STRONGEST – Document

Deliverable D3.1

Medium-term multi-domain reference model and architecture for OAM, control plane and e2e services

Version and Status:	Version 2.0, final	
Date of issue:	31.08.2010	
Distribution:	Public	
Author(s):	Name	Partner
	Berechya, David (editor chapter 5)	NSN
	Bincoletto, Luca	TI
	Botham, Paul	BT
	Broniecki, Ulrich (editor chapter 3)	ALUD
	Casellas, Ramon	CTTC
	Castoldi, Piero	CNIT
	Corliano Gabriele	BT
	Cugini, Filippo (editor chapter 4)	CNIT
	Di Giglio, Andrea	TI
	Garcia Argos, Carlos	TID
	González de Dios, Oscar	TID
	Iovanna, Paola (editor chapter 2)	TEI
	Javier Jimenez Chico, Francisco	TID
	Lautenschläger, Wolfram	ALUD
	Loffredo, Tullio	TI
	Maier, Guido	CNIT

	Marchetti, Loris	TI
	Margaria, Cyril	NSN-G
	Martinez, Ricardo	CTTC
	Milbrandt, Jens (deliverable editor)	ALUD
	Morro, Roberto	TI
	Muñoz, Raul	CTTC
	Paolucci, Francesco	CNIT
	Pulverer, Klaus	NSN-G
	Rambach, Franz	NSN-G
	Sfeir, Elie	NSN-G
	Siracusa, Domenico	CNIT
	Vezzoni, Emilio	VECOMM
	Zema, Cristiano	TEI
	Zuban, Alexander	PRI
Checked by:	Vezzoni, Emilio	VECOMM
	Di Giglio, Andrea	TI
Approved by:	Iovanna, Paola (WP3 leader)	TEI

Abstract

The STRONGEST WP3 on “end-to-end solutions for efficient networks” aims at providing efficient solutions to support end-to-end service delivery crossing domains that are heterogeneous in terms of networking technologies, control plane models, and vendors/operators.

The activities reported here are focused on medium-term network scenarios which address the interworking between heterogeneous GMPLS-controlled networks such as WSON and MPLS-TP networks. The corresponding tasks are dedicated to the definitions of (1) the network reference scenarios, (2) the OAM reference model, (3) the control plane reference architecture, and (4) the end-to-end services.

Table of contents

Executive summary

The activities reported in this deliverable are focused on medium-term network scenarios which address the interworking between heterogeneous GMPLS-controlled networks such as WSON and MPLS-TP. The corresponding tasks are dedicated to the definitions of (1) the envisaged network reference scenarios, (2) the OAM reference model, (3) the control plane reference architecture, and (4) the end-to-end services.

In more detail, the tasks named above are each dedicated their own chapters and have the following contents.

As a first step towards the definition of new architectures and strategies for OAM, control plane and end-to-end services, a set of 3 network reference scenarios is defined in Chapter 2. Reference scenario 1 reflects a single-domain, multi-region, single-carrier network, while reference scenarios 2 and 3 represent multi-domain, multi-region networks belonging to either a single-carrier carrier or to multiple carriers. In order to consider the most typical and challenging network situations and to have a common starting platform for WP3 work, we focus on reference scenarios 2 and 3 throughout the remainder of this deliverable.

OAM (Operations, Administration and Maintenance) plays a noteworthy role in telecommunication networks, providing procedures for fault management and performance monitoring. Chapter 3 summarizes the state of the art of OAM in MPLS-TP and WSON networks, defines OAM requirements for the STRONGEST network reference scenarios, and gives first insights into the STRONGEST OAM framework and associated future work.

With the growing size of transport networks and their dynamic label switched path management, automatic provisioning coupled with traffic engineering has become essential to operate them cost-efficiently. To increase flexible transport capacity and improve traffic engineering, a STRONGEST control plane reference architecture based on generalized multi-protocol label switching is proposed. Therefore, Chapter 4 summarizes the state of the art of control plane standards and technologies, defines the control plane requirements for the STRONGEST network reference scenarios, and derives a proposal for a STRONGEST control plane architecture and resulting future work.

