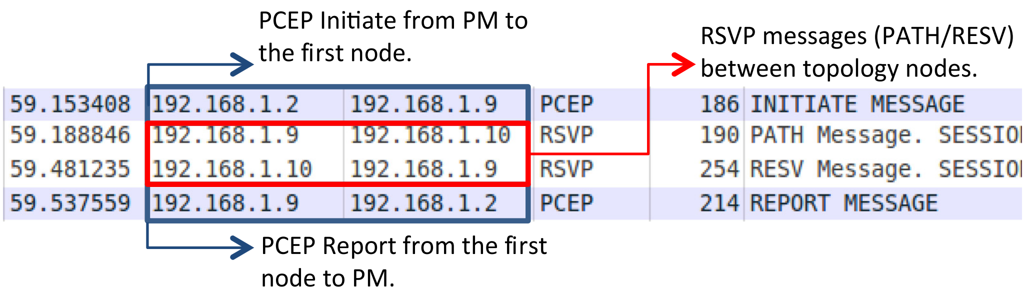


Figure 3 PCInitiate and RSVP messages in edge node

The next figure shows which are the RSVP messages that is redirected from the GMPLS nodes.

```
Resource Reservation Protocol (RSVP): PATH Message. SESSION: IPv4-LSP, Dest
⊞ RSVP Header. PATH Message.
⊞ SESSION: IPv4-LSP, Destination 192.168.1.16, Tunnel ID 3, Ext ID c0a80110.
⊞ HOP: IPv4, 192.168.1.9
⊞ TIME VALUES: 20000 ms
⊟ EXPLICIT ROUTE: Unnum 192.168.1.10/3, Label 1778384898, IPv4 192.168.1.11
    Length: 56
    Object class: EXPLICIT ROUTE object (20)
    C-type: 1
  ⊞ Unnumbered Interface-ID - 192.168.1.10, 3, Strict
  ⊞ Label Subobject - 1778384898, Strict
  ⊞ IPv4 Subobject - 192.168.1.11, Strict
  ⊞ Unnumbered Interface-ID - 192.168.1.11, 11, Strict
  ⊞ IPv4 Subobject - 192.168.1.16, Strict
⊞ LABEL REQUEST: Basic: L3PID: Unknown (0x0002)
⊞ SENDER TEMPLATE: IPv4-LSP, Tunnel Source: 192.168.1.9, LSP ID: 3.
⊞ SENDER TSPEC: IntServ,
```

Figure 4 RSVP PATH message in edge node

3.1.1 Interaction with control plane

Using the different stateful and stateless capabilities of the PCE and the single-domain or multi-domain scenarios, there can be differentiated four workflows where the ANM can interact with the control plane in L0 provisioning use cases:

- Single-domain with stateless PCE.
- Multi-domain with stateless PCE.
- Single-domain with stateful PCE.
- Multi-domain with stateful PCE.

Following each workflow is explained.

3.1.1.1 Single-domain with stateless PCE

On this scenario, we assume that there is a L0 PCE which is stateless. As the PCE is stateless, the ANM requires to run the Topology Module and the Provisioning Manager. The workflow follows the next process:

- A request from an end user via a NMS or in this case through the NMS application is sent.
- The NMS sends the information to the ABNO controller (two LSRs and the traffic parameters (e.g. bandwidth) required for the circuit).

- The ABNO controller asks the L0 PCE for a path in the L0 topology with a PCEP Request message and if there is path in the L0 topology.
- L0 PCE answers with a PCEP Response message including ERO object. If there is not a path in the L0 topology a NO PATH is sent to the ABNO. In this case, ABNO returns to the NMS an error message.
- The ABNO controller requests the Provisioning Manager to configure the L0 service/layer by establishing the path with a PCEP Initiate message.
- The Provisioning Manager queries the Topology Module for the description of each node.
- Depending on the technology and configuration mode, Provisioning Module selects a different protocol to complete the request. In this scenario only GMPLS is assumed, so the PM redirects the PCEP Initiate message to the GMPLS nodes. The GMPLS nodes send a RSVP Path message with the ERO sent by the PCE.
- Once the service has been established, Provisioning Manager notifies the ABNO Controller with a PCEP Report message.
- Similarly ABNO controller advertises the NMS.

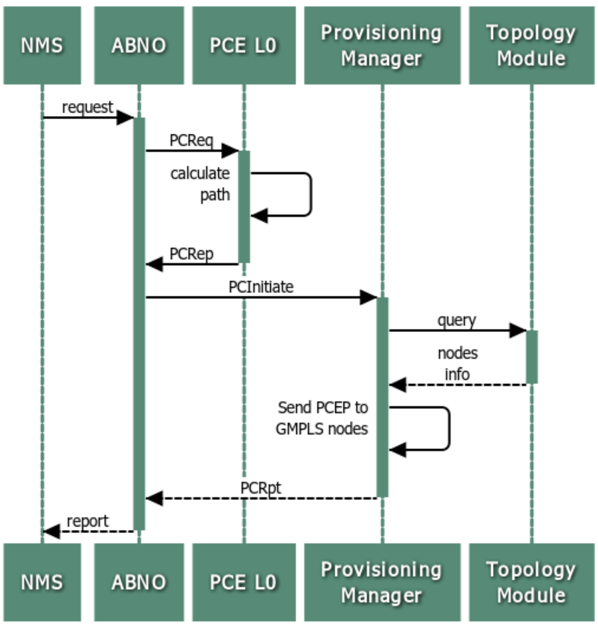


Figure 5 Single-domain with stateless PCE

3.1.1.2 Multi-domain with stateless PCE

On this scenario, we assume that there is a L0 parent PCE, which is stateless. As the PCE is stateless, the ANM requires to run the Topology Module and the Provisioning Manager.

- A request from an end user via a NMS or in this case through the NMS application is sent.
- The NMS sends the information to the ABNO controller (two LSRs and the traffic parameters (e.g. bandwidth) required for the circuit).
- The ABNO controller asks the Parent PCE for a path in the L0 topology with a PCEP Request message and if there is path in the L0 topology.
- Parent PCE request each Child PCE (in this case, L0 PCE1 and L0 PCE2) and, when they calculate intradomain path, they answer Parent PCE with a PCEP Response message with intradomain path information.
- Parent PCE, with all intradomain path information, creates a tree and prunes it to get full path and send to ABNO a PCEP Response message.
- The ABNO controller requests the Provisioning Manager to configure the L0 service/layer by establishing the path with a PCEP Initiate message.
- The Provisioning Manager queries the Topology Module for the description of each node.
- Depending on the technology and configuration mode, Provisioning Module selects a different protocol to complete the request. In this scenario only GMPLS is assumed, so the PM redirects the PCEP Initiate message to the GMPLS nodes. The GMPLS nodes send a RSVP Path message with the ERO sent by the PCE. Another option would be to send a PCEP Initiate to the head GMPLS node in each domain without progressing the RSVP through the E-NNI.
- Once the service has been established, Provisioning Manager notifies the ABNO Controller with a PCEP Report message.
- Similarly ABNO controller advertises the NMS.

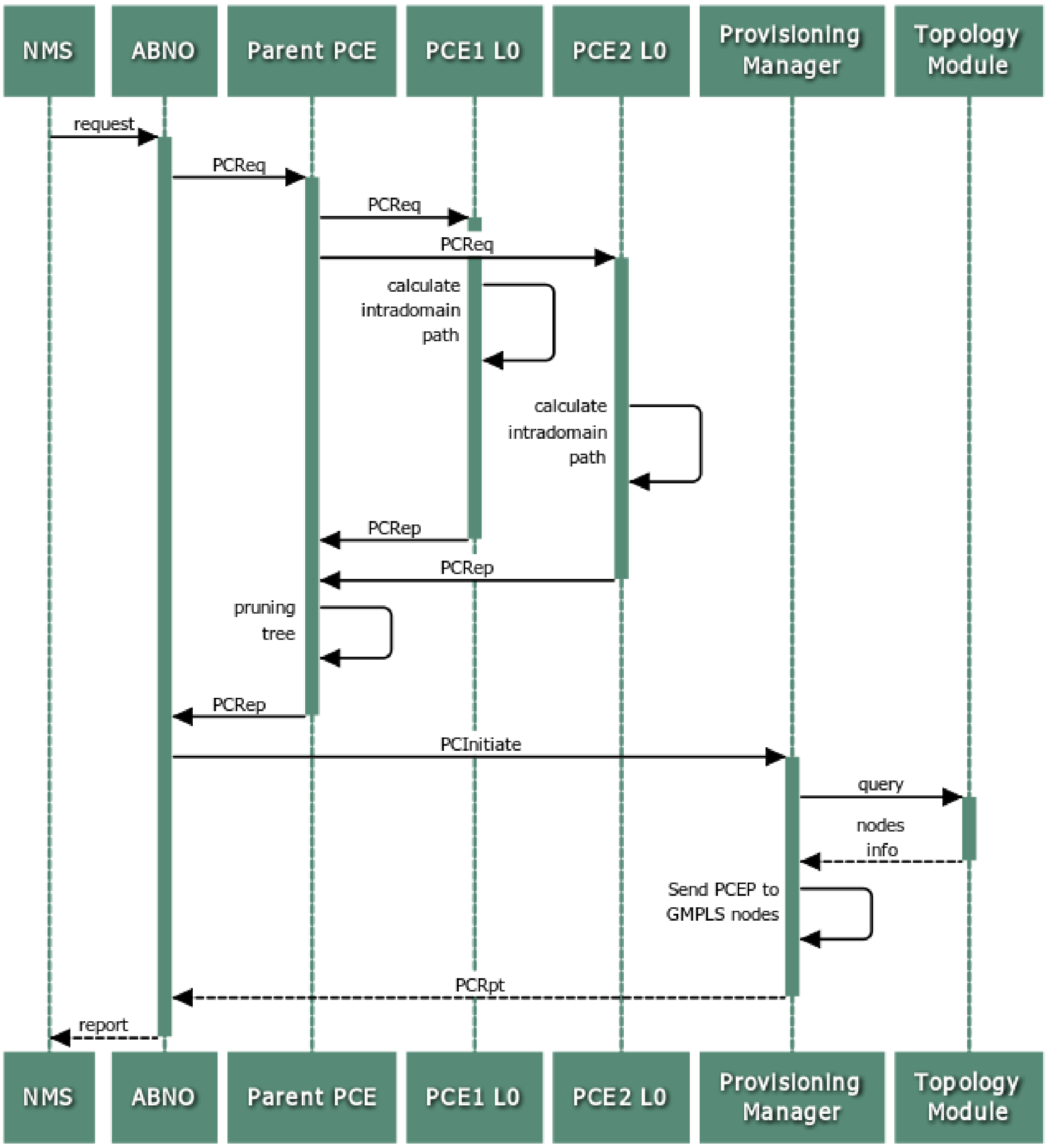


Figure 6 Multi-domain with stateless PCE

3.1.1.3 Single-domain with stateful PCE

On this scenario, we assume that there is a stateful L0 PCE. The stateful PCE acts as the Provisioning Manager, so there is no need to use it in the workflow.

- A request from an end user via a NMS or in this case through the NMS application is sent.
- The NMS sends the information to the ABNO controller (two LSRs and the traffic parameters (e.g. bandwidth) required for the circuit).
- The ABNO controller sends to the L0 PCE a PCEP Initiate message.
- L0 PCE determines the path and send PCEP message to GMPLS nodes. The GMPLS nodes send a RSVP Path message with the ERO sent by the PCE.

- Once the service has been established, L0 PCE notifies the ABNO Controller with a PCEP Report message.
- Similarly ABNO controller advertises the NMS.

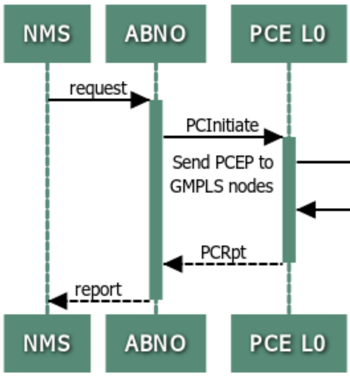


Figure 7 Single-domain with stateful PCE

3.1.1.4 Multi-domain with stateful PCE

A request from an end user via a NMS or in this case through the NMS application is sent.

- The NMS sends the information to the ABNO controller (two LSRs and the traffic parameters (e.g. bandwidth) required for the circuit).
- The ABNO controller sends to L0 PCE a PCEP Initiate message.
- Parent PCE request each Child PCE (in this case, L0 PCE1 and L0 PCE2) and, when they calculate intradomain path, they answer Parent PCE with a PCEP Response message with intradomain path information.
- Parent PCE, with all intradomain path information, creates a tree and prunes it to get full path and send to ABNO a PCEP Response message.
- L0 PCE spreads to each Child PCE the Initiate message with the selected path in each domain.
- The Child PCE redirects the PCEP Initiate message to the GMPLS nodes. The GMPLS nodes send a RSVP Path message with the ERO sent by the PCE. Another option would be that the parent PCE sends a PCEP Initiate to the head GMPLS node, but as the child PCEs are stateful it makes more sense to use their capabilities.
- Once the service has been established, each Child PCE notifies the Parent PCE with a PCEP Report message and this notifies the ABNO Controller with a similar PCEP Report message.
- Similarly ABNO Controller advertises the NMS.

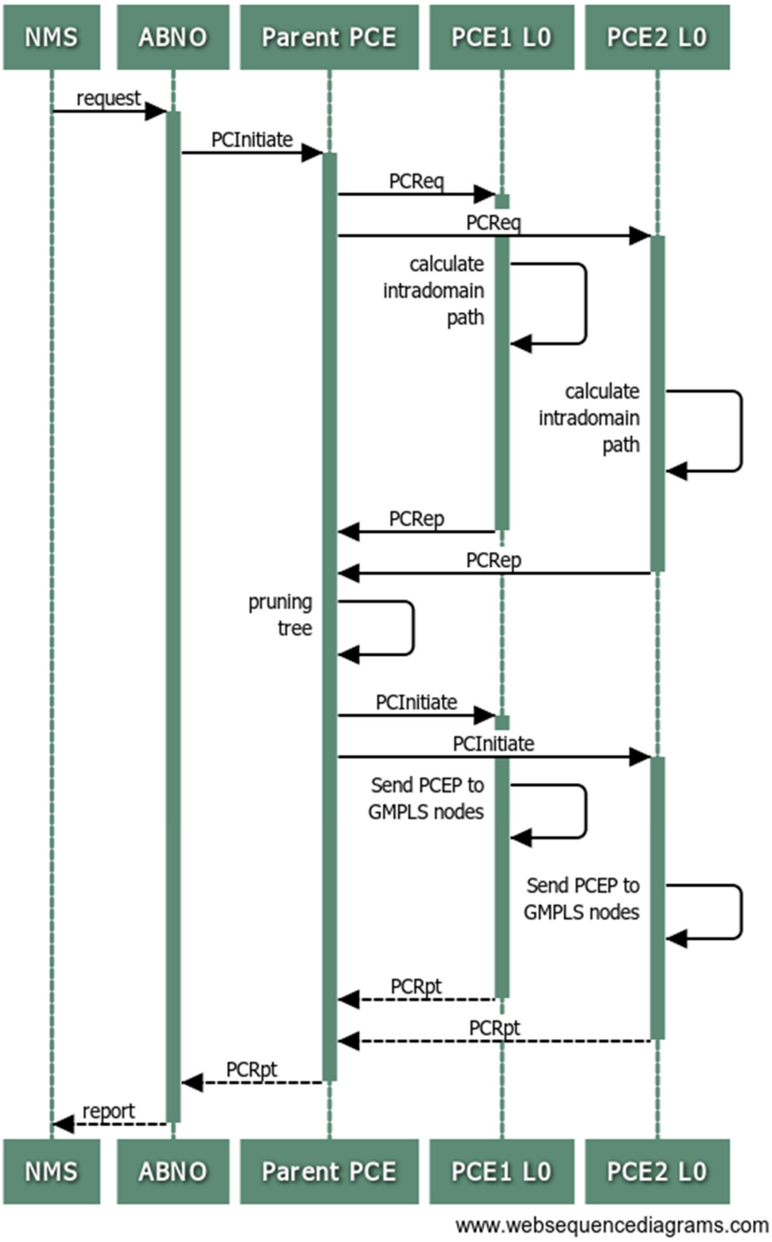


Figure 8 Multi-domain with stateful PCE

3.1.2 Interaction of the ANM with Planning Tools

Network capacity planning requires the placement of network resources to satisfy expected traffic demands and network failure scenarios. Today, the network capacity planning process is typically an offline activity, and is based on very long planning cycles (yearly, quarterly). Generally, this is due to the static and inflexible nature of current networks. This can be said for both the transport -optical and Ethernet- layer, as well as for the IP/MPLS layer, which should be inherently more dynamic compared to the underlying transport infrastructure. The latter might use automated Traffic Engineering (TE) techniques to place IP/MPLS traffic where the network resources are. However, the cycle to deploy any connection in the network is a long process, not only for the installation process, but also because operators does not have the information updated and the tools to automatically provision the services in the devices.

To support dynamicity in the connection deployment, current network architecture needs evolve to include the service layer and the network elements to support multi-service provisioning in multi-vendor and multi-technology scenarios. To do so, two standard interfaces are required. Firstly, the *north bound* interface that, among other functionalities, gives an abstracted view of the network, enabling a common entry point to provision multiple services and to provision the planned configuration for the network. Moreover, this interface also allows coordinating network and service layer according to service requirements. Secondly, the *south bound* interface covering provisioning, monitoring, and information retrieval. Moreover, operators should require some man-machine iteration, and new configurations have to be reviewed and acknowledged before being implemented in the network.

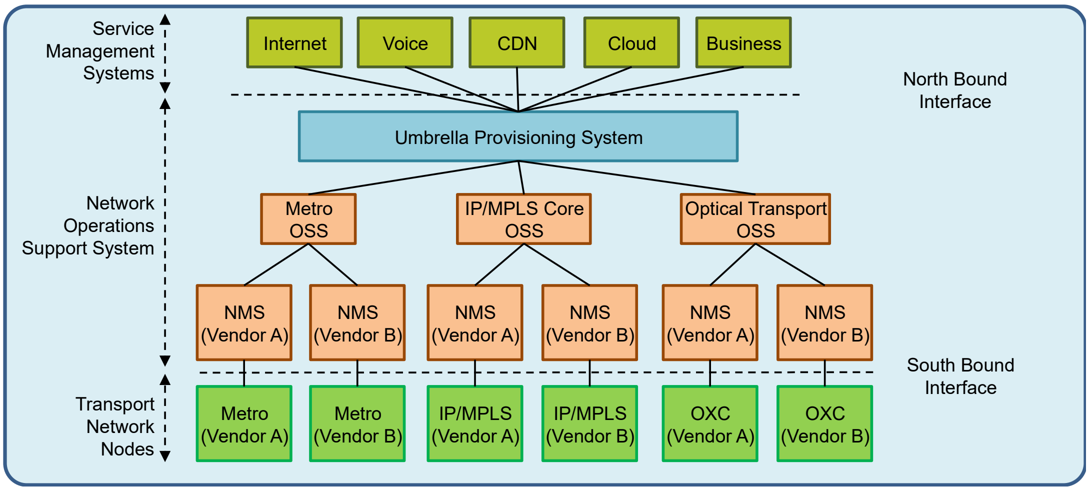


Figure 9 Network planning process in current networks

Once the need is clearly defined, there two options where the planning tool can interface the network: as a network tool and as a service. Figure 10 shows an architecture where the planning tools acts as network tool, which is a back-end PCE. On the other hand, the Figure 11 presents the planning tool as a service that is on top of the network control architecture.

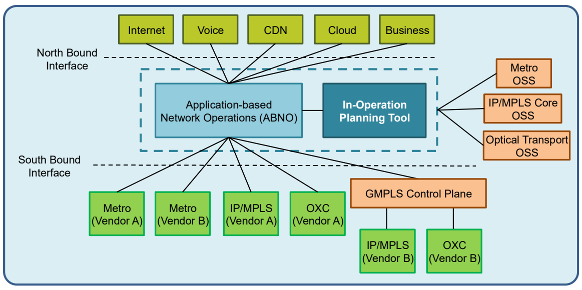


Figure 10 In-operation network planning as a network tool

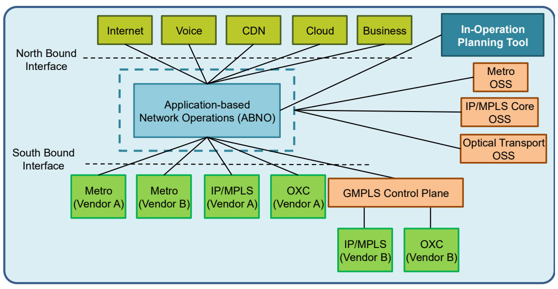


Figure 11 In-operation network planning as a service

There are three functionalities required on the north bound and/or south bound interface to support the in-operation network planning:

i. *Provisioning* enables the set-up, tear down and modification of connections in the network. Its most basic feature is to set up a point-to-point connection between two locations. However, there are other characteristics that a client interface can have like excluding or including nodes/links for traffic engineering, defining the protection level, defining its bandwidth or defining its disjointness from another connection. We include in this functionality the service synchronization.
ii. *Topology Discovery* requires, at a minimum, that the interface exports network topology information with unique identifiers. However, network identifiers (such as IPv4 or datapath-IDs) help to carry out path computation and to integrate the nodes for an end-to-end scenario.
iii. *Path Computation* is a critical and fundamental feature because each element can compute connections in a common way. Without a path computation interface, the ANM is limited to carrying out a crank-back process.

When considering the planning as a network tool, the ANM acts as a front-end PCE, while the planning tool is a back-end PCE. The front-end PCE acts on workflows where the computation is required for real time algorithms, while the back-end PCE is for algorithms more complex algorithms that require more computation resources. On this scenario where the planning tool is a back-end PCE, the interfaces to support this use case are clear defined based on the work done in IDEALIST. Figure 12 shows a detailed architecture of the ANM and the planning tool for the planning tool as a network service. The interfaces from the ANM to the planning tool are PCEP and BGP-LS. PCEP covers the functionalities of provisioning and path computation, while the BGP-LS enables the topology discovery information. The utilization of binary interfaces on this scenario is to interface on a faster time scale with the control plane.

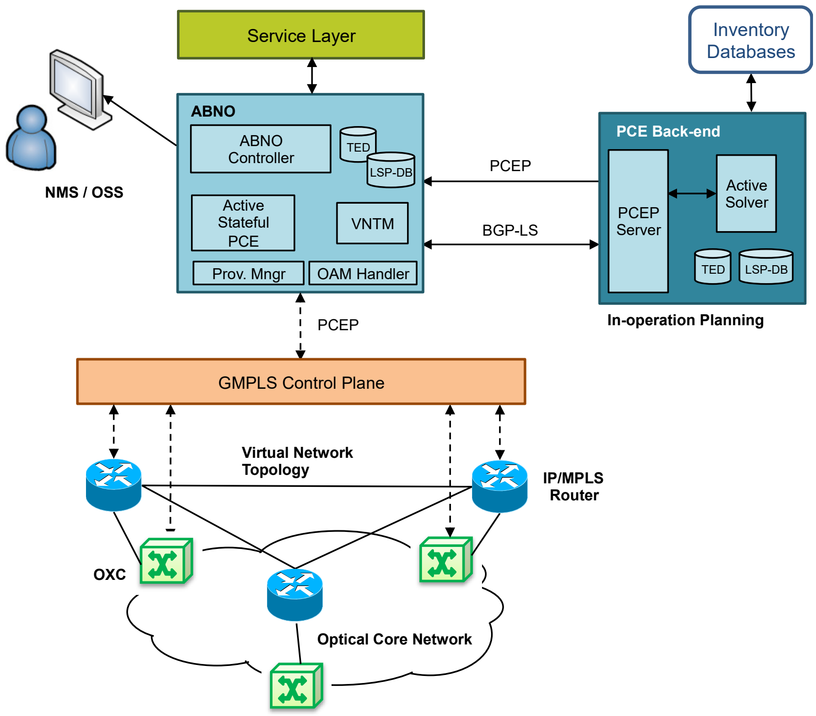


Figure 12 Protocols for an in-operation network planning as a network tool

When considering the planning as a service, the planning tool acts a service or application that is on top of the ANM. On this scenario, the planning tool works with a snapshot of the TED and the LSP-DB that retrieves from the network, operates with it and sends the commands to configure the network.