The goal of any transport network is to provide the infrastructure for network services which determine many network requirements and have a major impact on network technology and design. Therefore, Chapter 5 summarizes the state of the art of currently deployed network services and their attributes, defines new end-to-end services and their requirements with respect to the STRONGEST network reference scenarios, and concludes with a description of future work.

1 Introduction

In this report, the STRONGEST WP3 group presents their view of a medium-term multi-domain reference model and architecture for OAM, control plane and e2e services. The document is organized as follows:

As a first step towards the definition of new architectures and strategies for OAM, control plane and end-to-end services, a set of STRONGEST network reference scenarios is defined in Chapter 2 in order to consider the most typical and challenging network situations and to have a common starting platform for WP3 work.

OAM (Operations, Administration and Maintenance) plays a noteworthy role in telecommunication networks, providing procedures for fault management and performance monitoring. The STRONGEST OAM concept presented in Chapter 3 is applicable to a network's transport and service layers in order to improve its ability to support services with guaranteed and strict service level agreements while reducing their operational costs.

With the growing size of transport networks and their dynamic label switched path management, automatic provisioning coupled with traffic engineering has become essential to operate them cost-efficiently. To increase flexible transport capacity and improve traffic engineering, a STRONGEST control plane reference architecture based on generalized multi-protocol label switching is described in Chapter 4.

The goal of any transport network is to provide the infrastructure for network services. As shown in Chapter 5, the definitions of these services determine many network requirements and have a major impact on network technology and design.

2 Network reference scenarios

As a first step towards the definition of new architectures and strategies for OAM, Control Plane and end-to-end services, a set of reference scenarios was defined in the context of STRONGEST Project, in order to consider the most typical and challenging network situations and have a common starting platform for WP3 work.

When modeling the control plane reference scenarios, several aspects were considered, such as:

- **Number of Domains -** A domain is considered to be any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include IGP areas and Autonomous Systems;
- **Number of Carriers** – A carrier is an organization that provides communications and networking services;
- **Type of Regions** – An LSP region, as well as its boundaries, is constructed and identified by the information carried in the Interface Switching Capability Descriptor (ISCD), as stated in [RFC 4206].

According to the above mentioned aspects, three different scenarios were identified as the STRONGEST medium-term reference ones, with the aim of covering a larger set of possible ones.

As a matter of fact, the three reference scenarios arise from the consideration of:

- Single-domain vs. multi-domain contexts
- Single-carrier vs. multi-carrier contexts

Moreover, each reference scenario is a multi-region one, i.e. it is composed by elements belonging to different technologies and having different switching capabilities. Particularly, Multi-Protocol Label Switching – Transport Profile (MPLS-TP) and Wavelength Switched Optical Networks (WSON) were considered as reference Packet Switching Capability (PSC) and Lambda Switching Capability (LSC) networks respectively.

A more detailed description of the STRONGEST Control Plane Reference Scenarios is reported in the following sections.

2.1 Reference scenario 1: single-domain / multi-region / single-carrier

The first scenario to be considered is the *STRONGEST reference scenario 1*, shown in Figure 1.