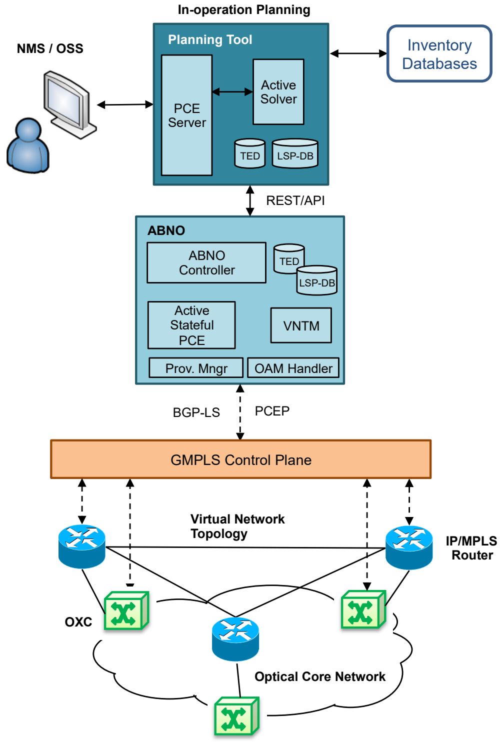


Figure 13 Protocols for an in-operation network planning as a service

To do so, an information model for the LSPs and the TED is required to exchange the information between the planning and the ABNO. Regarding LSPs, there is an initial work defined in the IETF [YANGLSPs] that cope with the information model required to sync the LSP-DB. As there are no standards defined as yet, the ANM and the planning tools syncs the LSP-DB using a YANG model as used in Figure 14.

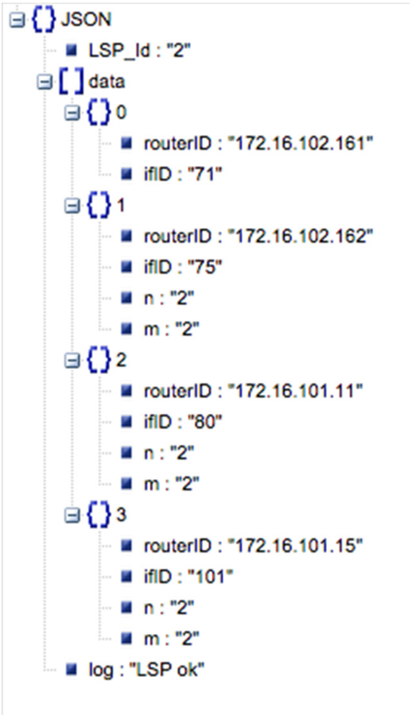


Figure 14 LSPs format to exchange information between the ANM and the planning tool

The format defined for the scope of the project is the following:

- LSP_Id. Unique identifier of the LSP in the network. It is given by the ANM.
- Data. It contains the LSP information with the ERO. Each of the elements in the data array is composed by the information of each hop. It can contain the following elements:
    - RouterID. Unique identifier of the network node. It is an IP on the IDEALIST context.
    - ifID. This optional value is the unnumbered interface of the element.
    - N. N is the frequency slot, as defined in flexigrid networks.
    - M. M is the spectral width, as defined in flexigrid networks.
- Log. This contains information about the state or the information during the provisioning process.

Within [D3.2], the YANG model proposed within IDEALIST was presented and explained in detailed. This work is being under standardization within the IETF [YANGFlexi]. With this information, the planning tool syncs the TEDB and the LSPDB information and using a REST/API. There is a detailed description of the YANG model within [D3.2].

3.1.3 Monitoring and OAM handler

In the proposed ANM architecture, the OAM Handler is key element for implementing the Dynamic Bandwidth Allocation use case. The ultimate goal of this use case is to show how flexi-grid networks can dynamically adapt themselves to changes in the traffic.

3.1.3.1 Monitoring architecture

The final architecture of the proposed monitoring scheme is shown in Figure 15. Passive monitoring probes are located in the edges of the network, at the gray interfaces of the routers. These probes are in charge of obtaining network flows, which are the main input to the bandwidth estimation algorithm described later. Flow records are forwarded to the OAM Handler, where the DBA Monitor obtains executes the bandwidth estimation algorithm and proposes the adequate bandwidth for each end-to-end link. According to the calculations of the DBA Monitor, the OAM Handler creates an MPLSProvisioning_DBA_WF message that is sent to the ABNO Controller.

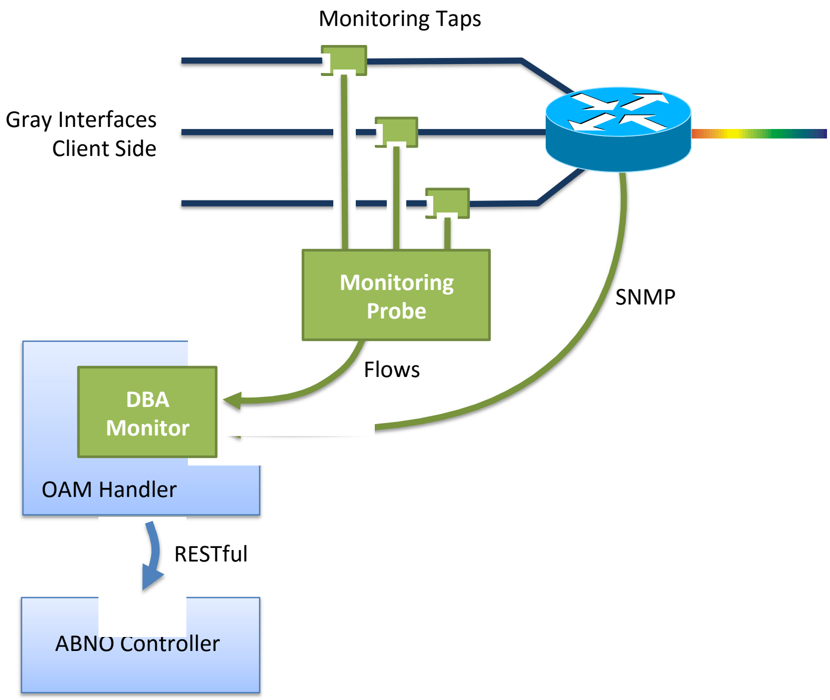


Figure 15 Architecture of the ANM Monitoring Scheme

Monitoring probes are required because the flow exporting capabilities of routers is limited. For example, for a Cisco CRS router the maximum packets per second for the flow exporter is 125,000 and the flow table is limited to 1,000,000 entries. Using just router-generated records could affect the accuracy of the bandwidth estimation algorithm, since it relies on the quality of the measurements being taken. However, the information provided by routers is also relevant, and in the proposed architecture it is envisioned as a complementary data source for the DBA Monitor.

As it was explained in deliverable [D4.3], monitoring probes are based on FPGA in order to ensure the performance needed to be able to collect network flows at 10+ Gbps. The latest version of the probes features a HW/SW solution where network traffic is captured and pre-processed by an FPGA in order to achieve an optimal equilibrium between performance and cost.

3.1.3.2 Bandwidth estimation algorithm

The main goal of the bandwidth estimation algorithm is to distinguish changes on the load of a link that are maintained over time from those who are spurious. The premises for designing the algorithm were:

- Ensure that the capacity of the link is bigger than the demand;
- Minimize the unused bandwidth; and
- Reduce the number of proposed bandwidth changes.

Unfortunately, a bandwidth time series typically exhibits high and unpredictable changes, which are usually neither related to changes in the number of users nor on how users exploit the link. An averaging process may be used as an attempt to remove such spurious changes [Pap03]. That is, the high frequency components are aggressively ruled out by means of iterative subsampling processes. In this regard, a wavelet-based processing provides more accuracy for networking-related time series, which show both temporal and frequency components [Gar08].

Alternatively, we support the hypothesis that the time series of concurrent flows reflects more genuinely changes on the sustained use of a link. To dig into this issue, we use traces obtained from a large corporate network where we provided our traffic analysis and monitoring services. Figure 16 shows the bandwidth time series for this reference trace.

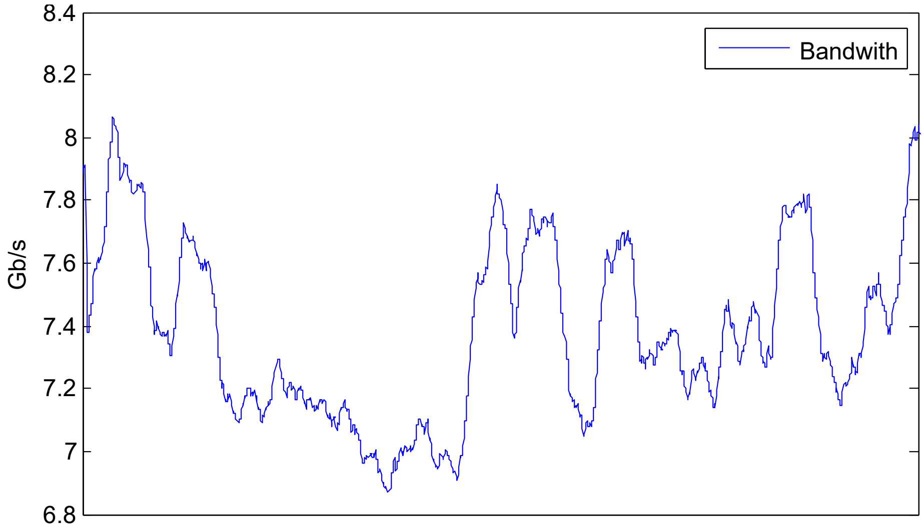


Figure 16 Bandwidth time series for the reference trace for the algorithm

Figure 17 compares the energy of both bandwidth and concurrent-flows time series, at one-second granularity, after an iterative wavelet processing. The figure shows that the bandwidth time series still exhibits a significant variance after 10-16 aggregation processes, while the concurrent-flows time series converges faster to a soft signal, roughly 6 times faster. In numerical terms, it means that only a little more than 1 minute of concurrent flows time series has to be aggregated in order to achieve a soft signal, while in the case of bandwidth such aggregation needs to be for intervals between 15 minutes and half a day in order to obtain a smooth signal. Finally, once concurrent-flows series is being used to detect sustained changes, the ratio flows/bandwidth is used to obtain link capacity.

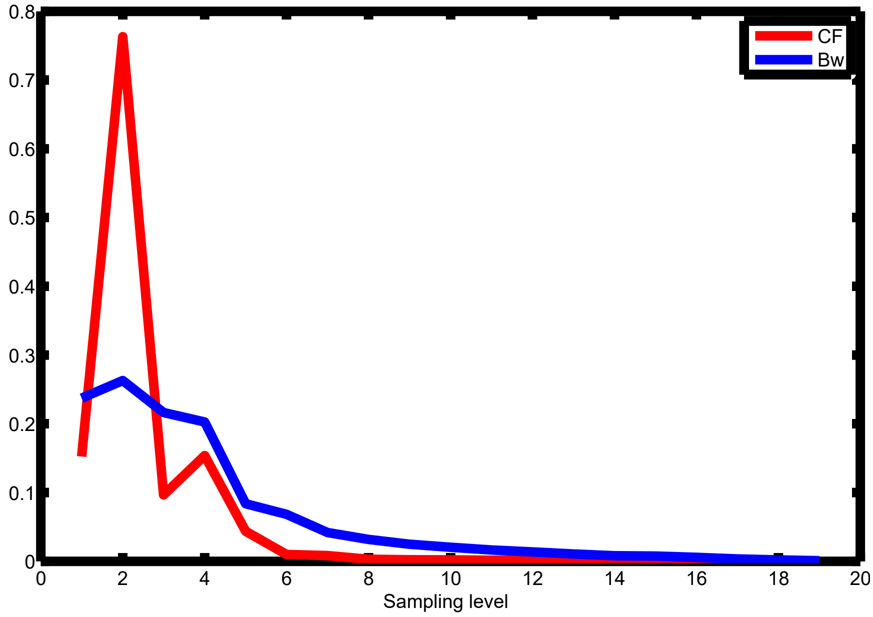
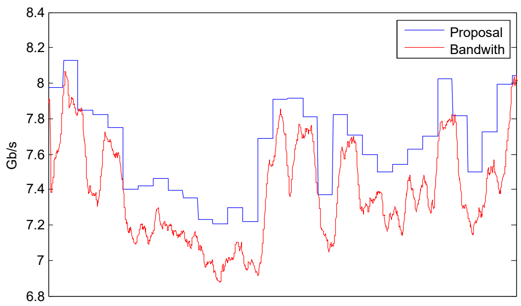
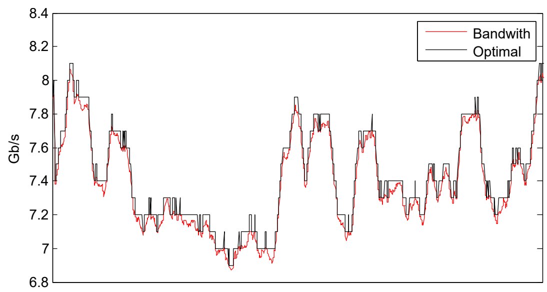


Figure 17 Energy of the concurrent flow time series (CF) and Bandwidth (Bw) as subsampling levels apply

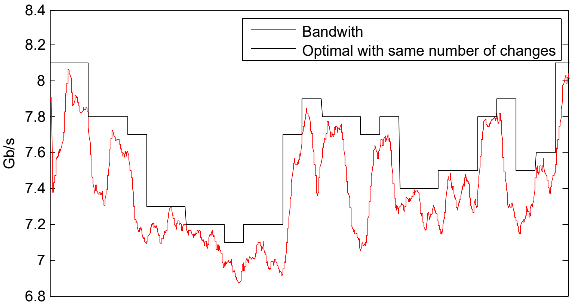
We have evaluated the goodness of our algorithm (Figure 18 a) by comparing it with three alternative approaches. First, we assume that the bandwidth can be known in advance and changes in bandwidth can be proposed at any time (Figure 18 b). Second, we also assume that the bandwidth is known but changes in bandwidth are limited in order to avoid frequent and costly network reconfigurations (Figure 18 c). Finally, we consider that bandwidth is not known beforehand and the estimation for capacity for the following time interval is the last sample after aggregation plus a block of bandwidth as guard interval (Figure 18 d). Of course, the best alternative is the unrealistic case of knowing the bandwidth in advance, with a relative error of 1.2%. However, when the frequency of changes is limited, error raises to 8.6%, higher than the one obtained by our algorithm, which is 5.9%. Finally, the approach based on previous bandwidth measurements shows an error of 13.6%, proving the benefits of the proposed algorithm based on the concurrent-flows time series.



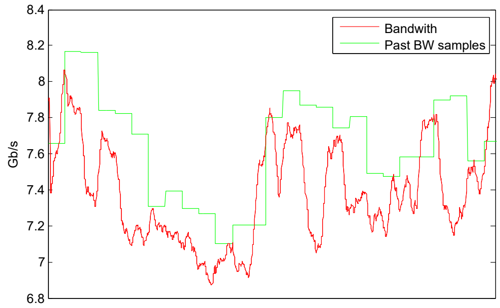
(a) Our proposed algorithm



(b) Optimal solution (bandwidth known in advance)



(c) Optimal solution (bandwidth known in advance) limited by the interval between changes



(d) Changes based on previous bandwidth measurements

Figure 18 Visual contrast between the original bandwidth time series and our proposal and the alternative set of approaches

3.1.3.1 Measuring throughput of links

As a complement to the bandwidth estimation algorithm described in the previous section, we have studied the problem of measuring throughput of high-speed network links. The goal is that the OAM Handler can actually check the quality of links, in a minimally invasive way. In order to fulfil this goal, we have evaluated the packet train technique, which is known to provide good estimates of throughput with a minimal overhead to the existing traffic. Usually, this technique is implemented using software techniques such as pktgen. However, the accuracy of software at relatively high speeds (>1 Gbps) is low. We tested FPGA-based implementations of the packet-train technique, finding that the accuracy is much better. Figure 19 shows a comparison between hardware (FPGA-based) and software implementations of the packet train for 10 Gbps. Further details can be found in [Ruiz15]

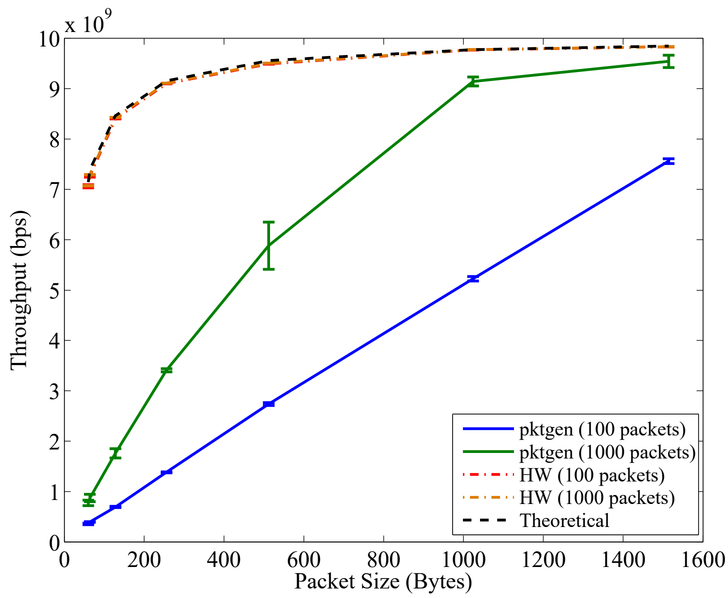


Figure 19 Comparison of hardware (FPGA) and software (pktgen) throughput estimations via packet trains for 10 Gbps

To summarize, the goal has been to probe that the FPGA-based monitoring probes can be used both to monitor traffic in order to estimate the optimal bandwidth (passive probe) and to measure the actual throughput of a network link (active probe)

3.1.3.2 Integration with ANM

IDEALIST's deliverable [D3.1] established the interfaces between OAM Handler and the ABNO Controller, and the workflow for the Dynamic Bandwidth Allocation use case. The RESTful API being used for the communication between the OAM Handler and the ABNO Controller were detailed in deliverable 3.2.

As a conclusion, the OAM Handler is integrated in the ANM by means of a RESTful interface to the ABNO Controller. The OAM Handler emits bandwidth change recommendations to the ABNO Controller, which executes them (or not) according to the ANM policies. As it was explained in D3.1, a decision to change a link bandwidth triggers a series of actions that finishes in the Provisioning Manager.

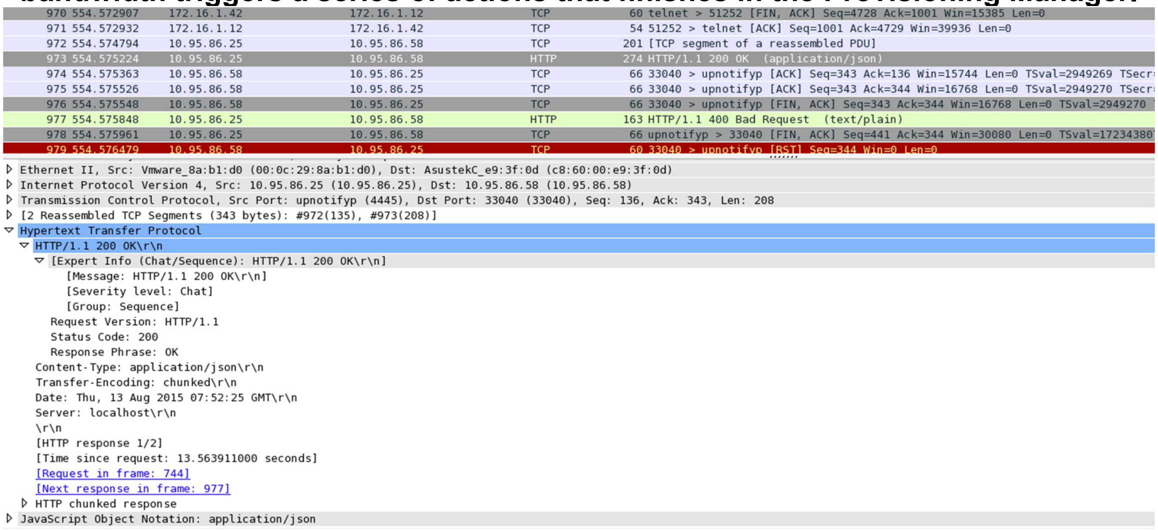


Figure 20 shows a Wireshark capture of the message being sent from the OAM Handler to the ABNO Controller in order to trigger a network reconfiguration.

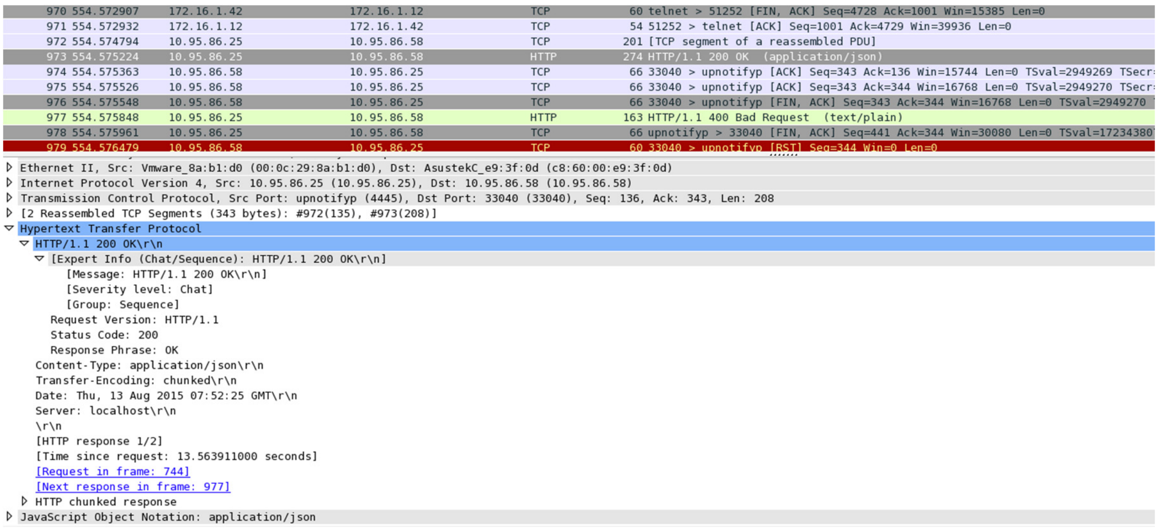


Figure 20. Wireshark capture of a bandwidth update message being sent from the OAM Handler to the ABNO Controller, and its response

3.1.4 ANM use cases final assessment

The use cases defined for the ANM was demonstrated in lab trials and using not only emulated GMPLS scenarios, but also commercial equipment. Following table presents n summary of the use cases and the mean of verification with the reports of the demonstration activities.

<table>
<tr><th>Use cases</th><th>Short Description</th><th>Mean of verification</th></tr>
<tr><td>Automatic IP Link provisioning</td><td>The use case describes how the ANM framework can be applied to the provisioning of an IP link between two routers. In this example, the photonic meshed network is composed of (elastic) ROADMs providing connectivity to several IP routers. This use case was tested in Telefonica labs and it was published in OFC and JOCN. Moreover, MPLS service configuration was demonstrated in the JOCN work</td><td>[Agu14] [Agu15]</td></tr>
<tr><td rowspan="2">Resilience Use Cases</td><td>Re-optimization after link repair use case considers that after some failures in the network, the backup paths can be intelligently adjusted by an ANM depending on the properties of the working paths. An experimental validation of the process was carried out.</td><td>[Gif14a]</td></tr>
<tr><td>For the multilayer restoration use case, there is not an evolution on the interface definition of the alarm system, so the test was not carried out in the lab. The ANM architecture supports the alarm management as demonstrated in ECOC 15.</td><td>[Agu15b]</td></tr>
<tr><td rowspan="2">Dynamic Bandwidth Allocation</td><td>Re-optimization process PCE-based strategy for LSP re-optimization in flexi-grid optical networks was experimentally validated. A front- and back-end stateful PCE architecture was used, automatic and coordinated operations were demonstrated using PCEP and BGP-LS protocols.</td><td>[Mar14]</td></tr>
<tr><td>Traffic variation use case was demonstrated in Telefonica lab using the Cisco equipment.</td><td>[D4.4]</td></tr>
</table>

3.2 Multi-domain control plane

Within the IDEALIST project, multi-domain E2E connectivity provisioning has been addressed by considering the H-PCE architecture and the GMPLS protocol suite.

In addition, the recently proposed BGP-LS protocol has been adopted to provide TE information to the pPCE, thus enabling efficient domain sequence computation and, according to the adopted policy, also intra-domain segment definition. Indeed, in addition to inter-domain links, selected topological elements can be provided to the pPCE, either representing an abstracted or complete view of intra-domain resources.

Specific extensions to BGP-LS for elastic optical networks are also introduced.

The proposed multi-domain control plane is designed to enable segment computation, selection and path concatenation, also including BVT end-point indication and configuration. Furthermore, instantiation, end-to-end RSVP-TE signaling and per-domain instantiation and stitching are addressed.