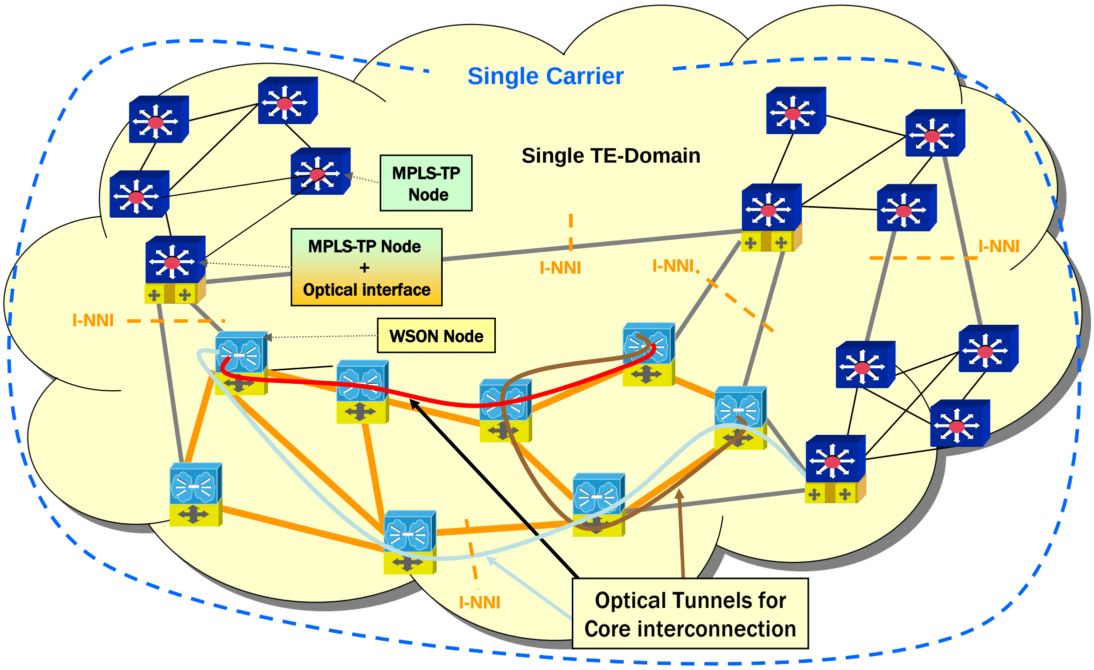


Figure 1 – Reference scenario 1: single domain / multi-region / single carrier

The considered network is a single domain one, composed by nodes belonging to different technologies, which are managed and administrated by a single carrier. Particularly, the network topology is a meshed one composed by a MPLS-TP region and a WSON region. MPLS-TP and WSON regions are interconnected by means of optical interfaces equipped on MPLS-TP border nodes.

Since the reference scenario 1 refers to a single TE-domain, all interfaces are considered to be Internal Network to Network Interfaces (I-NNI) and a full visibility of the network topology is allowed. Therefore, the exchanged routing and signaling information refers to both optical and packet impairments and characteristics.

Intra-region connections (i.e. packet LSPs connecting only MPLS-TP nodes or optical LSPs connecting only WSON nodes) are obtained by setting up LSPs of the corresponding technology, while intra-region connections (i.e. LSPs crossing both MPLS-TP nodes and WSON ones) are obtained by means of Forwarding Adjacencies (FAs) [RFC 4206].

In other words, when crossing different regions (and layers), a WSON LSP is announced as a TE Link into the same instance of the GMPLS control plane as the one that was used to create it. A MPLS-TP LSP, then can be set up using this FA (i.e. the higher layer LSP is nested in the lower layer one).

It is worth to make some additional considerations about optical LSPs. WSON LSPs can be considered as core ones for their high bandwidth granularity (i.e. wavelength bandwidth) and their set up time is higher than a packet LSP. Therefore, a possible approach is to pre-configure a set of WSON LSPs connecting the boundaries of the WSON region, and advertise them as FAs for the MPLS-TP region.

This approach can be followed also in case of multi-layer connections within PSC clouds, without involving LSC ones. As a matter of fact, a set of MPLS-TP LSPs can be nested in a lower level MPLS-TP one (exploiting the MPLS-TP hierarchy).

Since the reference scenario 1 refers to single-domain and single-carrier case, it should be considered as the simplest scenario among the ones considered within STRONGEST. It can also be considered as a starting point for analysis and solutions study that can be extended to the more complex multi-domain and/or multi-carrier case.

A set of possible activities that can be mapped into reference scenario 1 are:

- Analysis of routing information to be carried in a multi-layer network / multi-region network (MLN/MRN) (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- Analysis of signaling information to be carried in MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- OAM analysis for MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- Scalability studies, such as:
    - The correlation between number of nodes that can be managed in a single domain and its scalability impact
    - The impact of considering pre-configured WSON LSPs, or dynamically created ones, or a mix of the two, in terms of setup time, blocking probability, etc.

2.2 Reference scenario 2: multi-domain / multi-region / single-carrier

The second scenario to be considered is the *STRONGEST reference scenario 2*, shown in Figure 2. The considered network is a multi-domain one, composed by nodes belonging to different technologies, which are managed and administrated by a single carrier. The network topology is a meshed one composed by three MPLS-TP regions and two WSON region that are interconnected by means of optical interfaces equipped on WSON nodes and MPLS-TP border nodes.

Reference scenario 2 refers to multiple TE-domains, in order to improve network scalability and to make the network management more flexible. Therefore, two kinds of interfaces are considered: Internal Network to Network Interfaces (I-NNI), i.e. the interfaces connecting nodes belonging to the same domain, and External Network to Network Interfaces (E-NNI), i.e. the ones interconnecting border nodes of different domains. Due to the multi-domain approach, a full visibility of the whole network topology is not allowed. Each TE-domain has the full visibility of its own topology and resources, so that the routing and signaling information that need to cross different domains are exchanged by means of E-NNI interfaces.

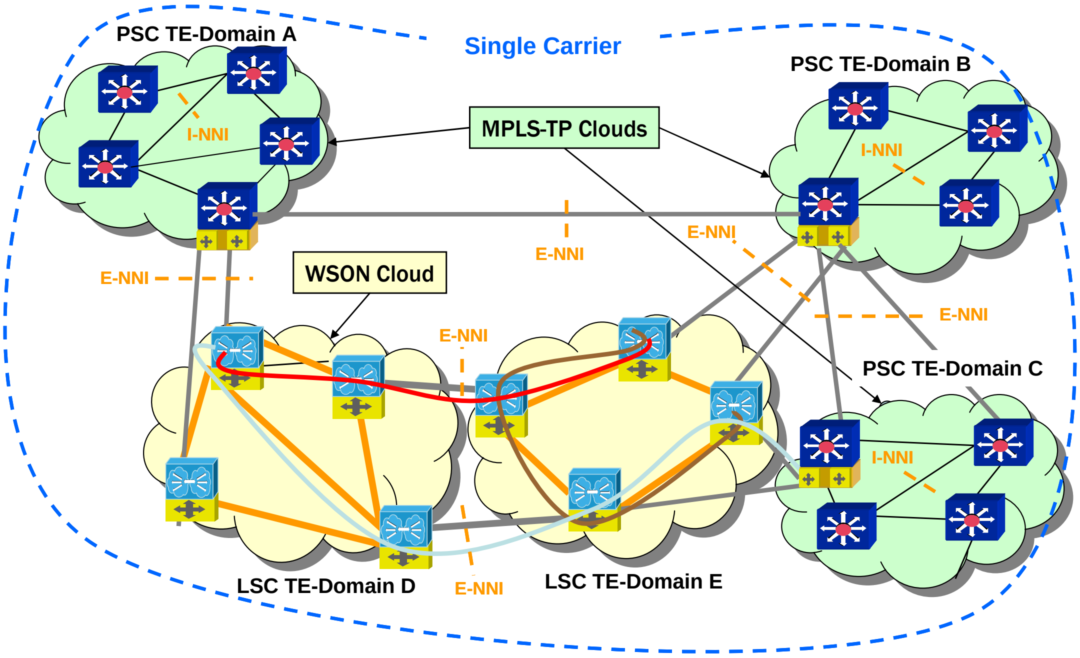


Figure 2 – Reference scenario 2: multi-domain / multi-region / single carrier

The reference scenario 2 can also reflect a control plane hierarchical architecture, where the first level is represented by all the single domains, exchanging routing and topology information with a higher level that has a summarized view of the whole network.