The proposed multi-domain control plane exploits the intra-domain solutions and extensions already defined in IDEALIST, such as the NCFS OSPF-TE extension (see D3.1 and D3.2).

Three main multi-domain scenarios have been specifically designed in WP3 for EONs.

In the first scenario, called StateLess-H-PCE with E2E signaling and instantiation (SL-E2E), stateless PCEs are employed to trigger, through PCEP instantiation, the end-to-end RSVP-TE signaling.

In the second scenario, called StateFul-H-PCE with Per-Domain instantiation and stitching (SF-PD), stateful PCEs with active capabilities are assumed, triggering per-domain instantiation exploiting intra-domain RSVP-TE signaling. This second scenario enables more advanced traffic engineering procedures (e.g. network re-optimization). However, this scenario may suffer for additional scalability issues at the pPCE. However, it is important to remark that the PCE architecture is not designed to be operated over the entire Internet and any PCE implementation has to be designed with an adequately dimensioned domain of visibility.

In the third scenario, optical / electrical / optical conversion is also considered at defined interfacing points (e.g. on domain borders), thus enabling the support of non-transparent E2E provisioning.

Additional details can be found in [Gon-JOCN15].The aforementioned scenarios have been successfully implemented and validated in WP4.

3.2.1 Hierarchical PCE

As detailed in previous WP3 documents, the H-PCE has been adopted to provide effective path computations for LSPs crossing multiple domains. In particular, the pPCE is responsible for domain sequence computation. Then, within each computed domain, a cPCE is responsible for segment expansion. Through BGP-LS, the pPCE is provided with an abstracted domain topology map, including the child domains and their interconnections. Moreover, besides reachability information, a mesh of abstracted links between border nodes can be introduced in the parent TED to improve the effectiveness of domain sequence computation.

The PCE has been considered with a stateless architecture, where only information on resource utilization is utilized. Within IDEALIST, the PCE architecture has been also enhanced to support the stateful capability, i.e. to store the attributes of the established LSPs (e.g. the route) within the LSP State Database (LSPDB). Furthermore, the stateful PCE architecture has been also enhanced to include the active functionality. The active stateful PCE issues recommendations to the network through PCEP, e.g. to dynamically update LSP parameters and enable advanced traffic engineering functionalities, including elastic LSP operations and global defragmentation in flexi-grid networks.

Both stateless and stateful pPCE architectures have been designed and investigated within the IDALIST project for multi-domain operations.

The two pPCE architectures are suitably combined with either the contiguous or the stitched inter-domain LSP provisioning technique, as detailed in the following.

3.2.1.1 Scenario I: single RSVP-TE session with transparent data plane

Figure 21 shows the main procedure for the provisioning of a multi-domain LSP when the stateless pPCE is assumed, and E2E signaling over transparent data plane is considered.

In this scenario, a StateLess-H-PCE with e2e signaling and instantiation (SL-E2E) is designed.

Upon ANM request (1), multi-domain path computation is triggered (2).

A two-step procedure is implemented. In the first step, the pPCE obtains the domain sequence; in the second step, each involved children is requested to expand the related segment.

This way, the pPCE obtains an end-to-end explicit route. To account for the interconnection with ingress ports and client interfaces, the ERO object including the unnumbered interfaces describing the outgoing TE links can be suitably augmented with additional IPv4 or unnumbered prefixes. The adopted Explicit Label Control (ELC) is applied to define the frequency slot to be used by the downstream node.

Upon the two-step path computation procedure has been completed (i.e. the E2E route identified), the actual LSP provisioning is performed (4). The PCEP PCInitiate and PCEP PCRpt messages are employed. More specifically, the PCInitiate includes the SRP, LSP, ENDPOINTS, ERO objects, and instructs the ingress node to initiate the signaling procedure, based on the Path/Resv RSVP-TE message exchange with an end-to-end session (5).

Upon completion of the signaling process, the PCRpt message is sent back to the provisioning manager, additionally including the route object RRO and the allocated frequency slot.

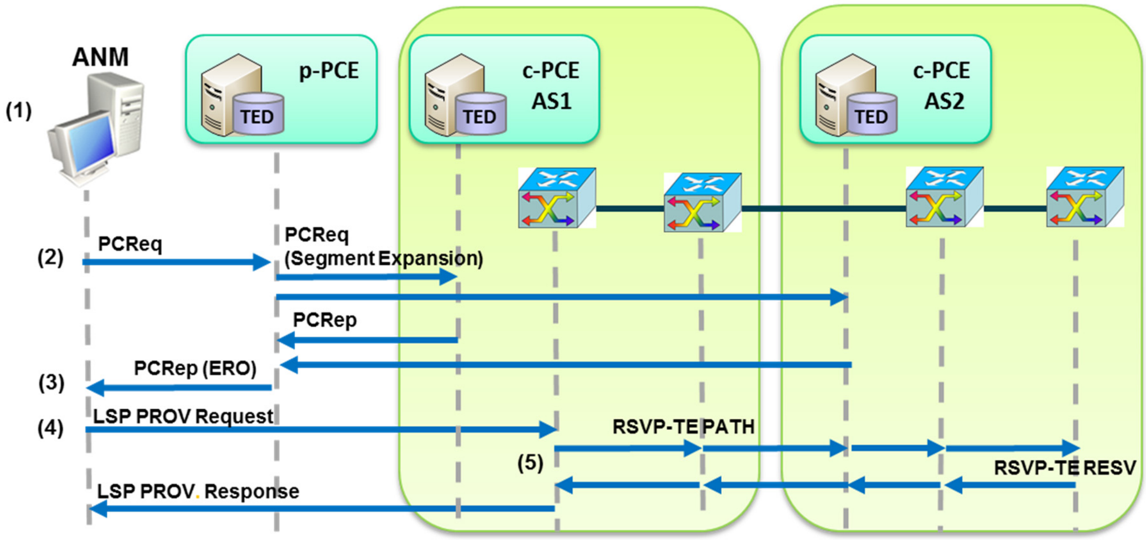


Figure 21 Stateless H-PCE with end-to-end signaling and instantiation (SL-E2E): architecture and message flow

3.2.1.2 Scenario II: segment concatenation and stitching

In the second scenario, Stateful-H-PCE with Per-Domain instantiation and stitching (SF-PD) is designed.

In this scenario, the transparent data plane connectivity is achieved by the concatenation of media channels at each domain border. Figure 22 shows the procedure in the SF-PD scenario: upon ANM request (1), the pPCE, by relying on its local TED, computes the domain sequence (2). Then, all the involved cPCEs are requested to expand the domain segment within their respective domains (3).

According to the information provided by the cPCE, the pPCE identifies a frequency slot accounting for the requested constraints in terms of slot width, the available network resources and the continuity constraint among the traversed links. The pPCE then performs Explicit Label Control (ELC) by adding a label subobject after every hop in the overall EROs. To properly perform the route and spectrum assignment phase, new extensions have been defined to enable the pPCE to receive from the cPCEs the spectrum availability of the inter-domain links via the BGP-LS protocol and the availability of intra-domain segment resources through extensions of the PCRep PCEP message utilized during segment expansion.

The obtained EROs representing each domain segment are then used inside the PCInitiate message sent by the pPCE to the respective cPCE (5). Each cPCE forwards the PCInitiate message to the domain ingress node (6) that acts as the signaling source, starting a RSVP-TE session limited to the local domain. Once the signaling procedure is completed, PCRpt message, conveying the session's status, is sent back to the domain cPCE that, in turn, forwards it to the pPCE (7).

Finally, the pPCE notifies the ANM with the results of the procedure (8).

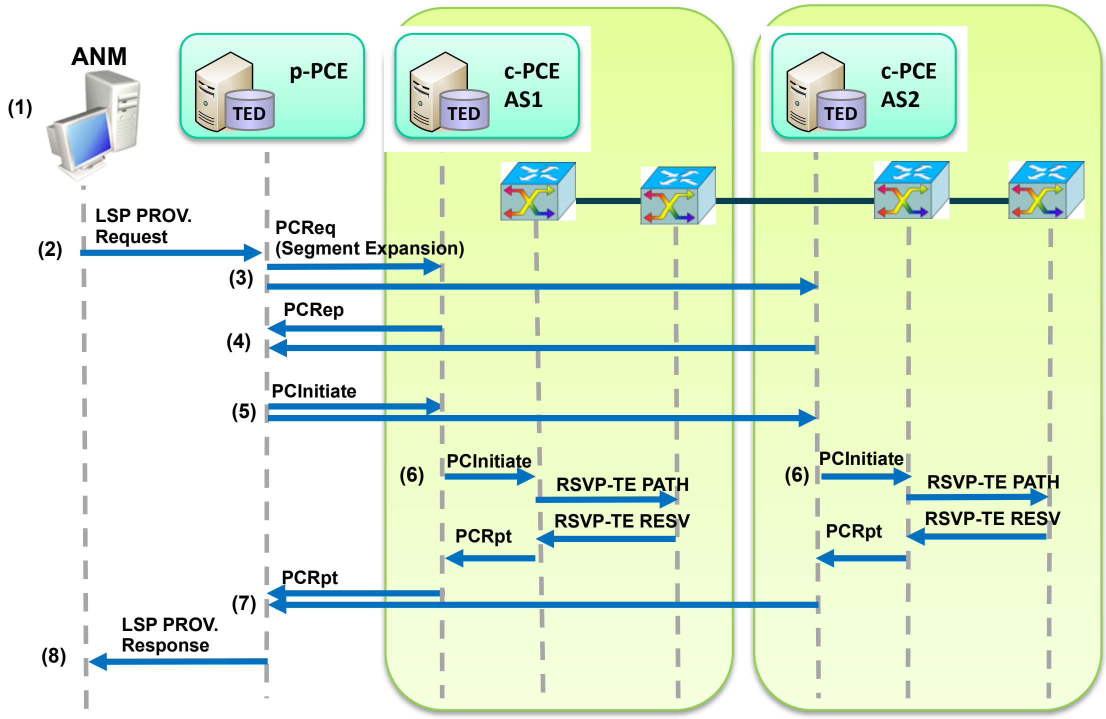


Figure 22 Stateful H-PCE and per-child instantiation (SF-PD): architecture and message flow

3.2.2 Layered approach with abstraction

This section addresses the multi-domain approach in a complementary way to the hierarchical setting, and combines centralized and distributed aspects, assuming, for this, that in each domain there is an "entry point" that can be a SDN controller or a PCE as above. Simple SDN control deployments are scoped to a single domain, where a (logically) centralized controller has full topology visibility of the network. Even if the network is segmented, e.g. for scalability reasons, a common deployment still considers a single controller. However, new use cases and scenarios call for more robust and scalable solutions, considering the concept of orchestration, abstraction and the use of multiple controllers, commonly arranged in hierarchical or flat configurations. Orchestration, roughly defined as the coordinated control of heterogeneous systems usually involving multiple interfaces and its related workflow, is becoming a key component of control and management solutions due to common use cases such as: i) multi-domain scenarios, networks segmented in vendor islands or heterogeneous control plane interworking (with SDN/OpenFlow and Generalized Multi-Protocol Label Switching, GMPLS potentially involving multiple controllers), ii) integration of cloud computing IT and networking resources with joint and dynamic provisioning of resources or iii) coordination of network segments such as access, aggregation and core transport. Abstraction, defined as the selection of an entity relevant attributes, is considered since although Traffic Engineering (TE) information is only contained within each domain for scaling and confidentiality, end-to-end services may require a constrained TE path, so it is efficient to expose a limited amount of TE information.

We consider a hybrid solution (Figure 23), based on a combination of centralized and distributed elements, where a mesh of controllers collaborate running a dedicated control plane instance using GMPLS protocols as their East/West (E/W) interfaces. Within each domain, different control approaches are possible. This approach complements, where we proposed a multiple controller model for overarching control and orchestration based on a hierarchy. We adopt the concept of TE reachability, so a subset of TE reachability information should be provided from each domain so that a client can determine whether they can establish a multi-domain TE path. The proposed architecture addresses several use cases such as multi-domain overarching control, migration scenarios, or support for GMPLS User Network Interface (UNI) for a SDN controller. It combines the benefits of the GMPLS and Automatically Switched Optical Networks (ASON) routing (robustness, scalability, stability) while locally enabling network operators to deploy SDN features in their scoped and controlled domains (control and data plane separation, logically centralized controllers, open interfaces and modular software designs, etc.) in view, for example, of progressive migrations.

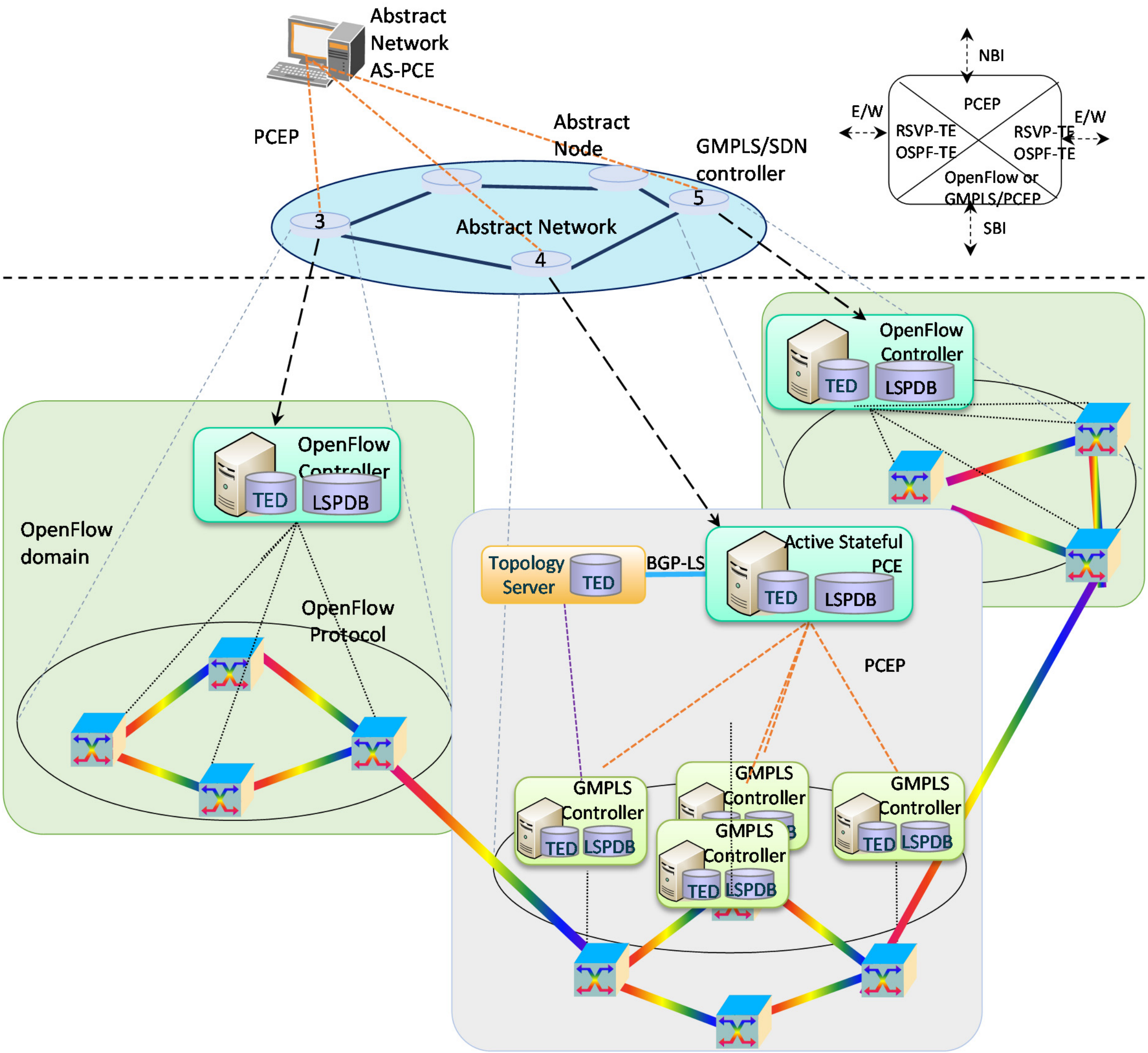


Figure 23 Overview of the hybrid architecture combining elements from distributed and centralized control

We consider a domain as a set of data plane elements (such as Reconfigurable Optical Add Drop Multiplexers or ROADMs) under the control of a technology specific control plane, assuming a deployment model of a single SDN controller per domain, with full topology visibility. Even if the domain deploys a distributed control plane such as GMPLS, we assume that an Active Stateful Path Computation Element (AS-PCE) acts as a proxy single point of entry. If we consider basic set of functions associated to an SDN controller, related to the separation of control and data planes, its centralized nature, the call/connection provisioning capabilities and aspects such as topology management, let us note that an AS-PCE is an SDN controller. It is able to provision Label Switched Paths (LSPs), delegating to an underlying GMPLS control plane, provided, for completeness, that it exports a certain North Bound Interface (NBI) Application Programming Interface (API) to applications, such a Network Management System (NMS) or orchestrator that requests the provisioning from the PCE.

The design of the overarching control and orchestration architecture is based on the following underlying principles and initial requirements, involving a trade-off between flexibility, scalability and availability. First, the main target for deployment of the architecture is a reduced number of domains (the applicability to a large number of domains is not considered), in the scope of one or few number of operators with peering agreements, the most common being a single operator that spans multiple domains. Second, each domain is constrained to its own control technology, due to several factors such as return on investment of existing deployments or vendor lock-in in which the SDN controller is tightly coupled to the domain elements. Within each domain, intra-domain control is up to the domain, abstracted to the higher layers: either OpenFlow, GMPLS or even based on a traditional NMS. Third, is the use of domain abstraction to enable scalability and hierarchical TE. Key to this is TE network abstraction, which is the synthesizing of reported TE attribute information for each domain and inter-domain links (abstract topology). This representation does not include all possible connectivity options, but instead provides a connectivity matrix based on attributes as reported from within each domain. The abstraction makes use of available TE information, but is subject to network policy and management. By selecting relevant TE attributes, a hierarchy of topologies can be constructed. Note the recursive property of the architecture, where the abstract network can be in turn placed behind a SDN controller or AS-PCE and be part of a wider orchestration and overarching control layer (i.e. to jointly provision IT and network resources).

Next, in view of the number, disposition and resulting mesh domains, the distributed GMPLS architecture and framework is selected for the abstract network layer. The choice is motivated by the maturity, robustness, and available feature set, including the ability of resource and neighbor discovery, path provisioning and recovery. A potential drawback of the approach is related to inter-operability. It has been argued that the interworking of GMPLS implementations of different vendors is problematic, due in part to proprietary vendor extensions; the complexity of the protocols -- notably when dealing with transport data planes such as Optical Transport Networks (OTN) -- and the different interpretation of standards. We believe that these issues can be mitigated to a large extent when considering network topology abstraction, in which the selected relevant TE attributes are mature and well understood. That said, inter-operability between vendors is not likely to disappear, in spite of efforts related to the definition on common hardware models and open and standard protocols.

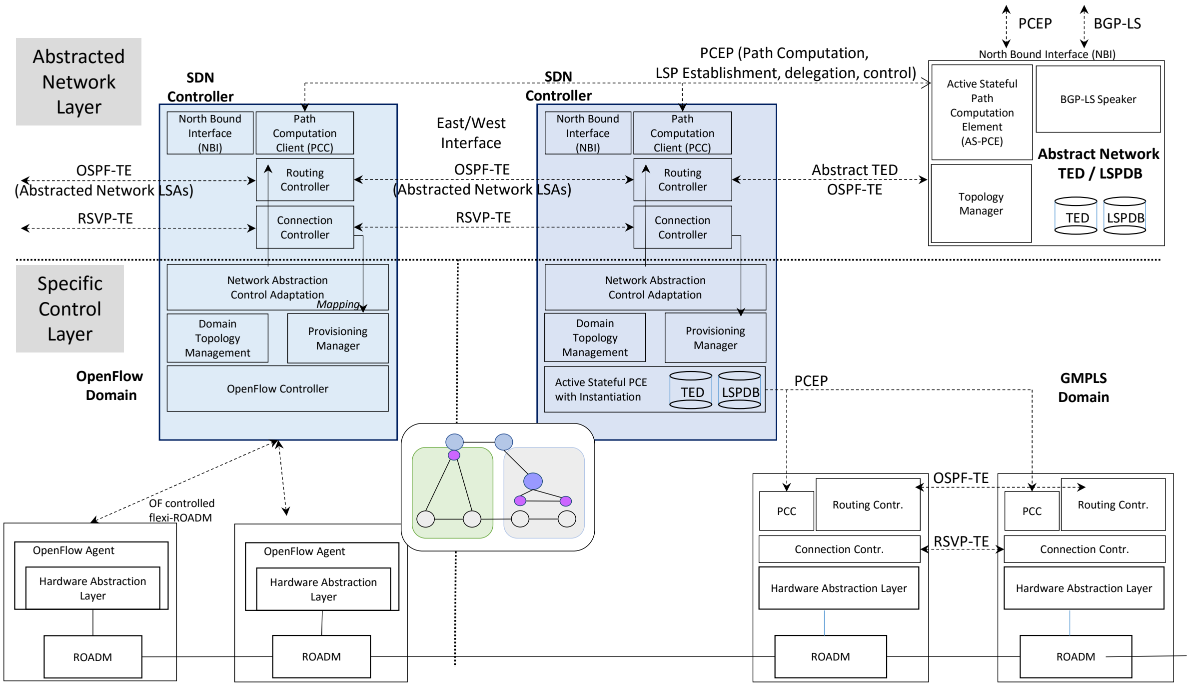


Figure 24 Hybrid control and orchestration architecture based on a mesh of SDN controllers forming an abstracted network layer

In view of the aforementioned considerations, the architecture (cfr. Figure 24) is thus based on a mesh of SDN controllers that implement GMPLS routing and signaling as their East/West interfaces. In other words, the GMPLS Open Shortest Path First with Traffic Engineering extensions (OSPF-TE) routing protocol and the Resource Reservation protocol with Traffic Engineering extensions (RSVP-TE) signaling protocol are selected as Inter-SDN controller communication. In the proposed architecture, cooperating SDN controllers are responsible for the abstraction of their domains (in terms of announcing an underlying abstract topology) and for adaptation of the control procedures, which remains technology specific. Such GMPLS enabled SDN controllers (GSC) maintain a domain abstract topology map and execute a separate GMPLS instance in which routing and signaling adjacencies are congruent to the domain connectivity. Each GSC disseminates the abstraction of its domain by means of dedicated OSPF-TE Link State Advertisements (LSAs). An AS-PCE placed on top of the abstracted network provides the functions of multi-domain path computation and multi-domain network service provisioning, operating on the abstracted Traffic Engineering Database (TED). Let us note that there may thus be several independent AS-PCE instances: one for the abstract network layer, and one for each domain that is controlled by a GMPLS control plane.

In view of the choice of this E/W interfaces, each GSC/SDN controller has thus a Path Computation Client (PCC) module that interacts with the AS-PCE (for path computation, LSP establishment and LSP control delegation), a Routing Controller (RC) that uses OSPF-TE and a Connection Controller (CC) that participates in signaling sessions. The part of the GSC corresponding to the specific control layer is responsible for network abstraction and control adaptation, address and resource identifier mappings and process adaptation.

3.2.2.1 Control Plane procedures and message flows

The involved GSCs disseminate OSPF-TE LSAs describing the abstract topology, as defined by their domain internal policy. This allows the AS-PCE in the abstract network to obtain a

TED of the abstract network layer. The workflow for the provisioning is described within Figure 25. An end-to-end provisioning is triggered by the AS-PCE in the abstract network, e.g. by sending a PCE communications Protocol (PCEP) PCInitiate message to the ingress abstract node, and relying on distributed signaling over the Data Communications Network (DCN) formed by the controller mesh. The ingress GSC uses an end-to-end RSVP-TE signaling session, indicating the Explicit Route Object (ERO) as provided by the AS-PCE, and the SESSION address being set to the destination abstract node. Each GSC processes RSVP-TE signaling messages, mapping resource reservations to actual segment connectivity requests, which are, in turn delegated to the underlying, specific technology (e.g. a cross-connect at an abstract nodes between input and output ports induces a segment connection in the corresponding domain).