According to hierarchical control plane architecture, the topology and resources of each domain can be advertised to the higher level with a certain level of abstraction, in terms of both resource and topology information. Methods and algorithms to perform topology summarization should consider both packet and optical impairments and characteristics, in order to summarize in a simple way all the needed information for an optimized MLN/MRN path computation.

On the contrary of what happens in reference scenario 1, within reference scenario 2 also intra-region connections can span different domains. For the sake of simplicity, but without loss of generality, in the reference scenario 2 a domain is composed by only one technology (i.e. region), while a region can be composed by more than one domain (WSON region is composed by two WSON domains and MPLS-TP region is composed by Three MPLS-TP domains).

In the same way as the reference scenario 1, intra-region connections are obtained by setting up LSPs of the corresponding technology, while intra-region connections are obtained by means of Forwarding Adjacencies (FAs) [RFC 4206]. It is worth to notice that the FAs approach applies whenever the considered end-to-end connection is obtained in a multi-layer fashion, while E-NNI interfaces and/or topology summarization are used whenever an end-to-end connection spans two or more domains.

WSON domains can be considered as core ones according to the same considerations made for the reference scenario 1. Therefore, also in the reference scenario 2, a possible approach is to pre-configure a set of WSON LSPs connecting the boundaries of the WSON region, and advertise them as FAs for the MPLS-TP region. This approach can be followed also in case of multi-layer connections within PSC clouds, without involving LSC ones, exploiting the MPLS-TP hierarchy.

Reference scenario 2 adds a degree of complexity to the networks represented by reference scenario 1. As a matter of fact, multiple TE-domains are considered, but still belonging to a single carrier.

This scenario reflects the most challenging and interesting network architectures actually under deployment within vendors, providers and standardization bodies. As a consequence, most of the activities within the WP3 of the STRONGEST Project can be mapped in into reference scenario 2, such as the following ones:

- Analysis of multi-domain issues in the context of MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- Analysis of routing information to be carried across different domains within MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- Analysis of signaling information to be carried across different domains within MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)
- E-NNI interface analysis and development in terms of:
    - definition and/or extensions of technology-dependent impairments (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios) for hierarchical architectures

  - impact of the above defined/extended E-NNI on path computation

- Analysis and solutions development of hierarchical architectures (e.g. Hierarchical PCE, TISPAN RACS, etc)

- Mapping of hierarchical approaches (as the ones described in Chapter 4) into the considered domains' technologies (i.e. MPLS-TP and WSON)

- Topology summarization methods, in terms of:

  - Algorithms

  - Technology-dependent constraints

  - General summarization criteria

- OAM analysis for multi-domain MLN/MRN networks (particularly MPLS-TP and WSON technologies, considered for all STRONGEST Reference Scenarios)

- Scalability studies, such as the correlation between the domain division of a network (i.e. the number of nodes within a domain and the number of domains within a hierarchical architecture) and its scalability impact

2.3 Reference scenario 3: multi-domain / multi-region / multi-carrier

The third (and last) scenario to be considered is the *STRONGEST reference scenario 3*, shown in Figure 3.

The considered network is a multi-domain one, composed of nodes belonging to different technologies, which are managed and administrated by different carriers. The network topology is a meshed one composed by three MPLS-TP regions and one WSON region that are interconnected by means of optical interfaces equipped on WSON nodes and MPLS-TP border nodes.

As in the reference scenario 2, also reference scenario 3 refers to multiple TE-domains, in order to improve network scalability and to make the network management more flexible. Therefore, both I-NNI and E-NNI interfaces are considered.

Due to the multi-domain approach, a full visibility of the whole network topology is not allowed. Each TE-domain has the full visibility of its own topology and resources, so that the routing and signaling information that need to cross different domains are exchanged by means of E-NNI interfaces.