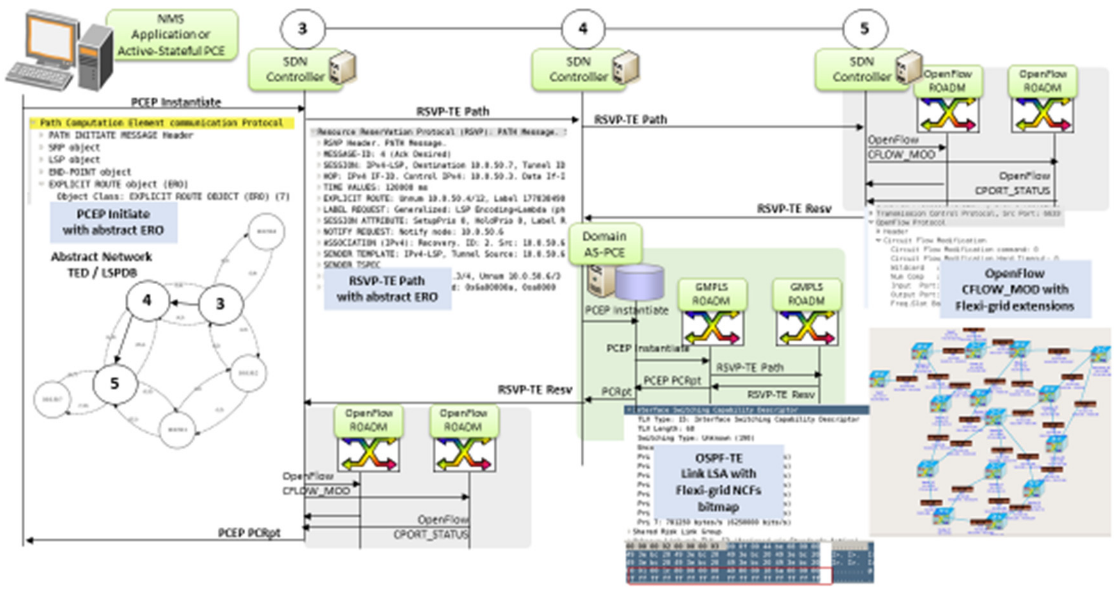


Figure 25 Example of message flow for the provisioning of end-to-end connections, showing the end-to-end RSVP-TE session and the intradomain segment establishment (simplified with 3 domains)

In particular, a receiving GSC processes the RSVP-TE PATH and RESV messages and identifies the involved abstract TE links, from the PATH RSVP_HOP and ERO objects, mapping such abstract link identifiers to its domain client port addresses, as defined in its topology aggregation policy. The sender descriptor traffic parameters (SENDER_TSPEC) conveys the requested frequency slot width and, if available, SUGGESTED_LABEL objects may specify the upstream domain preference, whereas the RESV GENERALIZED_LABEL contains the downstream allocated frequency slot. The underlying segment is established either using GMPLS/PCE control plane or OpenFlow, since the GSC South Bound Interface (SBI) is technology specific.

For an OpenFlow domain, the GSC configures the switching behaviour of the network elements using a modified OpenFlow for flexi-grid networks, directly programming cross-connections in the domain network elements. For a GMPLS/AS-PCE domain, the approach is applied recursively, thus having an additional RSVP-TE session that is scoped to the domain. To follow standard GMPLS procedures, the RESV message at the abstract layer is sent upstream upon successful completion of the establishment of the segment. The architecture does not preclude additional methods within a domain, as long as the sequential ordering is respected. The end-to-end service is thus the result of stitching the segments established within each domain. In case of errors, standard methods apply: a GSC may send a PATH_ERROR message upstream and a PATH_TEAR message downstream to release the resources. GMPLS Recovery applies without change at the abstract layer (e.g. supporting break-before-make rerouting). Finally, the successful provisioning of the end-to-end service is reported via a PCEP Report (PCRpt) message back to the AS-PCE.

3.2.2.1 Experimental performance evaluation

In order to validate the proposed architecture and to obtain performance indicators, we have deployed a control plane testbed with emulated optical hardware. This implies that each GMPLS controller or OpenFlow agent located in a data plane node accepts configuration requests via its Hardware Abstraction Layer (HAL) and assumes the underlying hardware is properly configured without error. Each of the involved AS-PCEs, GMPLS controllers as well as the OpenFlow controllers and OpenFlow agents is running in dedicated (physical) GNU/Linux servers with Core2 Duo processors and 1-4 Gb RAM, a hardware that suffices for the purpose. The servers are running in a LAN, with dedicated VLANs and GRE tunnels emulating the DCN.

The testbed has been configured as a multi-domain network with 5 domains. Each domain is controlled by a GSC that performs single TE node network abstraction, using OSPF-TE LSAs to disseminate node identifiers (domain X has 32 bit address 10.0.50.X) and link TE attributes. Additional nodes representing GMPLS clients/endpoints - 6 and 7 - have been added, resulting in the abstract topology. Domains 3 and 5 are optical OpenFlow islands with 4 core nodes each and three additional OpenFlow datapath ids that represent endpoints of the OpenFlow network. On the other hand, domains 1, 2 and 4 are GMPLS domains with an AS-PCE to instantiate LSPs. The biggest domain, domain 4, has 14 nodes representing a Spanish-wide optical mesh network, as shown in Figure 26. Let us note that the arrangement of OpenFlow - GMPLS - OpenFlow domains is motivated by a particular use case in which e.g. remote data centers controlled by OpenFlow are interconnected by a GMPLS controlled core transport network. Domains 1 and 2 in this context provide path diversity. All links are considered basic flexi-grid optical links with finest granularity in the selection of nominal central frequencies and frequency slot widths. All links are homogeneous having 128 usable nominal central frequencies, ranging from with n parameter n=0 to n= 127. The BV-ROADMs do not present any asymmetry and are assumed to be able to switch from any incoming port to any outgoing port.

Each GSC is responsible for mapping addresses and identifiers (such as endpoints or unnumbered interfaces) from the address space of the abstract network to its underlying domain-local address space.

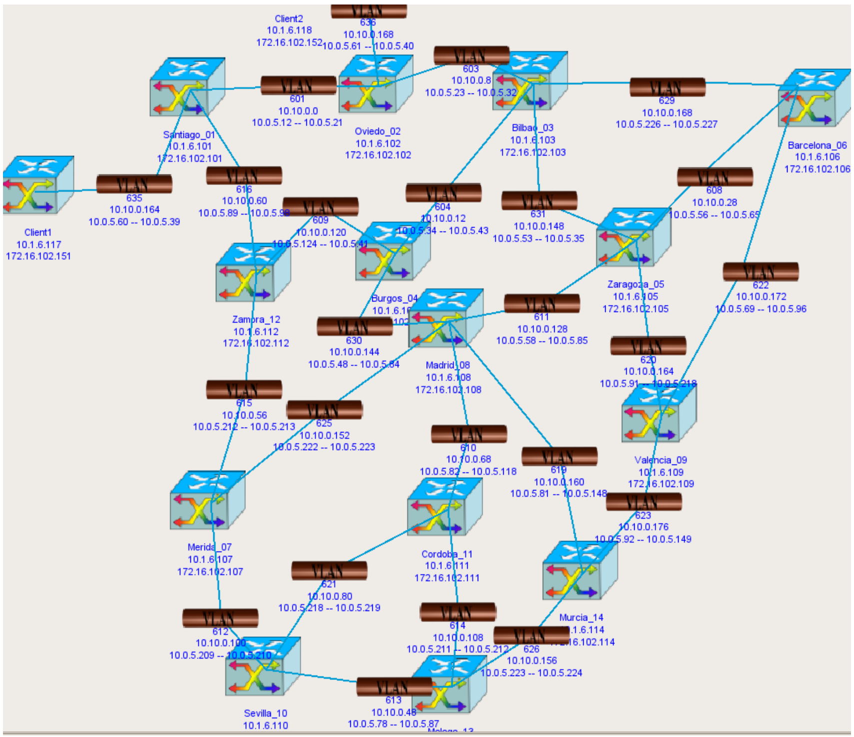


Figure 26 GMPLS controlled network emulating a Spanish 14-node topology

To test the system, connections are set up from Endpoint A (represented in the abstract network as node 10.0.50.6 and corresponding to OF datapathid 00-05 in domain 3) to Endpoint B (represented in the abstract network as node 10.0.50.7 and corresponding to OF datapathid 01-06 in domain 5), for a frequency slot of width m=1. At each domain, upon reception of the RSVP-TE RESV message, the GSC needs to actually expand the segment and establish the underlying connection, mapping the procedures to the underlying control plane. In particular, when the GSCs located in Domains 3 and 5 receive the RESV message, it sends OpenFlow CFLOW_MOD messages to the OF agents located in the nodes and gets OpenFlow CPORT_STATUS messages acknowledging the configuration. When the GSC located in Domain 4 receives the RESV message, it applies the initial process recursively: the GSC requests the AS-PCE in that domain to setup the segment, which proceeds by sending a PCInitiate to the head-end node of that domain, and sending a PCRpt upon completion. It is worth noting that the configuration of ROADMs in OpenFlow domains can be parallelized, as opposed to GMPLS in where it is serialized by the order of processing of RESV messages.

3.2.2.1 Setup delays

Regarding setup delays, there are several significant values. Starting from the service delay seen by an application or NMS on top of the AS-PCE of the abstract network layer, averaging 180ms, which includes two main components: the PCEP interface to the ingress node, with the PCInitiate and PCRpt exchange, and the actual signaling delay seen by the ingress node, measured from the sending of the PATH message to the reception of the RESV message, which is, on average 155 ms. This value is, in turn, the result of the sequential provisioning that happens at each domain. In particular, the highest value corresponds to Domain 4 (109 ms) seen by GSC 4, that covers the whole PCEP handshake with the AS-PCE in the domain and the reception of the PCRpt (note that this value could be further reduced if a persistent PCEP connection was used). The setup delay purely from the point of view of the AS PCE in the Domain 4 (from the PCInitiate to the PCRpt from the head end node in the segment) is 30 ms in average, with the actual segment provisioning involving 4 GMPLS nodes (16 ms of PATH/RESV signaling exchange). Segment provisioning in OpenFlow domains is the lowest value, averaging 12 ms and a result of the inherent parallelization in the domain.

The histogram of the path setup delay and the Cumulative Distribution Function (CDF) can be seen in Figure 27; the average setup value seen by an application using the NBI of the PCE in the abstract layer is around 180 ms (of which 155 ms correspond to the actual signaling delay seen by the ingress/head end abstract node). Let us note that flexi-grid optical hardware may require a significant amount of time (can reach the order of seconds) for configuration and optical connections may need to be validated before actual traffic is transported on the provisioned optical channel. Consequently, values are provided as a performance indicator, but real deployments would definitely show higher values.

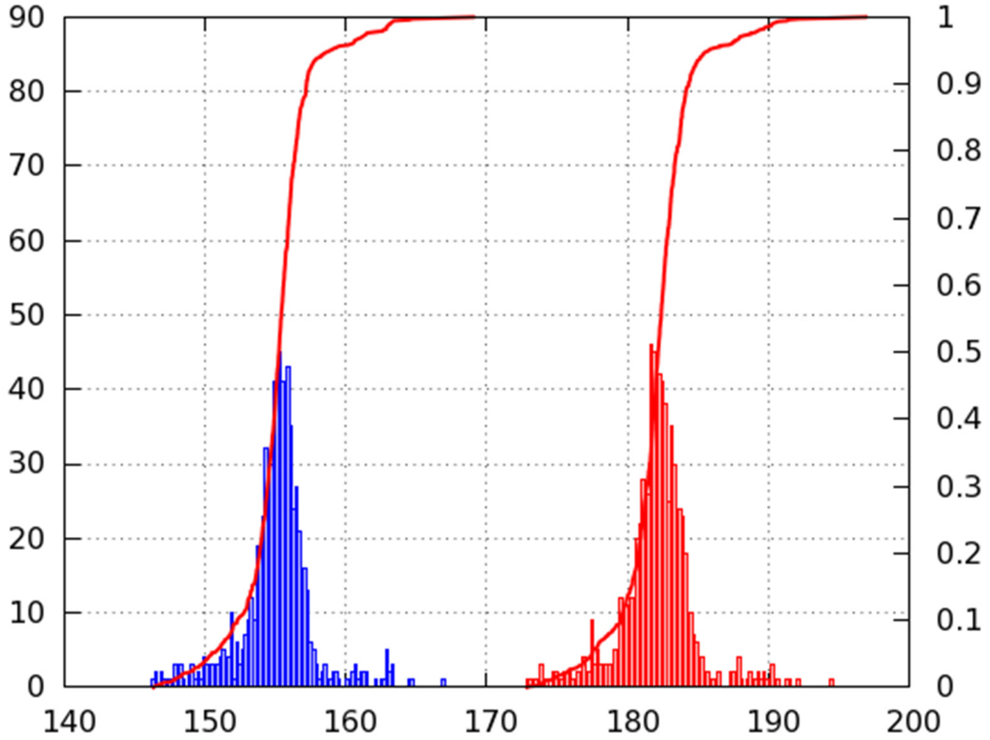


Figure 27 Setup delay (as seen by the abstract layer GMPLS instance and by the PCE) and CDF

3.2.2.2 Control plane overhead and processing requirements

Finally, another performance indicator is the control plane overhead, including not only the overhead in terms of required bandwidth in the control plane network due to message exchanges, but also in the processing requirements and resources related to session management notably in centralized elements. Generally speaking, the main factors to determine the overhead are the rate of arrivals of requests, the centralized or distributed nature of the involved elements and aspects such as the intra-domain topology aggregation policies.

Starting with the abstract layer, and for service requests arriving every few seconds, there are no significant issues regarding control plane overhead, but it is worth noting the effect of the topology dissemination delay which may cause the AS-PCE to act upon outdated information if the rate of changes is high. In this case, a conservative policy of aggregating domains as nodes significantly reduces the overhead when compared to policies (e.g. virtual links reflecting domain internal connectivity) that require complex or frequent re-computation of exported TE attributes due to internal changes. The overhead is then due exclusively to the periodic refreshes of LSAs, changes in the inter-domain links and the signaling sessions. For targeted scenarios involving a few domains the system scales reasonably. Generally speaking, assuming dedicated control plane links, the required control plane bandwidth is of a few Mb/s, resulting from the having message sizes rarely bigger than a few hundred bytes. To illustrate this, a single LSP provisioning results in the head end abstract node having to process PCEP initiate and report messages (272, 252 bytes) RSVP-TE PATH (284 bytes) and RESV (228 bytes) with ACK (56 bytes each) in 155 ms, with an average of ~40 packets per second resulting in 0.06 MBit/s throughput. In the GMPLS domain 4, during the window of a segment provisioning upon a reception of an abstract layer RESV message from domain 5, abstract node 4 sees 9 OSPF-TE LS Updates (248-444 bytes) resulting from the reception of LSAs generated by downstream abstract nodes and the generation of their own interdomain links totaling 2832 bytes with an aggregated throughput of 0.183 MBit/s.

Within each domain, there is a clear trade-off: for GMPLS domains, message overhead and control bandwidth is notably higher than the centralized OpenFlow domains due to the additional synchronization required for OSPF-TE LSAs and the signaling exchanges, but OpenFlow domains impose a higher stress in centralized controllers due to the need of processing all the messages and updating the internal databases. The overhead and scalability limitations are increasing with the domain size due, notably, to the number of managed sessions: for the AS-PCE it has to maintain N PCEP sessions with the edge nodes and for the OpenFlow controller it must maintain M OpenFlow connections with all the network elements. In the considered topologies, OpenFlow domains must typically configure 2 flexi-ROADMs involving 6 messages (2 cross-connection requests and 4 link notifications) which, in the worst case, happens in a 2ms window exchanging 924 bytes and resulting in 3.84 MBit/s peak rate.

In summary, the obtained values are within expected ranges for the considered scenarios and topologies, and dedicated control interfaces commonly implemented in terms of Ethernet interfaces are largely sufficient. The main bottlenecks remain in the processing capabilities of the nodes and the propagation delay of TED updates and synchronization, although mitigated by the expected traffic patterns.

3.3 Multi-Layer solutions

As a key target within IDEALIST, introducing automation and dynamicity on the provisioning and release of multi-layer and multi-domain end-to-end (E2E) connectivity services is a major concern of network operators in order to increase the revenues of their deployed network infrastructure. Several approaches such as based on Software Defined Networking (SDN) or Segment Routing (SR) have appeared as new networking paradigm to drastically reduce the cost of operation and reduce the time-to-market of introducing a new service, by increasing the level of automation on the control and configuration of the network infrastructure.

The need of assure the Return-Of-Investment (RoI) of existing network infrastructure suggests in the mid-term an hybrid scenario where traditional control and management systems, such Generalized Multiprotocol Label Switching (GMPLS), will coexist with the emerging technologies.

As mentioned in the previous section, the approaches to address multi-layer are several: from the relationship between the signal and media optical layers in the ITU architecture, to more research oriented approaches commonly based on orchestrating the different packet/optical layers via e.g. ANM/ABNO and interacting with each of the layer control plane instances. This covers using an IP/MPLS control plane, using OpenFlow as a data plane configuration protocol or Segment routing, as detailed in the following sub-sections.

3.3.1 Signal mapping

The Elastic Optical Transport Network envisioned in Idealist is built of several layers. From a bottom up approach, the lowest layer is the so called “media” layer, which deals with the pure spectrum. The connections handled by the control plane are media channels, which are explained in detail in WP3. The next layer is the so-called “signal layer” that handles the optical signals. Thus, the media channels carry Optical Tributary Signals, the term used by ITU-T to refer to the optical signal that conveys the information. The next layer in the hierarchy is the digital layer, i.e. the electrical signals. For example, this layer considers OTU signals or Ethernet frames.

In summary, the data plane of an elastic Optical Transport Network is, at a minimum, a three layer problem. The control and management plane has deal with this complex layering scenario. However, this does not imply that each layer has its own routing and signaling mechanisms. In theory, each layer could maintain its own topology and signaling state, but that would be a big overhead for the control plane devices. What is required is a clear mapping between the layers. Two main mappings are needed:

- Data rate (Gbps) to signal mapping. A request of X Gbps between two transponders has to be mapped to a given signal, characterized by the modulation format. WP1 and WP2 have defined a set of possible mappings between data rates.
- Signal to media channel mapping. Each signal has a requirement in terms of media channel. In the most basic form, the signal needs a minimum spectrum width. More complex mappings would set requirements in the filter shape.

3.3.1.1 End-to-end Signaling

The control plane can handle requests of:

- Media channels, characterized by an ”m”
- Connections from transponder to transponder (BVT to BVT, or S-BVT to S-BVT)
- With a fixed signal (modulation format, speed, etc),
- With possibility of selecting the modulation format.

For the first case, the route just needs to contain details of the media channel. For the second case, in the signaling session it is needed to include in the Explicit Route Object (ERO) the transponder information at the edges, and in the middle the sub-objects that describe the media channel. Section 5.1 has shown the set of extensions to describe the properties of a multi-flow transponder, which include details about the signal (or signals) generated by the transponder.

3.3.1.2 Path Computation and routing

The path computation becomes a multi-layer problem. The entity in charge of the path computation needs to have an access to the capabilities of the transponders and mapping of the specific optical information:

- Which modulation formats are supported by each transponder?
- Which data rates are supported by each transponder?

Such information needs to be distributed by the routing protocols. The PCE needs to have access to the mapping information.

3.3.2 Multi-Layer provisioning based on ANM/ABNO

The multi-layer provisioning use case on the ABNO was defined as automated IP link provisioning and as MPLS Service provision. The automated IP link provisioning between two routers which are connected via an optical network composed of (elastic) ROADMs providing connectivity to several IP routers. The IP link provisioning is a basic operation done by network operators. This operation is used to provide all customer services, for instance Internet connectivity, VPN or IPTV.

The provisioning of an MPLS service is a complex process for operators today. Nowadays, operator creates these connections to provide MPLS tunnels between two routers in the network. This process can be automated using ABNO architecture, but even when there is no bandwidth in the IP/MPLS layer, the ABNO can request to the optical layer to create a new IP link between two locations and provide this capacity to the MPLS layer. The provisioning of an MPLS tunnel between routers giving and administrative bandwidth is a second use case, but covers the IP link provisioning use case when there is not enough bandwidth for the MPLS tunnel. Figure 28 presents the interaction of the modules in the ABNO architecture to set-up an IP/MPLS Service.

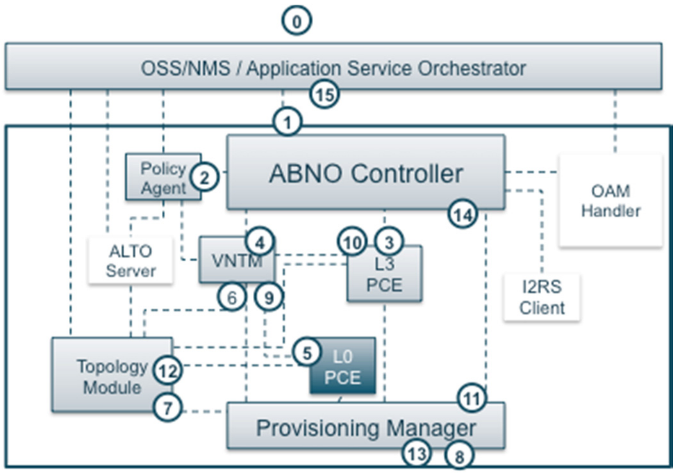


Figure 28 Interactions of the ABNO architecture for IP/MPLS provisioning use case

Each step is following explained:

1. The initial trigger for this use case is a request from an end user via a NMS.

2. The NMS sends the information to the ABNO controller with the two LSRs (R1 and R2, i.e. the head end and the tail end of the optical LSP) and the traffic parameters (e.g. bandwidth) required for the circuit.
3. When the ABNO controller receives the request, it asks the L3 PCE for a path with enough bandwidth in the MPLS topology. If there is not a path in the MPLS topology a NO PATH is sent to the ABNO. If there is a path move to step 10.
4. As there is no path or not enough bandwidth in the MPLS topology, the ABNO controller tells VNTM to create a link between the two LSRs (R1 and R2).
5. The VNTM requests the L0 PCE a path between the two optical nodes connected to the two routers (R1 and R2). If the L0 PCE is able to find it, it sends the path back to the VNTM.
6. VNTM has the interlayer information and the L0 path. This information is sent to the Provisioning Manager to configure the link.
7. The Provisioning Manager queries the Topology Module for the description of each node.
8. Depending on the technology and configuration mode, Provisioning Module selects a different protocol to complete the request. Same configuration interfaces are used than for the IP Link provisioning.
9. Once the path has been established, the Provisioning Manager notifies the VTNM. Similarly, VTNM notifies the ABNO controller.
10. ABNO controller asks the L3 PCE for a path with enough BW in the MPLS topology. Now there is enough bandwidth to cope with the request, so the L3 PCE responds to ABNO with the path.
11. The ABNO controller requests the Provisioning Manager to configure the MPLS service/layer by establishing the path.
12. The Provisioning Manager queries the Topology Module for the description of each node.
13. Depending on the technology and configuration mode, Provisioning Module selects a different protocol to complete the request. Same configuration interfaces are used than for the MPLS/IP Link provisioning. Figure 29 shows the case for multi-layer, where there are IPs of the routers for the tests (10.95.X.X) and IPs of the optical nodes 172.16.1.X. The provisioning manager splits the operations for IP and for optics, based on the topology module information. First the optical layer is configured and, secondly, the IP nodes.
14. Once the path has been established, Provisioning Manager notifies the ABNO Controller.
15. Similarly ABNO controller advertises the NMS.

```
▼ EXPLICIT ROUTE object (ERO)
    Object Class: EXPLICIT ROUTE OBJECT (ERO) (7)
    0001 .... = Object Type: 1
  ▶ Flags
    Object Length: 68
  ▶ SUBOBJECT: IPv4 Prefix: 10.95.73.72/0
  ▶ SUBOBJECT: Unnumbered Interface ID: 172.16.1.40:1
  ▶ SUBOBJECT: Label Control
  ▶ SUBOBJECT: Unnumbered Interface ID: 172.16.1.36:1
  ▶ SUBOBJECT: Label Control
  ▶ SUBOBJECT: IPv4 Prefix: 172.16.1.34/32
  ▶ SUBOBJECT: IPv4 Prefix: 10.95.73.74/0
```

Figure 29 ERO detailed for the IP/MPLS provisioning use case

3.3.3 Multi-Layer TE with OpenFlow and GMPLS

This section covers the work related to multi-layer studies considering Traffic Engineering enforcement in multi-domain SDN orchestration of Multi-Layer (packet/optical) networks, and how an SDN architecture based on ABNO can provision and monitor data services combining packet and optical layers. The work here was published in [May15].

We can assume a network comprised by multiple control instances which need to be orchestrated in order to achieve the so-desired dynamicity on the management and control of E2E services.