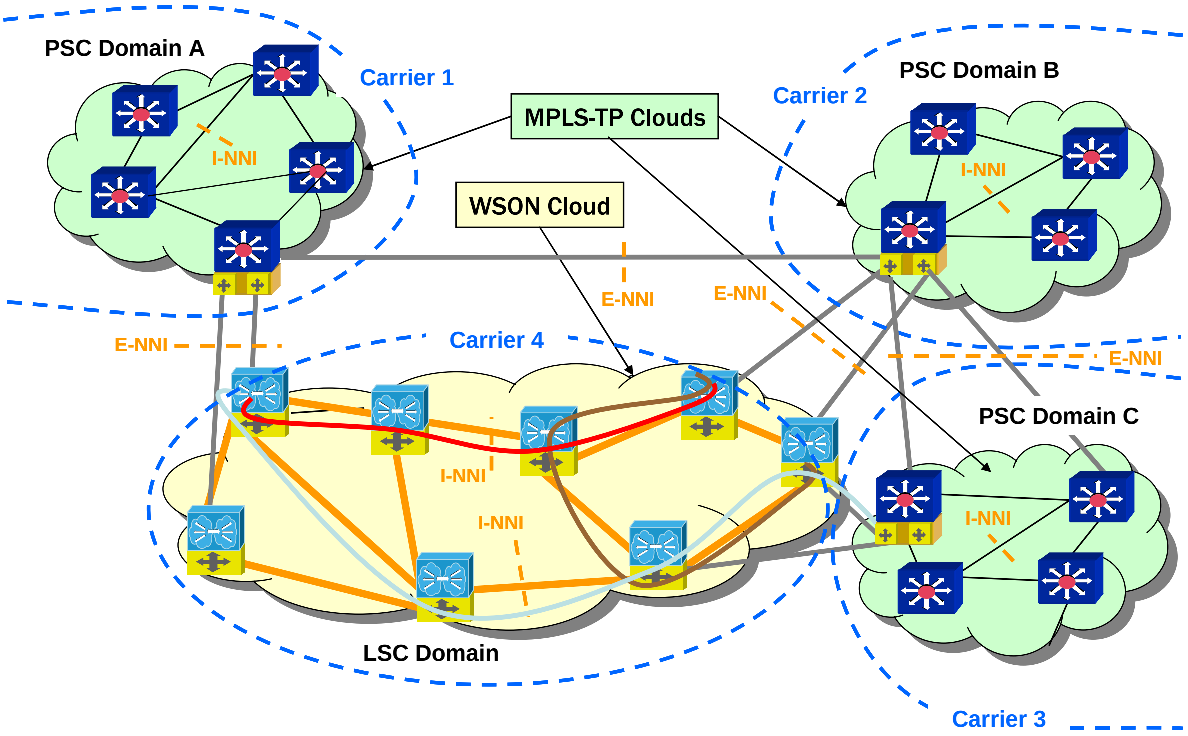


Figure 3 – Reference scenario 3: multi-domain / multi-region / multi-carrier

In the same way as the reference scenario 1 and 2, intra-region connections are obtained by setting up LSPs of the corresponding technology, while intra-region connections are obtained by means of Forwarding Adjacencies (FAs) [RFC 4206]. It is worth to notice that the FAs approach applies whenever the considered end-to-end connection is obtained in a multi-layer fashion, while E-NNI interfaces and/or topology summarization are used whenever an end-to-end connection spans two or more domains.

Reference scenario 3 adds another degree of complexity to the networks represented by reference scenarios 1 and 2. As a matter of fact, multiple TE-domains are considered, each one belonging to a different carrier. For the sake of simplicity, but without loss of generality, in the reference scenario 3 each carrier administrates a single domain composed by only one technology (i.e. region), while a region (MPLS-TP one in the reference scenario 3) can be composed by more than one domain (and so administrated by different carriers). In the context of the STRONGEST project, the reference scenario 3 will be only explored and the multi-carrier related issues will be investigated.