Network infrastructure monetization depends of efficient control of the traffic flow. Traditionally, this has been achieved by the introduction of Traffic Engineering (TE) policies to take benefit of highly optimized routing algorithms which fully exploit the network capacity. In SDN, the TE-enabler mechanisms have been not fully researched, neither the impact that TE policies can introduce into the jointly orchestration of multiple network domains comprising different transport technologies. For instance, in OpenFlow (OF)-based SDN deployments, traffic differentiation and traffic rate limitation were not enabled until 1.3 version.

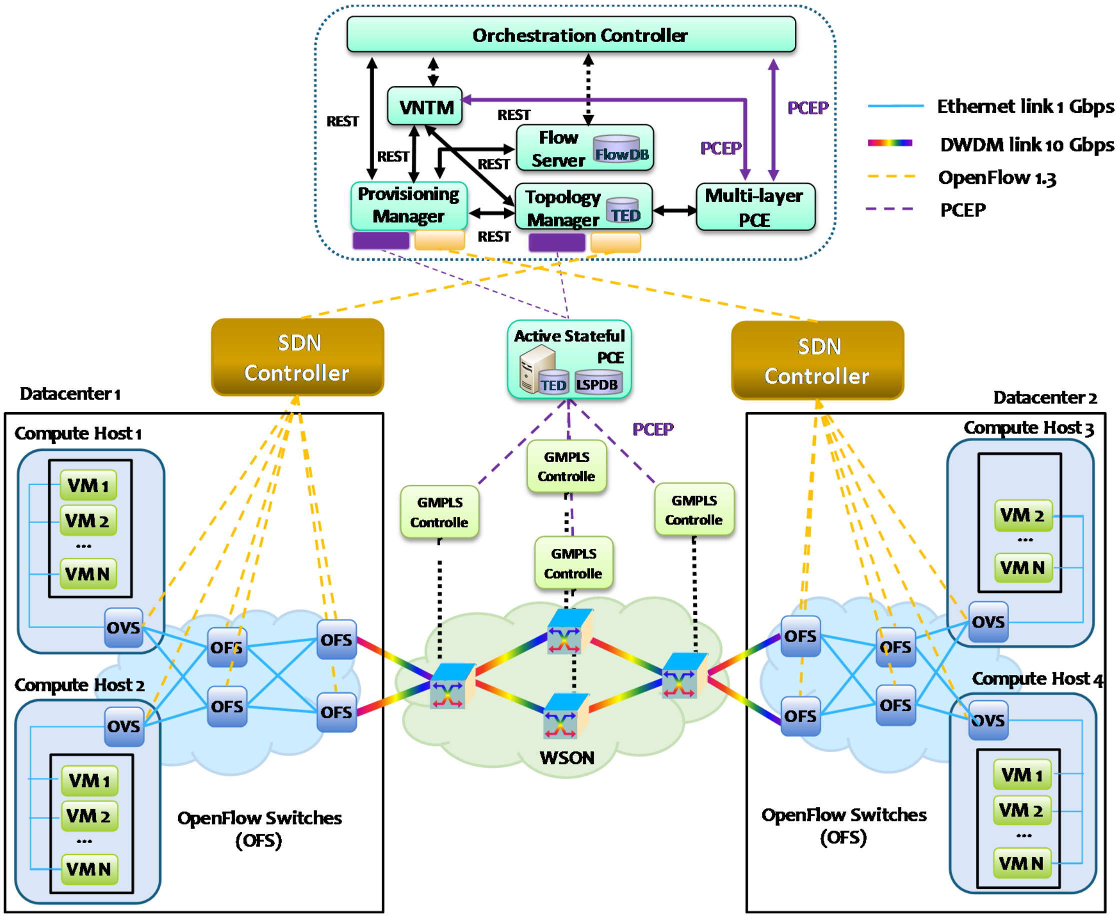


Figure 30 Orchestration architecture for the TE enforcement in packet over optical networks

Based on the IETF proposed Application-based Network Operations (ABNO), an SDN orchestration architecture (Figure 30) is introduced to provide E2E connectivity services in a multi-domain, multi-vendor and multi-layer network scenario. We focus on addressing the enforcement of TE metrics at the orchestration layer, with two principal objectives: (1) Multi-Layer E2E traffic aggregation; and (2) E2E Quality-of-Service (QoS) enabling. The three major challenges which must be solved by the orchestration architecture are: a) the multi-layer and multi-domain topology management; b) the end-to-end, TE-aware path computation; and c) the QoS guarantee into the E2E service provisioning through different provisioning interfaces, using per-flow service classification.

To fully exploit the network utilization, TE-strategies should be applied to place the traffic where the network resources are available. Our proposed SDN includes TE parameters (bandwidth, delay sensitivity, TE metric, etc.) into the network topologies managed by orchestration layer. Specifically, in the proposed architecture, the Topology Manager (TM) is the responsible of the Traffic Engineering Database (TED) management, where the multilayer, multi-domain topologies are stored. The TE enforcement strategy consist in the proactive reservation of the network capacity and its dynamic representation into the TED network topologies. The relevant topological parameters necessary to include at orchestration layer TED are: the maximum reservable bandwidth and the unreserved bandwidth. The former is obtained from the topology information retrieved by the per-domain controllers and the latter must be dynamically updated after every new connection is effectively established.

A multilayer PCE using the aforementioned TED, is included within the ABNO to perform the E2E path computations. The unreserved bandwidth on every link is taken as a constraint by the algorithm to calculate the E2E path across the multi-layer network. The CSPF output returned within the PCEP Response includes a multilayer Explicit Route Object (ERO) representing the path combining packet links (physical or virtual) and/or optical hops where the layer adaptation is ensured.

An E2E service may trigger the provisioning of one or more lower-layer connections depending on the need for transport the traffic through. The reservation of lower-layer resources (i.e. an optical circuit) must be advertised into the higher-layer topology as a new logical Virtual Link (VLink). The Virtual Network Topology Manager (VNTM) is the ABNO component responsible of the management of VLink creation, which involves the provisioning of lower-layer connections and the creation of the logical representations into the multilayer topology. When a new VLink is created, it is characterized with the TE information of the underlying connection (i.e. total available bandwidth, aggregated delay).

TE-aware traffic aggregation requires the guarantee of the effective bandwidth reservation into the E2E service provisioning. While in an optical domain frequency slot is reserved for each Label Switched Path (LSP), in a packet-based OF domains it is necessary to limit the bandwidth assigned to each flow (flow service classification) in order to preserve a certain QoS (i.e. bandwidth) assured to each E2E services. OF v1.3 introduces a new message (i.e. OFPT_METER_MOD) which enables the specification of traffic meters into the OF-switches, with an associated Band-rate and a QoS-enabling strategy, such dropping packets at a determined Drop-rate or Differentiated Services Code Point (DSCP) packet tagging to allow DiffServ. Flows can be attached to these predefined meters, associating a maximum rate to each flow. A Meter instruction has to be included inside the OFPT_FLOW_MOD message indicating the Meter-Id desired to be attached to.

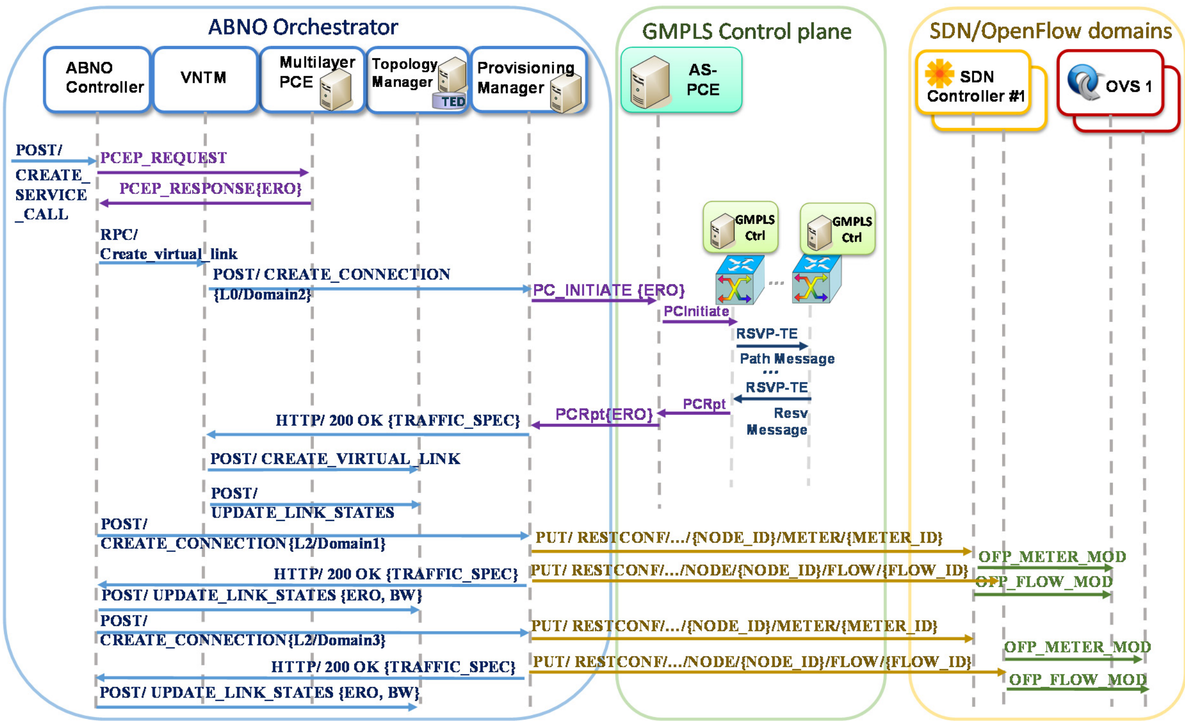


Figure 31 Multilayer E2E service orchestration workflow

Figure 31 shows a complete E2E service orchestration workflow including the effective bandwidth reservation into the packet domains employing OFv1.3. After every connection is established, the Traffic Specification containing the Peak Rate of the connection is returned to the orchestrator and the affected links are updated with the certain amount of network capacity that has been reserved.

The proposed architecture has been experimentally validated in the ADRENALINE Testbed. The network scenario consists of two packet SDN/OpenFlow domains based on Ethernet transport technology and an aggregation optical transport network. Each packet domain consists of four OF Switches deployed using standard Custom Off The Shelf (COTS) hardware and running OpenVSwitch (OVS) and Lagopus, which are all controlled by a per-domain SDN controller (based on OpenDaylight) using OF 1.0 and 1.3. Each OF border switch has been deployed into COTS hardware, a 10 Gb/s XFP tuneable transponder and OVS technology. The core network consists into a GMPLS-controlled optical network controlled by an Active-Stateful PCE (AS-PCE).

Figure 32 shows the traffic capture of the control overhead of the E2E orchestration of a single SERVICE_CALL request, demanding a 100Mbps Ethernet (L2) service. Firstly, the capture validates the multilayer orchestration by showing a virtual link creation supported by a bidirectional optical LSP requested to an AS-PCE (PCEP Initiate and Report messages). Secondly, the UPDATE_LINK_STATE message is shown, which is the responsible of update the link states after a connection (at any layer) is created. Finally, the per-flow traffic limitation in the SDN/OF domains is illustrated. The RESTCONF messages, sent from ABNO to the per-domain controllers, requesting the creation of a METER_BAND with 100 Mbps band-rate and the subsequent flow creation request with the associated METER_ID are shown in detail in the lower part of the capture. The meters are only introduced at the first and last nodes of the E2E connection, limiting the incoming traffic from both endpoints.

Figure 32 Multilayer E2E TE-aware service orchestration Wireshark capture

Finally, we demonstrate the per-flow meter traffic limitation concept in the data plane by showing a real traffic capture of the throughput reach by three different flows after traversing the OFSwitch, with and without attaching a meter band to them (Figure 33). The Iperf Bandwidth Measurement Tool has been employed to obtain the results placing a client into the machine connected to the input port and a server measuring the bandwidth achieved into a second machine, connected at the output port of the switch. The three meters are set to 100, 300 and 600 Mbps, we can observe how after applying the method previously described differentiated QoS levels to each flow have been successfully achieved.

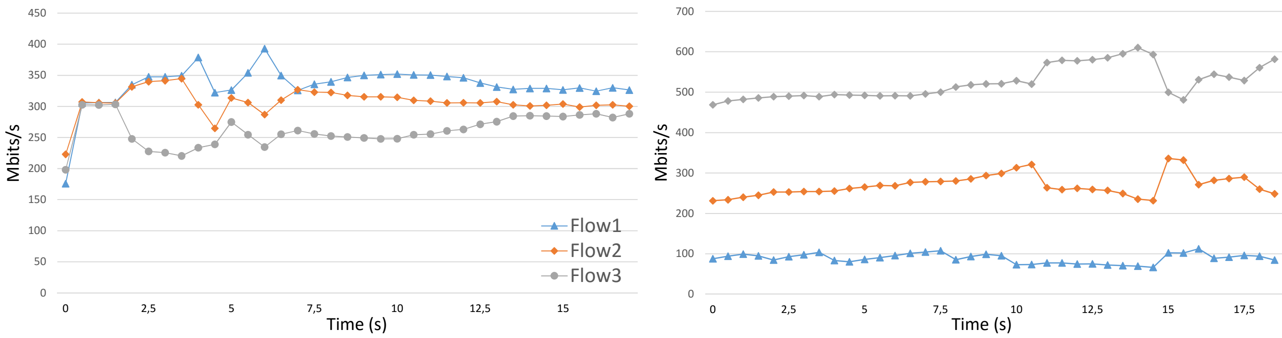


Figure 33 Per-flow bandwidth limitation; a) IO graph OF Switch output_port without meter limitation; b) IO graph OF Switch output_port throughput with 600, 300 and100 Mbps meter limitation

3.3.4 ML Segment Routing

The Segment Routing (SR) technology has been recently proposed, aiming at providing effective traffic engineering solutions while simplifying control plane operations [SR, SR-MPLS].

SR can be applied to Multiprotocol Label Switching (MPLS) networks by exploiting the label stacking functionality. In particular, SR relies on the source routing paradigm to enforce a packet flow through a path by applying, at the ingress node, a specifically designed stack of labels. This stack is compatible with the MPLS data plane and consists in an ordered list of segment identifiers (SIDs). During packet forwarding, only the top label in the stack is processed. That is, the packet is forwarded along the shortest path toward the network element represented by the SID. For example, a SID can be related to an IGP-Prefix, e.g. the loopback address of a node.

Unlike traditional MPLS networks, SR maintains per-flow state only at the ingress node, where the label stack is initialized. No signaling protocol is then required to populate the forwarding table of transit nodes, thus reducing the control plane load and simplifying the complex and time-consuming provisioning procedures. This is particularly relevant in multi-region/layer networks, where SR can eliminate the need to establish and maintain hierarchical instances of Label Switched Paths (LSPs).

So far, no experimental validations and scalability assessments of the SR technology have been reported in the scientific literature.

In IDEALIST, the SR technology has been considered and assessed, aiming at efficiently controlling the routing process in a multi-layer network.

In particular, a novel software defined networking (SDN) solution is adopted to control the label stacking configuration and successfully implement dynamic traffic adaptations without requiring GMPLS operations. Thus, optical bypass upon traffic load variations is applied and successfully demonstrated, also assessing the scalability performance of the SR-based solutions.

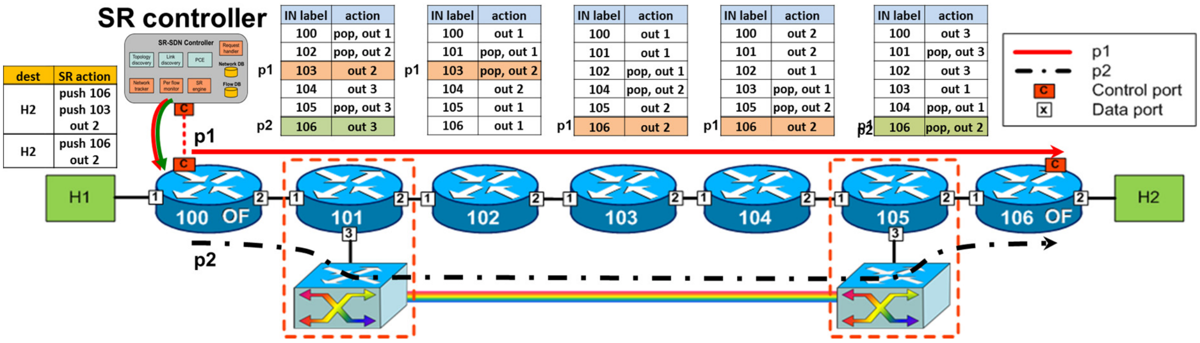


Figure 34 Testbed for optical bypass driven by Segment Routing operations

Figure 34 shows the considered multi-layer packet over optical network. The network includes nodes equipped with packet switching capable (PSC) interfaces supporting MPLS data plane forwarding. In particular, the PSC interfaces support pop and push operations on labels. Such operations are utilized during SR-based forwarding. The routing tables within the PSC nodes enable the selection of the next hop along the shortest path. This way, each node is able to autonomously determine the forwarding table used by the SR technology, i.e. based on SID (reported in the figure above the nodes).

The edges of the PSC domain (e.g. node *100*) also support SDN-based SR configurations. A specifically designed SDN-based SR controller is introduced to operate just on the edges of the PSC domain.

The SR controller encompasses network topology and traffic engineering information. This way, it is able to perform path computation by accounting for the available network resources.

The network is multi-layer as it includes reconfigurable optical add/drop multiplexer (ROADM) technology which is transparent to SR forwarding. A lightpath between node *101* and node *105* is activated to enable the optical bypass of the sequence of PSC nodes *102-104*.

PSC nodes have been implemented within Intel Core 4 servers (CPU 2.40 GHz, Kernel 2.6) equipped with 4 Gb/s Ethernet interfaces and running OpenvSwitch version 2.61, supporting MPLS-based forwarding. The OpenvSwitch software implementation has been modified to increase the maximum stack depth from 3 to 15 labels.

OpenFlow 1.3 has been utilized for the communication between the SR-Controller and ingress PSC node. Commercially available ROADMs equipped with 10 Gb/s OTN muxponders have been connected to PSC nodes as in the figure. The SR-Controller has been implemented by enhancing the SDN RYU controller. In particular, a specifically designed algorithm has been introduced to select the sequence of SIDs to be included as SR label stack. Moreover, the APPLY_ACTIONS type of command has been used in the OFP_FLOW_MOD message to push all the required labels as a single entry.

To perform optical bypass, the controller includes a routing policy to discriminate between packet and optical resources. The policy is here based on a bandwidth threshold (e.g. 40% of the PSC interface capacity): if the total bandwidth is below the acceptable threshold, the controller provides a routing preference for packet resources, otherwise for optical ones.

A first traffic request from node *100* to node *106* with bandwidth below the policy threshold arrives. Two possible paths can be considered by the SR controller: the path *p1* through PSC nodes only and the path *p2* through the optical link *101-105*. Given the considered routing policy, although path *p2* represents the shortest path, the path *p1* is selected for the considered request. To apply the computed route, the SR controller then configures the forwarding table of the ingress node *100* by exploiting a specifically designed OFP_FLOW_MOD message. The message, shown in Figure 35, includes the label stack information, i.e. *106* as bottom label and *103* as top label.

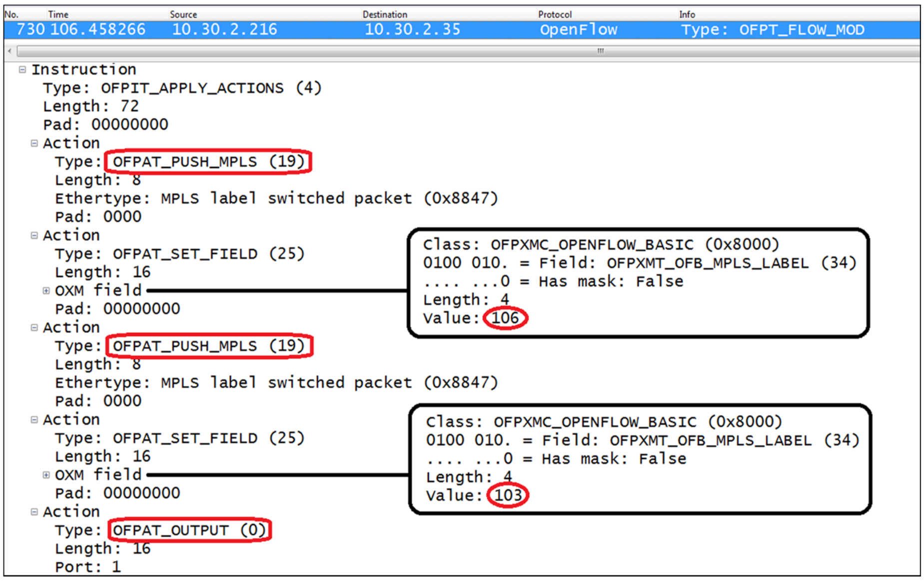


Figure 35 Captured OFP_FLOW_MOD messages sent from SR controller

The OpenFlow message is processed by the ingress node *100* which configures the forwarding table to apply the provided label stack configuration. Upon configuration, the label stack will be applied to the incoming packets, which will be sent to the adjacent node *101*. According to the included top label, the packet will be forwarded to node *102* where penultimate hop popping is performed. Then, following the new top label indicating Segment ID *106*, the packet will be forwarded toward the destination (with penultimate hop popping at node *105*).

Now let's assume that the aggregated bandwidth increases above the considered threshold due to a new traffic request from *100* to *106*, the SR controller performs a new constraint-based path computation selecting path *p2* including the optical link. To apply the computed route, a new OFP_FLOW_MOD message is generated by the SR controller and sent to node *100*. The message includes the recalculated label stack information, which encompasses just one label with value *106*.

Node *100* updates the forwarding table, applies the new label and sends the packet to *101*. Node *101*, in this case, forwards the packets through the ROADMs. The dynamic optical bypass of PSC nodes *102-103-104* is then successfully performed, with no packet loss and without requiring any signaling process through the network. This has the potential to significantly simplify interoperability operations.

Additional details, e.g. in terms of scalability performance, can be found in [Sga-OFC15]. For example, scalability tests have demonstrated that if a deep label stacking is applied, the time required to perform the overall flow configuration increases to an average value of about 1 ms, mainly due to path and label stacking computation at the SR controller. However, no performance degradation has been experienced after flow configuration, i.e. the packet forwarding time is almost independent on the label stack depth.

A further SR Controller implementation, in the context of the PCE architecture, is reported in [Sga-NOC15].

4 Future Extensions

4.1 Multi-Layer asymmetric load balancing

This section deals with possible use cases and extensions. One of the main points in aiming at a better utilization of e.g. flexible interfaces and exploiting all the advantages of the optical layer is to satisfy the demands of the client layer. Traditionally, this client layer is build up on IP and/or MPLS mechanisms, which shall be subsumed with the term "layer 3 mechanisms" in the following. This client layer is ignorant of any underlying structure and the details of how bits and bytes are transmitted over the transport layer. The terms of thinking appropriate to the layer 3 are "connectivity" and "reachability", found in well-known algorithms as in Dijkstra's shortest path determination or other routing algorithms and protocols. In this context, the physical properties of a link are hidden and the layer 3 is agnostic of any fine structure. Sloppily, in terms of IP or MPLS, a link is either "up" or "down" and nothing in between.

Furthermore, if given the chance to decide, IP/MPLS prefers a link rather to be shut down as to change its attributes. This is due to a specific layer 3 mechanism which is load balancing. Most notably this is seen by changes in the offered bandwidth (other properties of the transmission are still well hidden to the client layer).

In the following, we will shed some light on the details of this challenge and offer some solutions on how to deal with it. The overall goal of the discussion is a higher utilization of the IP interfaces (as opposed to the optical interfaces).

4.1.1 Layer 3 load balancing

One of the most prominent load balancing mechanisms is based on the so-called Equal Cost Multi Path scheme (ECMP).There, the cost of a path is measured as the sum over all link metrics along the path, where a link itself is a single (IP) hop between adjacent routers and a path consists of a series of such links. The link metric is usually taken simply as a number without units and is used in twofold way.

First, in a meshed network of routers with two or more alternative paths between source and sink router, one path can be preferred over others by assigning less costs to the links. This process is called traffic engineering and takes place on a time scale of days or weeks. The trigger for this might for example be the necessity of a router off-loading due to maintenance reasons or traffic steering in order to have an equally distributed network.

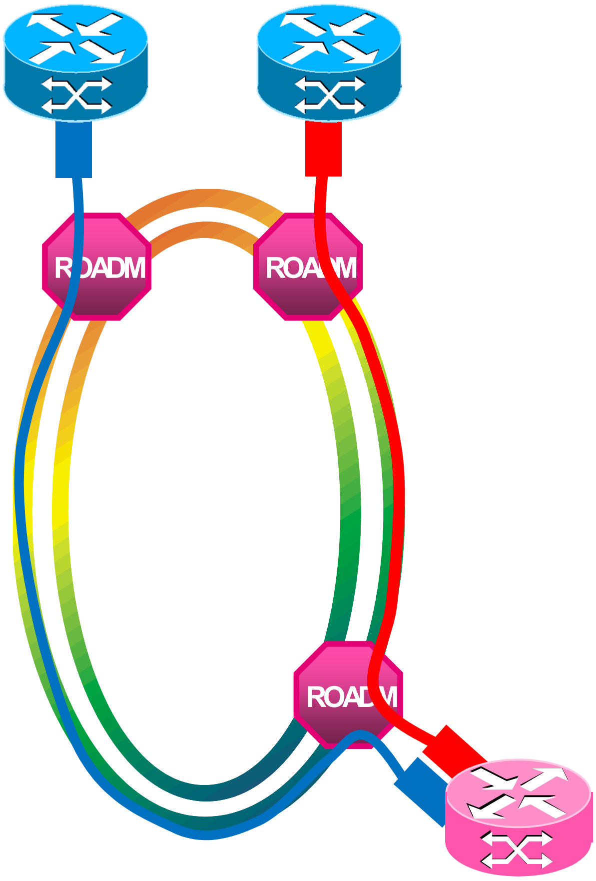


Figure 36 Example of metro/aggregation ring. The (pink) edge router has two IP links to two core routers (blue). The logical links are lead over two disjoint paths for resilience reasons and load balancing between the left and the right blue router.

The second application for the link metrics is as follows: in networks with a node degree equal to or higher than 2 (that is, each node is at least connected to 2 other nodes and at least 2 of the nodes have an extra connection), there exist at least two disjoint paths between any two nodes. This is best seen in a ring topology, where data can be sent over the left or the right direction. A prime example for such a topology in communication networks are metro rings or aggregation rings as shown in Figure 36 where an edge router (pink) is connected over optical infrastructure to two different core router (blue). In networks with an even higher node degree, many more equivalent paths are possible, e.g. in networks built up of triangles and a distance of N hops between source and sink router, there exist $2^{(N-1)}$ different paths (not necessarily disjoint to each other). If the link metrics are chosen such, that all of the possible paths yield the same cost function – which can be reached by assigning unit metrics for all links – the costs can be used to determine equivalent paths for each local router to all its counterpart routers. Given such a situation, the ECMP algorithm calculates a hash value for each packet to be sent to a certain destination. This hash value is taken over the header fields of the transport and network protocol (TCP/UDP, IP/MPLS) to account for requirements of their layer with regards to the OSI layer model (e.g. to prevent packet reordering and to match the TCP retransmission mechanism etc.). Also, by taking source and destination addresses and transmit and receive ports into account, an entropy is reached that is large enough to perform the hash ordering to so-called buckets by applying a modulo operation to the hash value and to avoid an imbalanced load of the links. In the ring topology of Figure 36 with only two paths, this comprises only a differentiation between even and odd (mod1).

In more complex situations, the modulo operation is taken with regards to the number of different paths over which a target can be reached. Modern hardware typically can discern between 2^6 or 2^7 different hash values and the calculation and interpretation of these for all practical purposes takes no extra time, because it is done in parallel to the determination of the out-going interface while finding a match in the routing or forwarding table for each packet.

In real world communication networks the situation will be neither the first nor the second extreme as far as the assignment of link metrics is concerned. Instead, the metric design is used as mixture between traffic engineering and provisioning of alternative paths. The latter is a key feature of protected networks. As long as there are more than one working paths, in case of a failure immediate action can be taken by rerouting the packet traffic over just these replacement paths. It is not easy to guarantee that the replacement path(s) can transport the extra load arising due to the failure, since the missing link might be bearer for many different traffic relationships. To accomplish this task complex calculations and a large amount of engineering experience are necessary.

4.1.1 Flexrate interfaces and IP/MPLS - the Challenge

As described above, load balancing and traffic engineering is a fine balanced dance at the edge of chaos. An important heuristic for the ECMP scheme to work as designed is the necessity to assign equal costs only to those paths that are not only equivalent in a topological sense, but are also comparable to each other w.r.t. the offered bandwidth. It is easy to understand, that load balancing between a 1Gbps link and a 100Gbps is not very reasonable. This stems from the fact, that IP is agnostic to link properties as data rate or OSNR or the like. As long as connectivity is given, a path is available and as such will be part of the routing algorithm independent of any further constraints. This in turn is due to the fact that the hashing and bucket functionality is buried deep within the hardware and at least at the moment cannot easily be adjusted.

Now, assume a situation as illustrated in Figure 37, where we consider the traffic between the two pink routers. In the working case, there are two possible paths, each of which consists of two optical hops (spans). In the case of a failure, as for example in Figure 38 one of those paths need to be restored. After the restoration, the new path for the wavelength might be considerably longer and thus cannot support the original modulation format due to a worse OSNR or the like. From a pure transport point of view, this does not result in a problem, thanks to the flexrate technology, which adjusts the data rate to the next smaller possible match. However, from a packet perspective this does indeed come as a challenge. The formerly equal paths are no longer equivalent although they still might have the same metric value assigned (for the IP/MPLS layer, the restored light-path still resembles only a single hop between adjacent routers. After all, that is, what restoration is all about – to leave the IP/MPLS topology unaffected and hence, to save interface costs!).

As described above, the usual symmetric load sharing between the undisturbed path and the restored path cannot be applied. The packet layer somehow needs to be informed about the change in the data rate. Else the ECMP mechanism just keeps distributing the arriving packets as before which might result in an overload for the constricted new path (of course, this only applies to a situation, in which the links are appropriately filled anyway. In an empty network, the failure has almost no impact as far as the traffic load is concerned).

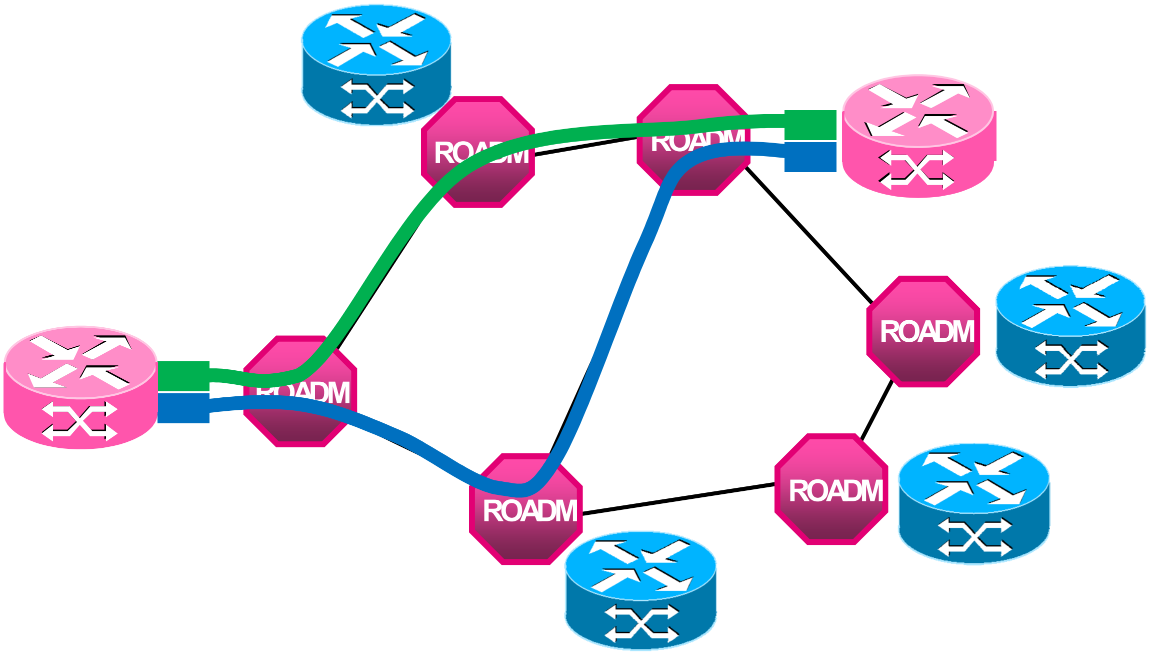


Figure 37 A typical network situation. Source and target routers (pink) have two possible working paths (blue and green line)

Unfortunately, to simply adjust the link metric for one of the affected links is not possible. The first problem is, whether the metrics of the green unperturbed path should be decreased or whether the metric of at least one of the links of the new blue path should be increased. Even if one could agree on a convention ("always increase affected links"), this would necessarily result in global changes: then a formerly absolutely not at all involved region of the network might be affected. For example, the traffic between one of the blue routers at the right and the single blue router in between the pink router will very likely be rerouted as well, This arises due to the fact, that link metrics are globally valid and aim at the whole set of all traffic relations. Thus, a change of the metrics has unforeseeable consequences. Also, if one would really follow this approach (which is, to adjust global parameter due to local impairments), one would introduce yet another time scale for a problem, which already has to cope with more than one time scales. To explain this, we need to consider a possible over-all resilience scheme which might be built up like this:

- The first and immediate reaction of the network to a failure detection takes place on the time scale of milliseconds (e.g. fast optical protection or fast reroute (FRR) on a packet layer)
- Broadcast distribution of the failure state within the network than happens within seconds by means of routing protocols (IGP, BGP)
- For the restoration of an optical light-path through the fiber network one usually assume a couple of minutes (the precise duration depends on physical constraints as number of affected wavelength, the gain parameter, the span length etc. So a absolute timescale cannot be given as concrete as could be done for the mechanisms on the electrical layer)
- Finally, as mentioned above, traffic engineering is a matter of hours or days and is not light-heartedly adopted into a network, because the global effects are not easily understood.

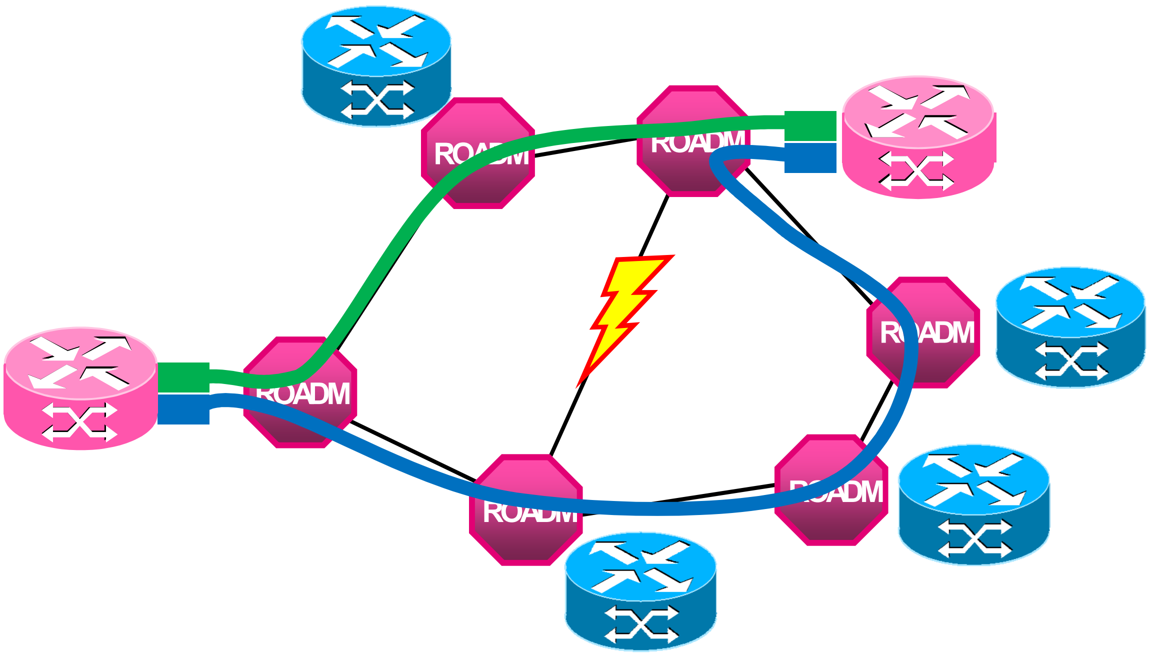


Figure 38 Network with the restored blue wavelength. The new light-path is considerably longer and might no longer bear the original modulation scheme.

So, in short, the usual way to cope with a situation as described above would be to completely shut down the affected link. This is necessary not to conflict with traditional mechanisms on the packet layer. However, this is not the optimal solution, since it just throws away the potential of the optical layer.

4.1.2 A possible way out

Now, after having learned, that flexrate has a specific impact on the packet layer, the requirements can be formulated and a possible way out comes into reach:

- The local IP interfaces need to be adjustable to different data rates offered by the optical flexrate devices
- In case of load balancing, the traditional usage of sharing the load between the different paths (i.e. local IP interfaces) needs to be extended
- This extension must not conflict with neither the mechanisms on higher layer (packet reordering is to be avoided), nor with the over-all global traffic engineering (non local effects are to be avoided)

The solution to the challenge is **asymmetric load balancing** between the involved links. Whenever the optical capacity changes on the line-side (due to whatever reason), this change must be accompanied by appropriate adaption at the client-site. This adaption comprises two steps:

- First, the nominal capacity from the IP interface need to be adjusted (probably in fixed steps of, say, 25 GbE granularity or whatever seems to be a suitable unit).
- Second, the load balancing algorithm need to be adjusted.

For the latter, one possible solution on the hardware level could be sketched like this: while finding a match in the routing/forwarding table for each packet, just as before calculate in parallel the hash value of the current package. Then, instead of taking the modulo operation with respect to the number of equivalent out-going interfaces, determine the bucket in which the hash value is falling. The size of these buckets are chosen according to the relative weight of all the links when compared in terms of data rate. for example, if the routing protocol finds three equivalent paths (measured in link metrics), with respective data rates of 100 Gbps, 40 Gbps, and 10 Gbps, their bucket weight should read 10,4, and 1 (or, when assuming a total hash-range of 2^7 bits, i.e. 64 possible values, 43, 17 and 4).

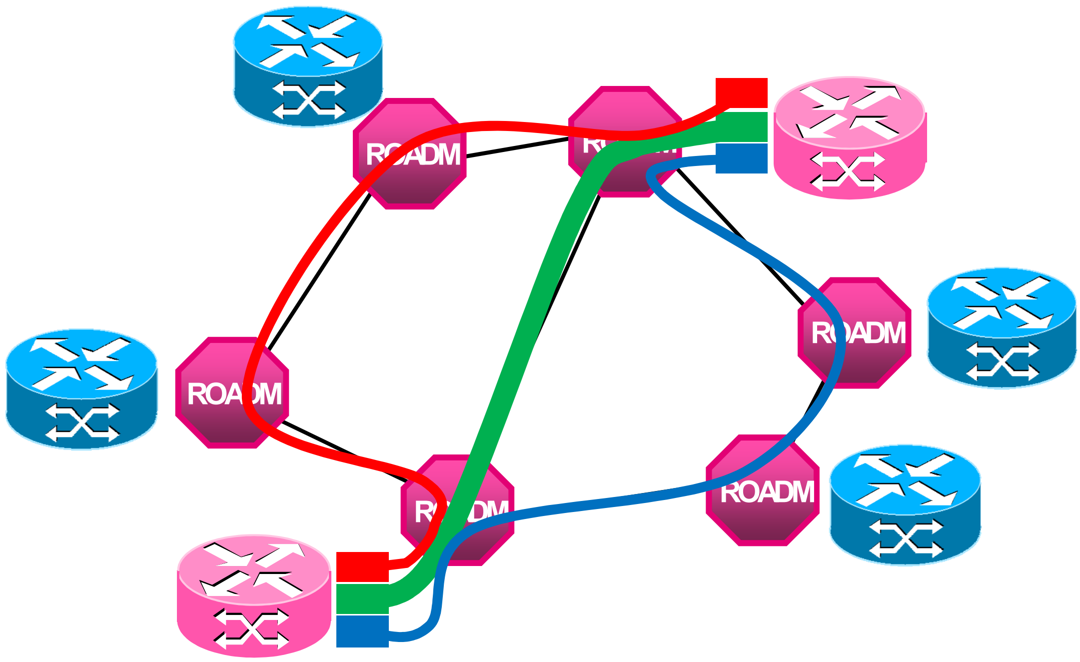


Figure 39 Three possible light-paths between the two pink routers. Only the green one allows a higher modulation format, while the red and the blue path should bear a smaller data rate. This could be achieved with local link weight in the pink router, according to the offered optical data rate.

These link weights are meant to be locally valid only. They are not to be intermingled with the link metrics for the global traffic engineering. In order to avoid conflicts, a possible migration path towards a full blown solution might be in a first step only to implement the possibility of asymmetric load balancing without using it in non-failure cases. Thus, the well-known processes and working schemes can be taken just as before. To be precise, this implies a distribution of load according to the ECMP mechanism. Only in case of a failure, the routers directly connected to the failing link are then allowed to switch to the asymmetric load balancing scheme. By doing this, the spread of the failure is kept rather small and manageable and the risk of propagation a wave of unforeseeable changes through the network is diminished. This phase could be used to make more experience with the system and to learn how it behaves in real world environments. In a second step, the asymmetric load balancing could also be adopted in situations, where no failure occurred as illustrated in Figure 39. But to finally judge this possibility, more research and discussion within the community is necessary.

5 Summary of control plane extensions

Control plane extensions affected all the protocols of the GMPLS suite together with the ones adopted as northbound interfaces between the cPCEs and the pPCE (i.e. PCEP and BGP-LS). The main purpose of these extensions is the handling of the media layer in which switching is based on a frequency slot described as central frequency and a slot width.

Most of them have already been documented in previous WP3 deliverables and will be shortly reported here only for completeness; others, covering the case in which transceivers in different domains must interwork, will be dedicated a detailed description.

5.1 RSVP-TE extensions

Label: new 64-bits label format defined by the following encoding:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |     Identifier    |              n              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             m                 |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- `Grid` specifies the grid type and has been assigned a value of 3 to indicate ITU-T Flex.
- `C.S.` is the Channel Spacing and the new value of 5 indicates a 6.25 GHz granularity.
- `n` expresses the slot central frequency according to the formula:
  Frequency (THz) = 193.1 + n * 0.00625
- `m` is the slot width according to the formula:
  Slot Width (GHz) = 12.5 * m

All other fields have the same meaning and values defined in [RFC6205], from which the new format is derived. The new format is used in all the objects carrying a label (GENERALIZED_LABEL, SUGGESTED_LABEL, LABEL_SET, ERO, etc.).

Traffic descriptors: new C-Type for the SENDER_TSPEC and FLOWSPEC objects to specify Spectrum Switched Optical Network (SSON) traffic parameters with the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               m               |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where `m` indicates the slot width according to the same formula used for the label field of the same name. Since IANA has not yet allocated a value for it, a C-Type = 10 was chosen.

It can be observed that 'm' is a parameter both of the GMPLS Flexigrid label and of the Flexigrid TSpec and Flowspec. The overlap comes from the fact that in a Flexigrid system the label value, that defines what is switched, indicates the slot width, therefore affecting also the bandwidth supported by an LSP.

Label Request: a new value (190, not yet allocated by IANA) was necessary for the `switching type` field of the GENERALIZED_LABEL_REQUEST object to specify spectrum switching (i.e. Flexi-grid) capabilities.

ERO: A new sub-object of the EXPLICIT_ROUTE object was defined to describe MF-OTPs or SBVTs capability of generating multiple optical flows. It is formed by a list of TLVs describing the sub-carrier attributes and appears only at the beginning and the end of the ERO to convey specific information about the configuration of the MF-OTPs at the path endpoints.
Such specific information includes:

- Sub-carriers IDs (2 or more if a superchannel LSP is being signalled).
- The modulation format per sub-carrier.
- The FEC information per sub-carrier.
- The Frequency Slot (FS) occupied per sub-carrier.
- The transponder class per sub-carrier.

To that end, the RSVP-TE ERO is extended with the so-called Explicit Transponder Control (ETC) identified by the ERO sub-object type 10 (not assigned by IANA) and formed by a variable list of Sub-Transponder TLVs. Each Sub-Transponder TLV is used for a specific sub-carrier allowing the configuration of a set of sub-carriers forming a superchannel LSP.

The Sub-Transponder TLV is set to type 1 and contains the following sub-TLVs (types: 5005, 5006, 5001, 5002 and 5020):

- **Sub-Transponder ID** (5005): the value field (32 bits) contains the local ID used for the sub-carrier being occupied by the LSP.

  Note: if it is not desired to maintain a global view of the sub-carrier IDs per MF-OTP within the network (i.e. TED), a solution is to set sub-Transponder ID to 0. This must be interpreted by the node processing the ERO as there is no preference on allocating a sub-carrier/sub-transponder, just select and allocate for the LSP being signaled one being available.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 5005            |          Length = 4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        SubTransponder ID                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **Sub-Transponder FS** (5006): the value field (64 bits) describes the Frequency Slot (FS) used by the sub-carrier in terms of grid, channel spacing, center frequency (n) and slot width (m).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 5006            |          Length = 8           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             m                 |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **Sub-Transponder Mod Format** (5001): In the value field (96 bits) it is specified
  - S: standardized format;
  - I: input / output (1 / 0)
  - Modulation ID: BPSK (1), DC DP BSPSK, QPSK, DP QPSK, QAM16, DP QAM16, DC DP QAM16
  - Symbol Rate: bauds/s formatted as an IEEE 32 bits float
  - Number of carriers
  - Bits/symbol

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 5001            |          Length = 12          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|I|      Modulation ID        |          (reserved)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Symbol rate                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Num carriers         |          Bit/Symbol           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **Sub-Transponder FEC** (5002): In the value field (32 bits) it is specified:
  - S: standardized format;
  - I: input / output (1 / 0)
  - FEC ID: reed-solomon (1), haming-code, golay, BCH

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 5002            |          Length = 4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|I|         FEC ID            |          (reserved)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The above listed TLVs are not able to describe all possible transceiver's parameters; for this purpose, it would be necessary to define many other TLVs, complicating the ERO structure and impacting negatively on the packet length. To overcome this limitation, other than specifying all the transceiver parameters, it might be convenient to summarize a set of parameters in a single proprietary parameter: the class of transceiver.

The transceiver classification will consider the main TX and RX parameters: Trunk Mode, Framing, Trunk Type, Channel Band, Channel Grid, Electrical Signal Framing at Tx, Minimum maximum Chromatic Dispersion (CD) at Rx, Maximum Polarization Mode Dispersion (PMD) at Rx, Maximum differential group delay at Rx, Pre-FEC BER, Q-factor, Q-margin, Constellation mapping, etc. To use this parameter, a transceiver vendor will be responsible to specify the class contents and values. The vendor can publish the parameters of its classes or declare to be compatible with already published classes

- **Sub-Transponder Transceiver Class** (5020): the value field (32 bits) specifies the selected class.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 5020            |           Length = 4          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Transceiver Class                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Another proposal for the Sub-transponder FS TLV has been analyzed, starting from the assumption that, in some cases, it may be required that the sub-carriers are tuned at a frequency having granularity steps of 0.1 GHz and a distance (or Grid) not multiple of 6.25 GHz (e.g. 3.125). In this case the above described type 5006 TLV format is not appropriate as the central frequency can be set at 6.25 GHz steps only. Moreover, it doesn't seem necessary to be able to define a different channel spacing for each sub carrier. Under these assumptions, the sub-carrier definition can be modified in order to allow setting the granularity in 0.1 GHz steps and to use a "relative addressing" instead of an "absolute addressing" by means of an off-set (k) from the media channel central frequency (n). In addition, the FS width (j) must be specified, always at the granularity of multiple of 0.1 GHz.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type=5006         |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Grid     |   Identifier  |               j               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               k                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The proposed solution has the advantage to really allow using all the fiber bandwidth in case of linear point-to-point network in which the user can define a network media channel the whole C-band wide and fill the subcarriers as a single data channel (not related to each other) thereby reducing the fragmentation.
One problem that can arise from the ETC approach involves the ERO sub-object length that may exceed 256 bytes which is the maximum allowed (length encoded in 8 bits).

In fact, for each subcarrier, the ETC must be structured as follows

Sub-Transponder TLV	Type 1	4 bytes
Sub-Transponder ID subTLV	Type 5005	8 bytes
Sub-Transponder FS TLV subTLV	Type 5006	12 bytes
Sub-Transponder Mod Format subTLV	Type 5001	16 bytes
Sub-Transponder FEC subTLV	Type 5002	8 byes

Sub-Transponder Class subTLV	Type 5020	8 byes

leading to a total of 56 bytes.