As a matter of fact, the very large set of challenging issues introduced by the multi-carrier context will be considered as low priority activities, subordinated to the ones related to Reference Scenarios 1 and 2, that will be deeply investigated and whose solutions will be developed and analyzed.

As the reference scenario 2, also the reference scenario 3 can reflect a control plane hierarchical architecture, but the multi-carrier context impacts very deeply the level of abstraction of the topology and resources information of each domain.

As a matter of fact, a set of activities within the WP3 of the STRONGEST Project can be mapped in into reference scenario 3, in order to fix the following issues:

- Analysis of Confidentiality issues – Each domain is administrated by a different carrier, so a set of domain information such as: topology, resources, objective functions and metrics used for path computation, cost of connections, etc. must not be disclosed to other domains (i.e. to other carriers)

- Analysis of security issues – Security strategies should be considered in order to avoid a set of possible situations, such as: malicious actions in order to retain sensible information from other domains, the advertisement of not fair information in order to favor the crossing of some domains, etc.

- Analysis of Topology summarization issues – Each domain can apply its internal policies and strategies, so that the summarized information can be obtained according to heterogeneous criteria, methods and algorithms

- Analysis of Economic issues – Service Level Agreement should take place between different domains (i.e. carriers) and path computation should take it into account. Moreover, malicious policies in order to favor a domain with cheaper connections (independently of their optimality) should be avoid

3 OAM reference model

OAM (Operations, Administration and Maintenance) plays a noteworthy role in telecommunication networks, providing procedures for fault management (detection and localization) and performance monitoring (loss/degrade of received information, delay). The OAM concept might be applied to both the transport and the service layers in order to improve their ability to support services with guaranteed and strict Service Level Agreements (SLAs), while reducing their operational costs.

Traditional transport networks (i.e. SDH and OTN) are endowed with very complete and highly performing mechanisms of Operation, Administration and Maintenance (OAM), including Performance Monitoring (PM).

On the other hand existing intra-office Ethernet offers very poor OAM and PM instruments; when the Ethernet technology was extended to long distance packet transport (especially in metropolitan networks) the development and standardization of OAM features for packet networks became mandatory, to ensure suitable quality.

For this reason, both in IEEE and ITU-T, respectively, Recommendations Y.1731 and 802.1ag have specified the OAM features for Ethernet networks, with primary aims to detect failures, to trigger the network protections and to locate faults. These Recommendations are the basis for OAM in emerging packet transport networks.

The advanced packet transport technologies (in particular MPLS-TP) have been designed to extend the concept of packet connectivity for transport purposes. For this reason ITU-T has developed a set of requirements for the transport profile of MPLS (MPLS-TP, TP = Transport Profile) and those regarding OAM represent an important section. Among the required functions, the most important are: a very fast fault detection and localization, the estimation of packet loss and the delay measurement.

The satisfaction of these requirements can be reached in two ways, differing essentially for implementation and not for functionality; this alternative is currently the object of a heated debate in the standardization bodies meetings.

The first proposal means to use the framework and the tool-set defined by the Bi-directional Forwarding Detection (BFD), a mechanism based on IETF RFC 5884 and widely adopted in routers. This approach is supported by traditional IP vendors and by the majority of American Carriers.

The competing approach is based on the extension to MPLS-TP of the tool-set defined by ITU-T Y.1731 that, as mentioned above, specifies the OAM features for Ethernet networks. This approach is supported by traditional vendors of transmission equipment and by the majority of incumbent European Carriers and emerging Chinese Operators.

At this stage, STRONGEST is not taking a position on the two competing OAM standardization proposals, but is rather studying OAM functions complying with its objectives and matching its target network architecture.

3.1 State of the art

Deployed transport networks based on consolidated technologies are endowed with complete and highly performing OAM features. On the other hand, networks based on emerging technologies still require a further, accurate definition of comparable OAM functions since, at the state of the art, they provide adequate OAM mechanisms, but only bounded inside a single, homogeneous technology; mechanisms allowing the monitoring of end-to-end transport services through multi-region (i.e. multi-technology) networks are not only lacking, but they are not even currently considered by standardization bodies and fora.