For example, if a 7 subcarriers superchannel is to be signaled, the total ETC length would be 394 bytes.
Two solutions can be envisaged:

- Concatenate more ETC subobjects: e.g. ETC1 for the first 4 subcarriers and ETC2 for the last 3 subcarriers.
- If the transponders transmission parameters for a superchannel are always the same, specify all sub-TLVs only for the first subcarrier and describe the remaining subcarriers in terms of ID + FS only.

While the second solution avoids duplication of information, the first one seems to be more general.

5.2 OSPF-TE extensions

The extensions to the OSPF-TE protocol have been defined for the dissemination of NCF availability by means of a new Switching Capability-specific Information (SCSI) field to be appended to the Interface Switching Capability Descriptor (ISCD). The SCSI is used to carry the technology specific part of the flexi-grid DWDM and may include a variable number of sub-TLVs called Bandwidth sub-TLVs. To advertise the available labels sub-TLVs type 1 is used. The format of such TLV is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      PRI      |                 Reserved                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Label Set Field                        |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where

- `PRI`: A bitmap used to indicate which priorities are being advertised

The `Label Set Field` is encoded using the bitmap format (`action = 4`) in which each bit in the bitmap represents the availability of a particular central frequency. Bit position zero represents the lowest central frequency and corresponds to the base label, while each succeeding bit position represents the next central frequency logically above the previous.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   4   |     Num Labels        |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          start Label                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          start Label(continue)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #1 (Lowest numerical labels)                  |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #N (Highest numerical labels)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

In this case, the `start Label` field specifies the lowest supported central frequency using the same label format already described for the RSVP-TE protocol (in this case, *m* has no meaning and is set to zero).

5.3 PCEP extensions

For the implementation of the multi-domain control plane based on an H-PCE architecture, the different PCEs must support stateful PCEP extensions [PCEPState] encompassing also LSP instantiation capabilities [PCEPInit] and hierarchical PCEP extensions [PCEPHier]. Specific extensions have been defined for the Routing and Spectrum Assignment (RSA) procedures in a hierarchical framework. Upon pPCE request, all cPCEs compute the node sequence inside their domains and forward this information to the parent together with spectrum availability. This is accomplished attaching two new objects to the PCRep message:

- a LABEL_SET object (class 130, type 1) that encodes the free NCF as a bitmap:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   4   |     Num Labels        |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |                n              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #1 (Lowest numerical labels)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #N (Highest numerical labels)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- a SUGGESTED_LABEL object (class 129, type 1) where the label is encoded with the same format as in the RSVP-TE protocol, suggesting (but not mandating) the label (specific frequency slot) to be used on the computed path.

The same extensions already discussed for the RSVP-TE protocol apply also to the ERO PCEP object. At last, the BANDWIDTH object of the type “Requested generalized bandwidth” implements the new traffic descriptor already presented in the RSVP-TE section to allow the PCE correctly managing the path calculation request, taking into account that, in a Flexigrid network, the LSP bandwidth corresponds to the FS width.

5.4 BGP-LS extensions

The BGP Update message includes new LINK_STATE attributes TLVs:

- **NCF availability** (1200): defined for the distribution of the available frequencies. The format is same used for the ISCD of the OSPF-TE protocol.

- Transceiver Class and application code (2498):

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2498            |           Length = 4          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Transceiver Application code                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Transceiver Class                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Transceiver Application Code: 32 bits

The code, of type DScW-ytz(v), allows to define the application scenario in terms of wavelength range (e.g. 1530-1565nm), type of fibre (e.g. G.652), dispersion compensation, presence of amplifiers, system rate (e.g. 10G), etc.

An example of considered code could be FN6.25U-7A2(c)F: flexi-grid, narrow spectral excursion, 6.25 GHz channel spacing, dispersion uncompensated link, terabit transmission, optical amplifiers included, ITU-T G.652 fiber, C band, FEC required.

Possible values: FN6.25U-7A2(c)F (1)

Transceiver Class: 32 bits

This field has the same meaning of the one of the same name within the ETC (see RSVP-TE extensions section)

Possible values:

1 (CNIT-Coriant agreed parameters)

2 (CTTC – ALU agreed parameters).

- **MF-OTP Encoding** (2499): It is formed by a list of TLVs describing the sub-carrier attributes. The TLVs are the ones for the ERO described in the RSVP-TE extensions section.

6 Integrated IDEALIST Scenario

This section reports the design of the scenario to be deployed and implemented in WP4 in view of the final integration and demonstration, in which control plane and data plane elements are integrated. It shows the overall macroscopic scenario, followed by specific details of the involved domains. This deliverable only covers the design of the scenario, selection of addresses and agreements on interfaces. The results of the integrated demo are reported in [D4.4].

6.1 Overall scenario

The scenario is built by the interconnection of different components, physically distributed within labs. The testbed encompasses three Flexi-grid domains with different capabilities, one hierarchical PCE, an ABNO Controller and the PLATON planning tool.

The first flexi-grid domain (102) is composed of two nodes. Each node includes: (I) a flexible and configurable digital cross connect (OTN fabric) for client mapping and centralized control of S-BVT; (II) real-time S-BVT modules carrying multiple OTU2 tributaries thanks to programmable FPGAs and a (multi-flow) optical front end that can adapt its symbol rate such as 107 Gb/s, 53.5 Gb/s per carrier and its number of carriers according to the reach, physical impairments or capacity demand; (III) an optical matrix that provides flexibility through architecture-on-demand (AoD) in terms of synthesis of fibre switching cross-connections, optical bandwidth switching with bandwidth variable WSS.

The second domain (101) comprises integrated data and control plane. The data plane setup consists of 4 nodes flexi-grid network, a CNIT /Ericsson DSP unit at the TX and two different ones (CNIT/Ericsson and Coriant) at the RX as part of an optical coherent test-bed.

The domain (105) is composed of six nodes running a GMPLS control plane performing data plane emulation at the media layer; no data plane devices are deployed. The pPCE and ABNO Controller are java-based applications performing multi-domain computation and orchestration, enabling multiple workflows.

The figure below shows the aggregated multi-domain topology seen by the parent PCE.

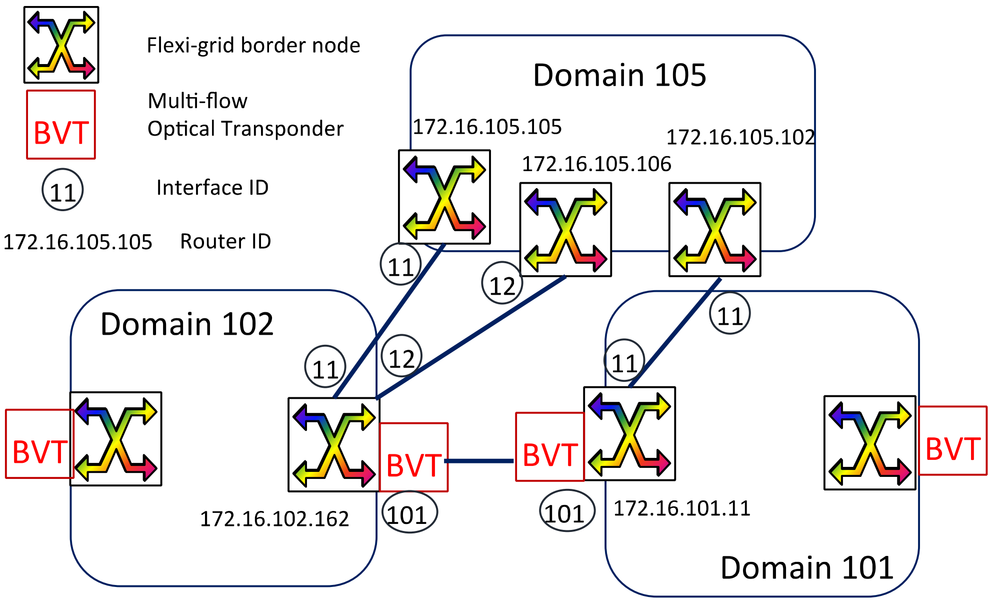


Figure 40 Inter-domain topology in the IDEALIST multi-partner scenario

6.2 CTTC/ALU/Bristol domain

This domain (Figure 41) is composed of two optical nodes, with the associated transceivers. A GMPLS control plane with an active stateful PCE is deployed, with two node identifiers being 172.16.102.161 and 162. PCEP is the provisioning protocol used by the PCE – which in turn, acts as a cPCE – and enables the establishment of optical connections. The hardware abstraction layer (HAL) via which a given node programs the underlying hardware is designed in view of the exported API by the hardware providers, which uses a REST API to configure the nodes.

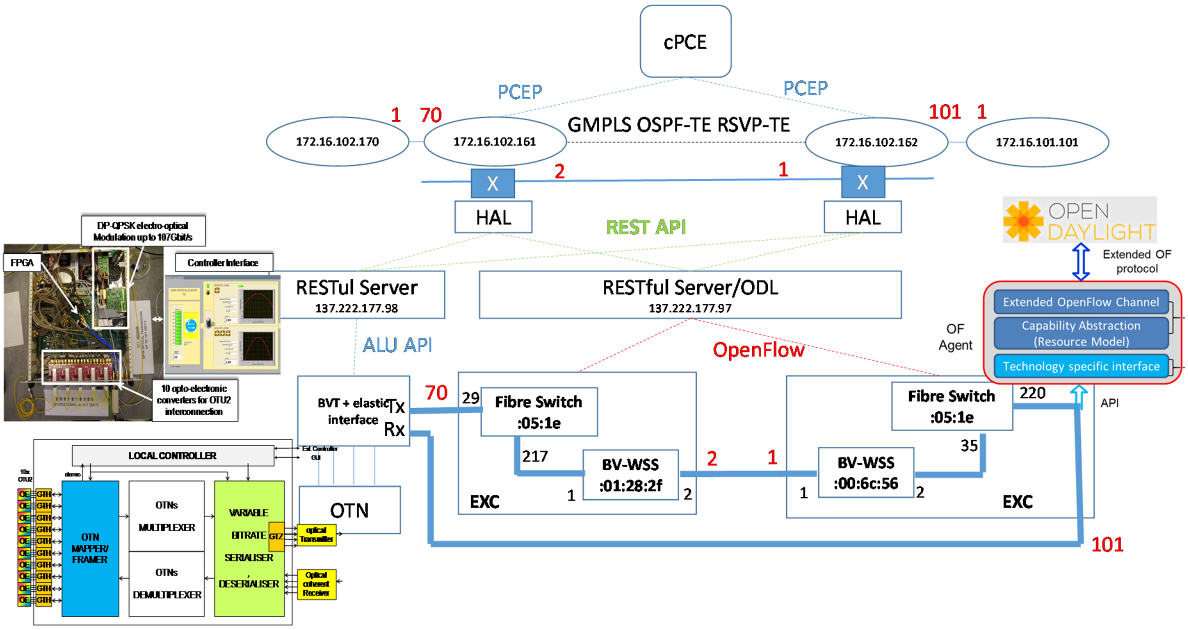


Figure 41 Final IDEALIST scenario, CTTC/ALU/University of Bristol Domain

6.3 CNIT/Ericsson domain

This domain (Figure 42) is composed of software-defined 1 Tb/s S-BVT and flexi-grid BV-OXCs. Also in this case, a GMPLS control plane with an active stateful PCE is deployed and PCEP is used as provisioning protocol.

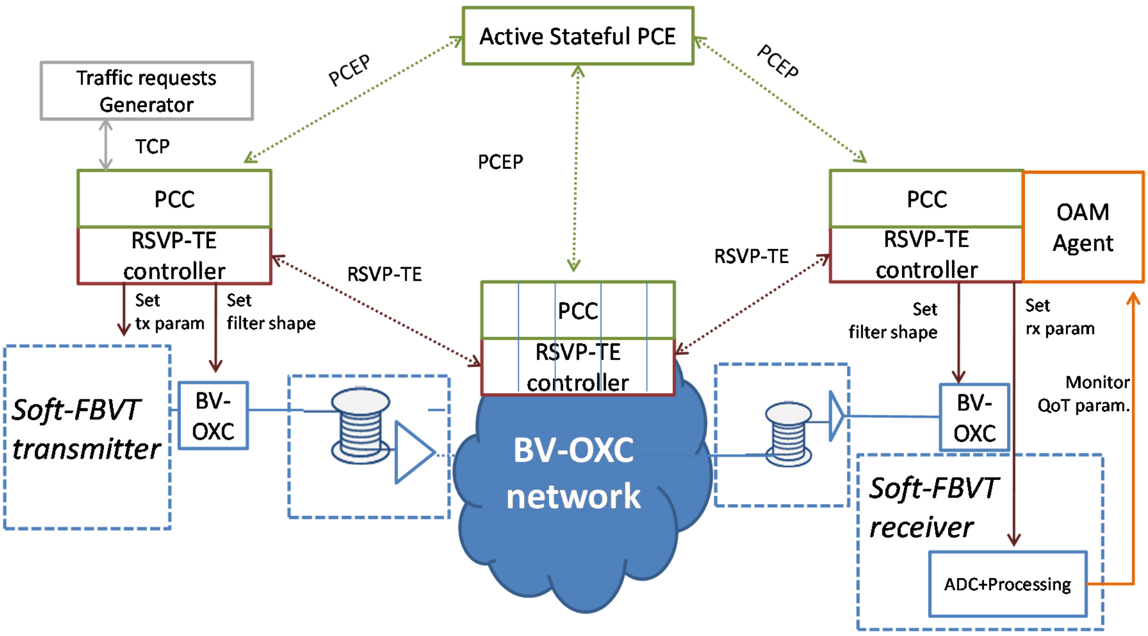


Figure 42 Final IDEALIST scenario, CNIT/TEI domain

6.4 Telecom Italia domain

This domain (Figure 43) is composed of 6 emulated nodes (i.e. without any data plane function) whose GMPLS control plane runs on dedicated Linux based boxes interconnected by means of dedicated point-to-point Fast-Ethernet interfaces emulating out-of-fibre control channels. Data plane emulation is performed only at the media layer, i.e. no transponder functionality is implemented. As in the other domains, an active stateful PCE is deployed acting also as the domain cPCE. PCEP is adopted as provisioning protocol and a BGP-LS speaker is implemented to export topology and resource availability information to the pPCE.

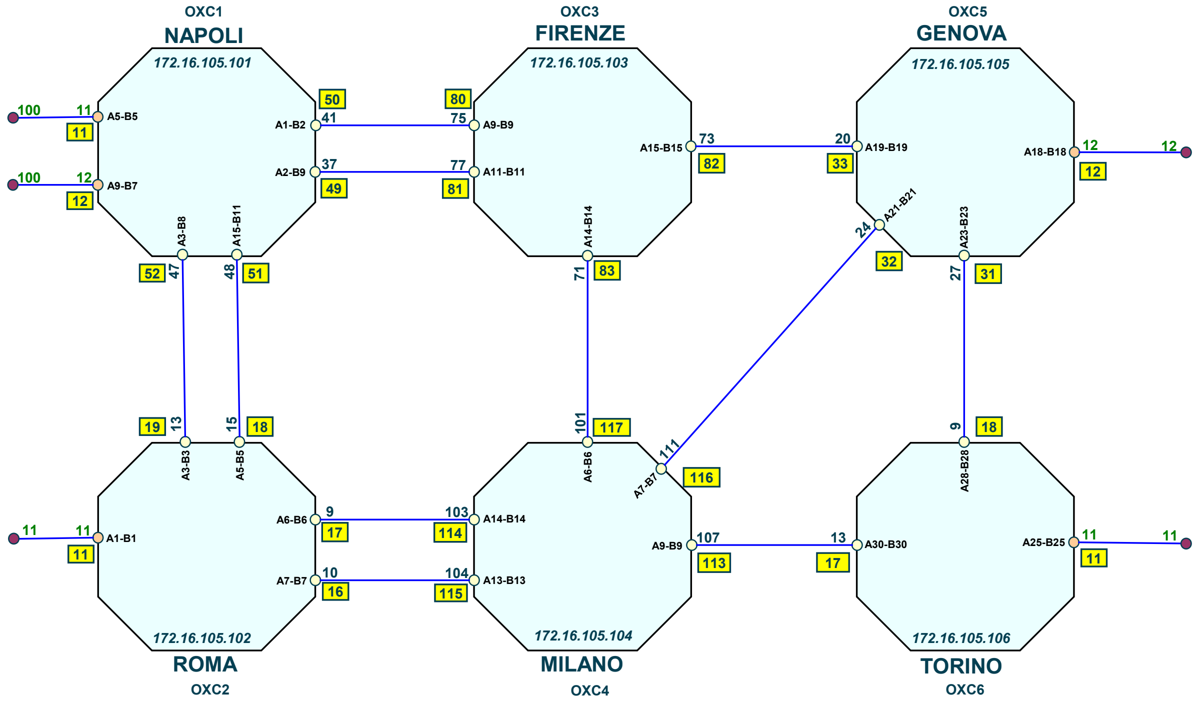


Figure 43 Final IDEALIST scenario, TI domain

7 Inter-work package feedback for WP1, WP2, WP4

7.1 Feedback from WP1

WP1 has three main relevant areas of work. First of all, WP1 defined the Elastic Optical network architectures that need the control and management functionalities developed in WP3. Second, WP1 has provided work on algorithms and optimization. WP3 has designed the control plane so as the necessary network information is used by the entities that perform the path computation functions. OSPF-TE and BGP-LS extensions in the case of the control plane, and REST APIs in the case of ANM, have been extended to include flexi-grid capabilities and transponder information. The signalling is also flexi-grid enhanced enabling WP1 proposals such as the spectrum defragmentation or re-optimization after failure links (in general, re-optimization strategies).

Finally, WP1 provides techno-economic analysis of the features developed in the project. This activity has provided relevant feedback to WP3 and is reported in depth in D1.5 [D15]. In particular, the impact of an adaptive network manager and an elastic-enabled control plane proposed in WP3 in terms of OPEX reduction (including reduced power consumption) in a network operator has been provided by WP1.

The adaptive network manager developed in WP3 provides a TCO (Total Costs of Ownership) reduction, in CAPEX thanks to the multi-layer service provisioning and restoration enabled by the ANM, and an OPEX reduction thanks to the IP-Optical coordination.

The control plane also allows a CAPEX reduction as it enables Spectrum Defragmentation, re-optimization after failure links, use of distance adaptive algorithms.

7.2 Feedback from WP2

WP3 received many relevant contributions from WP2. Dedicated joint activities were established to enable the careful control of the whole data plane architecture defined within WP2.

Three main aspects are here reported as examples of successful collaborations between WP2 and WP3.

The first aspect deals with the control of the spectrum resources according to the flexible grid. The concept of frequency slot, defined as the combination of the nominal central frequency and the slot width has been clarified from a technological perspective (e.g. filter configuration). Moreover, feedback has been provided by WP2 on the feasibility of practical data plane operations, including for example the implications related to the frequency slot configuration (e.g. on guard bands and filtering cascade effects due to the presence of non-ideal filters within bandwidth-variable WSSes).

The second aspect deals with the control of the S-BVT. A detailed list of transponder parameters has been identified by WP2 also considering multi-vendor interoperability aspects. The list includes modulation format, baud rate, sampling rate, FEC/coding type, DSP algorithm and other optical parameters (e.g. range of launch power, maximum received optical power, chromatic dispersion tolerance, maximum receiver reflectance, pre-FEC BER thresholds, Error Vector Magnitude, etc.). The difficulties in handling a large amount of parameters in the control plane has driven the joint WP2-WP3 decision to adopt the transponder class solution. This solution allows most of the parameters to be defined and specified in advance, without requiring to be included, distributed and agreed during control plane operations. Only a few subset of the defined parameters is then encompassed within the control plane, successfully achieving multi-vendor interoperability while satisfying scalability requirements.

The third aspects refers to the definition of advanced solutions for provisioning and re-optimization. The definition of such solutions also involved WP1. Defragmentation through hitless shifting (i.e. push-pull) is an example of effective interaction between control plane operations driving specifically designed data plane actions (e.g. laser retuning and filter re-shaping) according to re-optimization strategies identified within WP1. A further example refers to the DSP parameters identified by WP2 to be used by WP3 for monitoring purposes within the OAM Handler of the ABNO architecture.

7.3 Feedback from WP4

As reported in [D4.3], WP3 has benefitted from continuous feedback from the different control plane implementations (i.e. the test-beds with control plane only, and the ones with both control and data planes), as well as the interoperability tests and experiments executed in WP4. Implementing the proposed extensions by multiple partners ensures that the control plane definitions can be interpreted by separate teams with little to no ambiguity. Basic initial experiments have demonstrated that the IDEALIST control plane architecture proposed in WP3 is sound and can be implemented in small to mid-sized networks and domains reasonably well, with performance indicators that are expected as per the defined requirements. The feedback has covered aspects such as the architectures, the control plane protocol extensions, complexity of implementation and, most notably, feasibility assessment, validation and gathering of results and key performance indicators.

In all, WP4 has validated that a common control plane for a multi-vendor Flexi-grid network can be designed, implemented and validated in targeted scenarios covering a small number of domains and small to mid-sized domain sizes with expected results, and that protocol extensions can be defined where appropriate to cover the identified requirements. Issues related to scalability are mitigated by scoping the applicability of the proposed solutions. Feedback from WP4 has been identified in the sense that having only transparent connections spanning multiple vendors and domains is problematic when integrating both control and data planes, and variations of the proposed control plane architecture that support and use intermediate electrical conversion in intermediate points are required. This has motivated WP3 to extend to a third scenario that addresses this missing aspect (i.e. electronic regeneration).

Implementations and measurements have provided feedback in real, close-to-production networks in terms of scalability and feasibility, which have influenced the way WP3 defines the control plane: first, having a centralized PCE is feasible but presents a potential bottleneck if the domain size is large. This is not only due to the control plane protocol overhead, since that the number and size of messages is small related to the capabilities of the assumed control plane links, and that link data-rates/bandwidth are largely sufficient for this. The issues come from the number of supported TCP/PCEP/BGP-LS sessions and the processing requirements when such entities are executed in commodity hardware. This, in part, has motivated the use of hierarchical architectures based on segmentation and network abstraction, in which each PCE is responsible for its domain. The time it takes for a network state change (e.g. a Traffic Engineering update of a link) to propagate in the involved control plane functional entities has implications on the tolerable network dynamicity: experiments carried out have reported that an OSPF-TE LSA update can take on the order of second(s) to propagate to a national network, and that, in addition, this change needs to be reflected in the PCE and, in case of multi-domain networks with a H-PCE up to the parent PCE.

WP4 has provided feedback to WP3 in the sense that to have full control of the underlying hardware architectural and protocol extensions are required in aspects such as to control the diversity of hardware devices such as bandwidth variable transceivers, multi-flow transponders and Flexi-grid optical nodes, to report and disseminate the hardware capabilities to the upper functional entities and layers and to report on the status of physical layer nodes and links in what concerns the physical impairments.

Protocol extensions defined in WP3 have been subject to feedback from WP4 in terms of the implications when implemented and in view of standards conformance: when implementing a proposed set of protocol extensions, WP4 and WP3 have to carefully consider whether a proposed extension is in line with on-going similar efforts outside IDEALIST, and related to control plane definitions and implementations being addressed by third parties. In this context, it is of critical importance the coordination with standardization efforts carried in WP6. IDEALIST control plane solutions are specified in IETF internet drafts that are under the control of third parties and that are updated accordingly. WP4 implementations relying of the latest updated versions are aware of changes across versions and can report back. While the implications of these changes are usually of limited effect (change of message names, minor format changes, updates in the types or encodings of TLVs, etc.) they are commonly reported back to WP3.

Complexity of implementation, in terms of not only the protocol extensions but also architectural aspects has been a common criterion to choose from multiple options when deciding how to address a specific requirement. LABELSETs as defined in WP3 and in alignment with existing normative documents can be defined in multiple formats, including inclusive lists, exclusive lists, ranges or a using more recent bitmap encoding. Feedback from WP4 addressed what concerns complexity of implementation and scalability. For instance, defining a long list of labels (frequency ranges) may overflow object formats.

From a research perspective, one of the most useful feedback provided by WP4 to WP3 is related to the actual performance in real deployments, ideally as close as possible to production networks, which complements the performance evaluations which are obtained only using theoretical analysis and/or simulations. This aspect goes beyond the simple feasibility assessment and proof of concept, since the system must be robust enough to sustain continuous operations and stress conditions.

During the execution of the experiments, data related to service setup delay, control plane overhead and connection blocking probability (where applicable) is systematically collected and processed. This allows us to report on concrete and measurable objectives included in the description of work such as the latency in the automated provisioning of Flexi-grid optical connections with variable-sized frequency slots through a control plane with an average setup delay on the order of seconds for a reference European transport network (not considering hardware latencies).

Examples of obtained KPIs are related to the control plane latencies, path computation times, control plane overheads measuring actual messages and bit rates, or blocking probabilities when considering different routing and spectrum assignments. That said, some remarks apply: in experiments involving several partners, each partner test-bed is located in its own facility/location, which limits the value of some data, since it is not reflecting the fact that involved components will be commonly collocated in real scenarios.

8 Dissemination of WP3 main outcomes

The dissemination activities of IDEALIST WP3 results have been mainly in the scope of scientific publications (both in peer reviewed journals and magazines as well as in international conferences) and in the form of inputs for standardization.

In this regards, since the start of the work, WP3 has covered activities that are more experimental and some activities that aimed at having a direct impact on standardization. It is worth noting that, by policy, some SDOs do not address standardization activities (e.g. control plane aspects at the IETF CCAMP working group) for hardware and data plane elements that are not fully standardized (e.g. by the ITU-T). This is why some control plane procedures and protocol extensions covering more advanced hardware elements (such as targeting the dynamic control and configuration of S-BVTs) are not subject to standardization at this point.

8.1 Publications and dissemination

IDEALIST has carried out scientific dissemination in conferences such as the Optical Fiber Communications (OFC) conference, or the European Conference on Optical Communications (ECOC). Considering peer reviewed journals and magazines, WP3 has disseminated main results in journals such as IEEE/OSA Journal of Lightwave Technologies (JLT), IEEE/OSA Journal of Optical Communications and Networking (JOCN) or IEEE Communications Magazine.

A significant amount of publications are the result of the joint work of multiple partners, combining work performed in WP1 related to on-line planning and algorithms, WP2 for the underlying hardware models, WP3 for control plane aspects and WP4 for implementation and performance values.

Amongst the large number of publications stemming from WP3, it is worth mentioning the following, marking important milestones within IDEALIST WP3: first, the initial definition of the GMPLS/PCE architecture for a single domain [Funems13], the sketching of dynamic and adaptive control plane solutions for flexi-grid optical networks based on stateful PCE [Mun14] and the extension to the multi-domain context [Eucnc14].

Moreover, the ANM was implemented and demonstrated for multi-layer scenarios in [Agu14] [Agu15], while the integration of WP1 algorithms and planning tool, WP3 ANM and control plane using a front-/back-end architecture has been published in [Vel14a][Gif14a].

An example of successful WP2-WP3 collaboration is reported in [Cug15]. In this work, novel monitoring solutions on DSP parameters (e.g. symbol variance) are defined by WP2 and utilized to detect transmission degradations. Then, specific control plane procedures defined by WP3 are activated to enforce adaptation techniques of transmission parameters. Novel adaptation techniques are also defined by WP2 (e.g. LDPC code adaptation) and successfully applied to perform hitless reconfiguration.

8.2 Standardization activities in cooperation with WP6

Standardization has been a key target for IDEALIST WP3, not only as a means to enable internal inter-operability between partners implementations, but also to reach a broader community and to start building consensus about the design and implementation of a GMPLS/PCE control plane for Elastic Optical Networks based on flexi-grid.

This section presents the main highlights regarding WP3 driven standardization activities that have been done in the scope of IDEALIST, feeding task T6.3, which coordinates the standardization activities carried out in international standards organizations as part of IDEALIST. In cooperation with WP6, selected results from WP3 have been inputs for standardization, notably in the scope of the IETF CCAMP, PCE and TEAS working groups.

In short, WP3 has achieved a very good progress with standards activities for architecture and control plane, work principally within the IETF. IDEALIST has positioned WP3 control plane work excellently, built on very early start and continued drive. Where applicable, WP4 implementation is based on and following draft standards work. The full list of standardization outputs is given in WP6 [D6.3], but it is worth mentioning the following contributions.

Adaptive Network Manager and Control Plane Architecture standards

- [PCEFAQ] Unanswered Questions in the Path Computation Element Architecture Updated after discussions within the IDEALIST project about the

applicability of PCE to a number of control plane scenarios, Published as RFC 7399
- [ABNO] A PCE-based Architecture for Application-based Network Operations, Integrated use-cases and architectural concepts arising from IDEALIST
- [FlexiFw] Framework & Requirements for GMPLS-based control of Flexi-grid DWDM Networks

Control Plane Protocol standards

- [FlexiLabel] Generalized Labels for the Flexi-Grid in Lambda Switch Capable (LSC) Label Switching Routers Specification of the label to be used in all control plane protocols
- [FlexiRSVPTE] Signaling Extensions in support of Flexible Grid Specification of the signaling protocol extensions Adopted by CCAMP working group and almost ready for working group last call
- [FlexiOSPFTE] GMPLS OSPF-TE Extensions in support of Flexible Grid, Specification of the routing protocol extensions Adopted by CCAMP working group

A most notable feedback has been given by IDEALIST members in what concerns the implementation report of BGP-LS, which is considered a quasi-mandatory step in the Inter-domain Routing (IDR) working group of the IETF. In order to share network link-state and traffic engineering information collected with external components using the BGP routing protocol a new BGP Network Layer Reachability Information (NLRI) encoding format is required. IDEALIST implementations are reported in the current IETF internet draft: "BGP Link-State Information Distribution Implementation Report" (https://datatracker.ietf.org/doc/draft-ietf-idr-ls-distribution-impl), which has been submitted to the IESG for publication, and is expected to become an RFC shortly.

Other drafts worth mentioning are the PCEP extensions for GMPLS [PCEPGMPLS] and Extensions to PCEP for Hierarchical Path Computation Elements [HPCE] and a YANG data model for WSON and Flexi-Grid Optical Network.

9 Assessment of WP3 Objectives

This section presents the main IDEALIST objectives as stated in the Description of Work (DoW). The first part, refers to measurable objectives and how they have been achieved. The second part refers to objectives defined on a yearly basis and how these objectives have been completed. For the former, examples of metrics are provided in selected scenarios. For the latter, each entry contains the description of the objective, its completion level and any applicable remarks.

9.1 Measurable objectives

The following table refers to DoW table 2, objective 4: Control plane for elastic optical Networks

Parameter	Metrics	Remarks
Automated provisioning of flexi-grid optical connections with	Average setup delay on the order of seconds for a reference	Provisioning driven by an ANM via a hierarchical PCE in a pan-

variable-sized frequency slots through a control plane	European transport network (not considering hardware latencies)	European testbed of the order of hundreds of milliseconds [OFC2015] Likewise, provisioning in a Spanish-wide photonic mesh on the order of hundreds of milliseconds [Mun14][Mar14][Goz15][Goz-JOCN15] and additional ones. Data measured without hardware latencies. Provisioning in the control plane WP4 testbed affected by partners' remote premises.
Automated modification of the parameters (e.g. frequency slot width, center frequency, modulation, etc.) of existing flexi-grid optical connections through a control plane.	Average modification delay on the order of seconds for a reference European transport network (not considering hardware latencies)	Modification and re-optimization of flexi-grid parameters performed in a hitless way (i.e. with no traffic disruption) in the order of hundreds of milliseconds [Cug15, Pao15-PN].
Automated restoration of existing flexi-grid optical connections and its modification/re-routing trough a control plane.	Average restoration delay on the order of seconds for a reference European transport network (not considering hardware latencies)	Restoration implemented in terms of "make-before-break", restoring existing connections either one-by-one or in batches using a front-end and back-end PCE architecture [Mar14] on the order of hundreds of milliseconds.
Automated and interoperable multi-vendor elastic optical connection service provisioning in multi-domain and multi-layer (packet/flexi-grid) optical networks	2 demonstrations of working control plane prototypes, and publication of IETF documents.	Existing control plane implementations and prototypes. Presented / demonstrated e.g. OFC2015. IETF documents (3 RFCs, several working group drafts and individual contributions)
Adaptive management of flexi-grid connection provisioning service based on monitoring the current network resource status and on traffic forecasts	Adaptive provisioning service delay of flexi-grid connections on the order of minutes	Reported in [Gif14a][Agu15b]
Adaptive management of flexi-grid connection modification service for concurrent optimization / defragmentation based on monitoring the current network resource status and on traffic forecasts	Adaptive concurrent optimization service delay on the order of minutes.	Global Concurrent Optimization (GCO) using a split front-end / back-end PCE in the order of seconds (not considering hardware latencies) for national networks. [Gif14a]
Traffic monitoring for proactive estimation of optimal bandwidth for links in a flexi-grid optical network	Lossless, sampleless network flow collection at 10 Gbps, as input for the bandwidth estimation algorithm	Demonstrated in TID/Cisco testbed (D4.4.)
Traffic monitoring for proactive estimation of optimal bandwidth for links in a flexi-grid optical network	Precise characterization of flows and accurate estimation of optimal bandwidth (error below 10%)	Characterized using real-world traffic scenarios

9.2 Yearly objectives

Complementary to the previous section, this section enumerates the yearly objectives as stated in Table 3 of the DoW, in what refers to WP3, control plane and adaptive service management.

9.2.1 Year 1

9.2.1.1 Objective: Control plane reference architecture

Definition of the framework for applying GMPLS and PCE control to flexi-grid networks based on the functional analysis of the flexi-grid node architectures defined in WP2 and the network architectures and on-line algorithms in WP1.

The objective has been completed in Year 1, based on ITU-T documents [G.694.1] and [G.872]. The framework for the GMPLS and PCE control of Elastic Optical Networks was completed and documented in deliverable D3.1 [D3.1]. It is worth noting that IDEALIST members are the main editors of the framework document published at the IETF [FlexiFw], expected to become an RFC. The framework adopts agreed models of, for example, flexi-grid ROADMs and their capabilities; e.g. [D2.1] from WP2 was considered in the model of optical hardware. Algorithms (WP1) are focused on RSA, in-operation network planning and de-fragmentation. The architecture and requirements was presented in the FUNEMS conference [Funems13].

As presented, the proposed architecture(s) involve a GMPLS/PCE approach with PCE stateful capabilities and instantiation & BGP-LS to convey and disclose topological elements and abstractions. At the latest stage (Year 3) the reference architecture has been extended to accommodate elements such as S-BVTs.

9.2.1.2 Objective: Functional requirements

Specification of the functional requirements for a reference GMPLS control plane and PCE- (…) Specification of the functional GMPLS control and PCE architecture to provide dynamic and flexible/elastic optical connection service provisioning, recovery and concurrent reoptimization

Functional requirements for a GMPLS/PCE control plane have been identified and agreed. Identified requirements are related to i) dynamic provisioning of elastic optical connections, ii) dynamic modification of elastic optical connections, iii) restoration and iv) sliceability. The requirements constituted the basis for WP4 implementation, joint interworking and testbed validation, detailed in [D3.1], mapping into specific control protocol requirements. From these high level requirements, IDEALIST identified the main requirements for a GMPLS/PCE control plane, including the dynamic provisioning of elastic optical connections where, for a given route, different client traffic demands, characterized by different requested bit-rates, may be satisfied through different transmission parameters, or (via hardware advanced functionalities) for the dynamic adaptation of existing data plane connections, which can be achieved through the modification of selected transmission parameters (for example, bit-rate and modulation format adaptation without changing the frequency slot or shift of the nominal central frequency or of the frequency slot).

Likewise, the functional requirements for the ANM listed, including IP link provisioning, bandwidth adaptation, or global concurrent optimization.

9.2.1.3 Objective: Protocol extensions

Preliminary definition of GMPLS signaling, routing and PCE-based path computation protocol extensions for single-domain flexi-grid optical networks.

Complete, protocol extensions that cover the main functionalities and requirements have been conceived, implemented and submitted for standardization. Extensions have been defined for OSPF-TE, RSVP-TE and PCEP. The extensions target not only the establishment of media layer network channels, but also additional considerations integrated during Year 2 and 3. The extensions enable, notably, the dissemination of Nominal Central Frequencies on a TE link; the signalling of sender Tspec and Flowspec for media layer LSPs; the use of 64 bit labels that characterize frequency slots. Since the beginning, IDEALIST has also provided advanced control plane extensions for advanced functionalities such as hitless defragmentation.

9.2.1.4 Objective: Use cases for adaptive service and network interworking

Definition of the scenarios of application and use cases of an adaptive network management of the control plane services of a flexi-grid optical network through an Adaptive Network Manager (ANM) in collaboration with the PCE.

Use cases include Automatic IP Link provisioning, Dynamic Bandwidth Allocation based on traffic changes, periodic Dynamic Bandwidth Allocation, network re-optimization after restorations and multi-layer restorations. Detailed scenarios and workflows were reported in [D3.1].

9.2.1.5 Objective: Adaptive Network Manager Architecture

High level functional design of the Adaptive Network Manager (ANM) was completed, including preliminary definition of the functional blocks and interfaces. Identification of the technical alternatives for the functional blocks and interfaces.

9.2.2 Year 2

9.2.2.1 Objective: Single domain flexi-grid control plane architecture

Final definition and evaluation of a feature-complete control plane (covering both functional GMPLS and PCE architectures and protocol extensions) for dynamic and flexible/elastic optical connection service provisioning, modification, recovery and concurrent re-optimization in single-domain flexi-grid optical networks.

The architecture for a single domain control plane is considered final, based on the protocol extensions defined for OSPF-TE, RSVP-TE, PCEP and BGP-LS. Protocol extensions have been documented, implemented in the scope of WP4 and, where applicable, have been object of standardization in cooperation with WP6.

CP Basics (in terms of architectures and protocol extensions) was done in Year 1, based on ITU-T documents G.694.1 and G.872. Year 2 work has extended this in specific aspects: i) new use cases and requirements appearing during the execution of the project and not originally considered. This is the case of composite labels and to setup connections that concern multiple, non-necessarily contiguous frequency slots; ii) additional activities that are clearly experimental, research oriented and not apt for standardization at the IETF (constrained by actual, in force ITU-T recommendations with regard to the data plane) WP2 inputs and iii) topics such as the direct control of filters, or the control of hardware elements such as sliceable bandwidth variable transceivers.

The solutions have been documented in [D3.1] and [D3.2]. Concurrent reoptimization and recovery has also been validated in cooperation with WP1/WP3/WP4 with the integration of ANM, dynamic planning tools and PCE front-end and back-end architectures.

Selected studies have addresses the GMPLS/PCE-controlled Multi-Flow Optical Transponders in Flexi-grid Networks, including CP Extensions and performance evaluation; the use of Active Stateful PCE for LSP adaptation, including Hitless LDPC code adaptation PCEP extensions for suggested code type and rate; Source-Independent Frequency Conversion and hitless defrag (both for GMPLS/PCE and OpenFlow – with the frequency conversion and defrag op driven by CP); Fast restoration techniques, by means of SDN and stateful controller to enforce one-shot global reconfiguration or SuperFilter technique consolidation SSS filtering cascade effect.

9.2.2.2 Objective: Multi-domain interfaces

Definition of extended interfaces, protocols and procedures within the GMPLS control plane and Hierarchical PCE (H-PCE) architectures for interoperable end-to-end multi-domain path computation and provisioning in multi-vendor optical networks.

Complete. The concept of multi-domain control plane has been the main driver for Year 2 T3.1. Architectures and protocols extensions have been defined, specific scenarios for demonstration have been agreed upon between partners, and provided to WP4 for implementation and testing.

Firstly, IDEALIST CP architecture for multi-domain multi-vendor interworking and interoperability was presented in the EuCNC 2014 conference [Eucnc14]: The IDEALIST control plane clearly addresses the multi-domain (notably in the context of vendor islands) aspect by building on top of the hierarchical PCE architecture with the ANM abstracting and providing the interfaces for network operation. The H-PCE architecture has been extended to cover the active and stateful capabilities.

Protocol extensions have been listed, reported in documents MS7 and, notably, D3.2. The extensions have been implemented by multiple partners in WP4, providing feedback from WP4 regarding the PCEP, RSVP-TE, OSPF-TE and BGP-LS protocols. Three scenarios for multi-domain have been considered and implemented, including the first scenario with a transparent network and having a single end-to-end RSVP-TE session and delegating the provisioning to the provisioning manager that directly requests the provisioning of the service to the head-end node; the second scenario in which provisioning was delegated to the parent and children PCE (pPCE and cPCE respectively) that are active and with instantiation capabilities and, finally, a third scenario that includes multi-layer considerations and the role of transceivers in such a way that connections are established end to end with optional regeneration at the inter-domain links.

The experimental validation of a first scenario as the result of WP3-WP4 joint collaboration was accepted and presented in OFC2015 [Gon15], an extended version including scenario 2 has been accepted for publication in an upcoming special issue of Journal of Optical Communications and Networking. Scenario 3 is part of the final IDEALIST demo.

9.2.2.3 Objective: Control plane performance analysis

Evaluation of the feature-complete control plane and PCE architecture, interfaces and protocols extensions for multi-domain flexible/elastic connection service provisioning in multivendor flexi-grid optical networks.

The evaluation of the control plane is considered complete. Evaluated with a combination of different methods: i) Specific activities include performance evaluation by means of dedicated simulations, such as the recovery KPIs comparing GMPLS/PCE and SDN approaches, ii) Performance analysis on specific architectures defined by WP2 (e.g. SBVT either employing multi-wavelength technology or multiple laser sources) with appropriate control plane deployment solutions and evaluating through simulation dynamic networking scenarios. Finally, iii) for protocol extensions defined in WP3 for WP4 experimentation, WP4 carries out performance evaluation in a real setting, close to a production deployment scenario, with real hardware implementing elements such as GMPLS controllers or PCEs.

IDEALIST has obtained meaningful data regarding performance analysis in view of validating that the project Key Performance Indicators (e.g. such as connection provisioning in a multi-domain scenario in a pan European testbed) are below the project defined threshold.

For example Table 2, DoW Automated provisioning of flexi-grid optical connections with variable-sized frequency slots through a control plane can be stated to yield an average setup delay on the order of milliseconds for a reference European transport network (not considering optical hardware configuration latencies which are vendor specific), below the initially identified order of magnitude.

9.2.2.4 Objective: Adaptive Network Manager

Design and evaluation (by means of simulation and theoretical studies) of the different functional blocks of the ANM. Design of the monitoring system, ready to be imported in WP4. Definition of the interfaces to external elements, including the control plane for flexible/elastic optical. Extension of the ANM and PCE architectures to support adaptive concurrent optimization for defragmentation. Definition of topology (i.e. TED) service requirements and interface (i.e. Topology Server embedded in the ANM) providing network resource status for global concurrent optimization, ready to be implemented in WP4.

Complete. This objective encompasses multiple sub-objectives: ANM final architecture and workflows is documented in [D3.2], and milestone M24. Interfaces extending e.g. PCEP or based on RESTful documented interfaces and APIs (e.g. the provisioning interface between the ANM and the control plane is based on PCEP with stateful extensions and instantiation capabilities). Yang models for network topology and inventory, as well as for connection database have been defined, reported both in [D3.2] and as initial set of contributions to the IETF. Experimental validations in WP4 covering the performance evaluation of the algorithms of the Monitoring subsystem, the implementation of the topology service to send the topology to the PCE and planning tool and the integration of the Adaptive Network Manager with the control plane, PCE and the rest of the elements.

9.2.3 Year 3

9.2.3.1 Objective: Multilayer control plane architecture

Definition of the signaling, routing and path computation protocol extensions to the GMPLS unified control and PCE for end-to-end provisioning and inter-layer path computation in multi-layer (packet/flexi-grid) optical networks.

Complete. The IDEALIST multi-layer control plane relies on the deployment of a GMPLS/PCE control plane that meets the requirements for the establishment of (network) media channels in the so called media layer, the configuration of the S-BVTs for the signal layer and the deployment of an ANM that coordinates both the provisioning of optical services and IP/MPLS services by either delegating to the MPLS (packet) control plane and/or directly configuring the IP routers that use the optical connections to cover the use cases identified in Year 1 (such as the IP link provisioning use case). WP3 has extended Multi-Domain aspects with regenerators and covering the cases of restoration. Considered the applicability of Interconnected TE info and abstraction to multi-layer aspects. Addressed Multi-Layer provisioning with ANM/ABNO with underlying mechanisms supported by GMPLS and/or SDN, including multi-Layer Path Computation with a PCE (single or multiple PCEs). Integration within the ANM of IP bandwidth estimation via the OAM handler.

9.2.3.2 Objective: Multilayer interfaces

Definition of extended interfaces and protocol extensions to the (hierarchical) PCE-based architecture for end-to-end provisioning and inter-layer path computation in multi-layer (packet/flexi-grid) optical networks.

Multi-layer extensions to the interfaces mainly address the packet layer in addition to the interfaces conceived during Year 2. The coordination of the layers is left as a task of the ANM controller, which coordinates the establishment of the (network) media channel and the configuration of the client signals.

The packet layer has been addressed by having the controller request the IP/MPLS layer (using interfaces such as PCEP or REST/JSON), directly using the OpenFlow [OpenFlow] protocol to configure OpenFlow enabled routers/switches to aggregate packet traffic into optical channels [May15].

New architectures such as the one based on Segment Routing have also been considered, so the packet layer is controlled by a PCE which uses optical tunnels (e.g. media channels) as specific segments in a multi-layer network.

9.2.3.3 Objective: Performance Analysis

Evaluation of the feature-complete control plane and PCE architecture, interfaces and protocols extensions for end-to-end flexible/elastic connection service provisioning in multi-layer (packet/flexi-grid) optical networks.

The implementations of the IDEALIST solutions within WP4 have enabled the performance analysis of the overall solution. Performance results are reported in IDEALIST deliverable D4.4 [D4.4].

9.2.3.4 Objective: Adaptive Network Manager

Final design of the ANM based on the feedback of the implementation from WP4. Final validation of the ANM, including scalability and adaptive concurrent optimization for defragmentation.

The ANM design is complete. The ANM has been implemented and tested in the scope of WP4, where the main use cases identified within WP3 and reported in [D3.1] have been validated, including IP/MPLS provisioning [Agu15].

The Resilience use case for multi-layer has been performed in Telefonica's lab in Madrid. The Bandwidth Re-Allocation use case including traffic variation has also been demonstrated including the integration with the probe systems and OAM Handler. Finally, the re-optimization process was also reported in [Gif14a].

10 Conclusions

IDEALIST WP3 has covered with the functional analysis, architecture, protocol design and evaluation of a control plane and a service management system, in order to introduce dynamicity, elasticity and adaptation in flexi-grid optical networks. As such, WP3 has completed the design and specification of the GMPLS control plane with PCE path computation for interoperable and scalable multi-vendor flexi-grid optical networks, including multi-domain and multi-layer scenarios, in order to provide dynamic and flexible, elastic optical connection service provisioning, recovery and concurrent re-optimization and defragmentation.

During the execution of the work, WP3 has considered both activities that target standardization such as the design of a control plane for the dynamic recovery and provisioning of (network) media channels assuming agreed and simple models of flexi-grid media layer matrices, as well as more research oriented not in the scope of current standardization charters (either being out of scope, or not having the underlying data plane in a closed standards), such as the direct control of transceivers which are non-standard or research prototypes. This has proven to be an efficient and flexible approach ensuring direct impact on standardization effort while still researching in advanced devices, nodes and network architectures.

Additionally, since a feature complete adaptive system requires functions that may only be available at the network management level (such as dedicated monitoring or the intelligence to react and self-adapt), WP3 has covered the design and the specification of a network management system for the adaptive management of the elastic optical connectivity services, complementary to the control plane, based on monitoring the current network resource status and on traffic forecasts to adaptively plan and decide on modifying parameters of an existing elastic connection (e.g. modulation and spectrum allocation), adding/removing elastic connections, or de-fragmenting the existing elastic connections.

IDEALIST WP3 has had an excellent impact, across scientific dissemination within peer reviewed conferences, journals and magazines; and in terms of contributions to standards. The main documents in the IETF CCAMP/PCE working groups related to a GMPLS control plane for flexi-grid networks and ABNO are authored by IDEALIST members.

All the objectives related to WP3 have been completed, both in terms of objectives per year as well as in terms of technical (measurable) parameters, in which metrics have been obtained by a combination of either simulations or feedback from real implementations in the scope of WP4. This has been reported in the deliverable, with details on a per objective basis and where the details are reported.

There are some issues that are harder to solve, and cannot be solved by IDEALIST project alone. In particular, not having standard data plane components across multiple vendors prevents the definition of a common control plane with advanced capabilities, and to some extent, vendor specific extensions can be required to fully exploit the capabilities of the hardware. The IDEALIST approach addresses this by scoping, as much as possible, such extensions to their own domains and exporting, in a global network, only as much abstracted details as required, and relying mainly in open and agreed formats and encodings. This of course may impact and limit the feature set that can be leveraged, but it is a trade-off between the abstraction and extensions.