This section describes the state of the art of OAM in transport networks, focusing on the transport technologies that STRONGEST is adopting for its mid-term scenario (i.e. MPLS-TP and WSON).

3.1.1 OAM for MPLS-TP

MPLS-TP is one of the two key transport technologies that STRONGEST takes into consideration for its mid-term network solution. In the following subsections the state of the art regarding OAM for MPLS-TP network layers is described, introducing requirements as defined by IETF, ITU-T and MEF, listing the OAM mechanism complying with such requirements

3.1.1.1 OAM requirements

Requirements for Operational, Administration and Maintenance have already been defined in details in by ITU-T, IETF and MEF, regarding the single-domain scenario. In particular, after deep debate, ITU-T and IETF reached an agreement having as a background the previous experience from SDH networks and as specific focus MPLS and Ethernet technologies. MEF has also defined OAM requirements, typically at a higher level, more focused on services monitoring.

In more details, IETF and ITU-T provided OAM requirements, identifying 2 main categories: architectural requirements and functional requirements.

The most important architectural requirements are:

- Independence of MPLS-TP layers (at OAM level) from service and underlying networks. In other terms, as reported in [draft-vigoureux] “The set of OAM functions must be a self-sufficient set that does not require external capabilities to achieve the OAM objectives”

- Bidirectional application of OAM functions

- Application of OAM functions to unidirectional point-to-point and point-to-multipoint connections

The functional requirements might be split into two categories with regard to the task they are facing with: fault localization and performance monitoring.

The main OAM mechanisms required by the joint ITU-T – IETF working group for fault management are:

- Continuity check / verification
- Alarm suppression
- Lock indication
- Diagnostic test
- Trace-route
- Remote defect indication

The main OAM mechanisms required by the joint ITU-T – IEFT working group for performance monitoring are:

- Packet loss measurement
- Delay measurement

On the other end MEF, more focused on service OAM, has specified the following list of requirements:

- Service OAM should discover other elements in the Metro Ethernet Networks (MEN)
- Service OAM should monitor the connectivity status of other elements (active, not-active, partially active).
- Performance monitoring should estimate Frame Loss Ratio (FLR) Performance, Frame Delay Performance, and Frame Delay Variation (FDV) Performance
- In a multi-domain environment OAM frames should be prevented from “leaking” outside the appropriate OAM domain to which they apply.
- The OAM frames should traverse the same paths as the service frames
- The OAM should be independent of but allow interoperability with the underlying transport layer and its OAM capabilities
- The OAM should be independent of the application layer technologies and OAM capabilities

The concept of Maintenance Entity (ME) is included in the OAM functions specified by MEF (similar to Y.1731), where a ME represents an OAM entity that requires management.

MEs are typically involved in different OAM domains. For the purposes of Service OAM, MEF focuses on UNI MEs, EVC MEs, and Subscriber MEs. The EVC MEs may span several domains. Requirements associated with E-NNI MEs are expected to be covered in future versions of MEF 17.

3.1.1.2 OAM mechanisms

The scenario for OAM in MPLS-TP is noticeably described in [oam-analysis]. This section reports an extract of the main concepts included in [oam-analysis]. The following OAM tools (either from the current MPLS tool-set or from the ITU-T documents) can be applied to meet the OAM requirements:

- **LSP Ping:** LSP Ping extends the basic ICMP Ping operation (of data-plane connectivity and continuity check) with functionality to verify data-plane vs. control-plane consistency for Forwarding Equivalence Class (FEC) and also Maximum Transmission Unit (MTU) problems. The trace-route functionality may be used to isolate and localize the MPLS faults, using the Time-to-live (TTL) indicator to incrementally identify the sub-path of the LSP that is successfully traversed before the faulty link or node.

- **BFD:** Bi-directional Forwarding Detection (BFD) is a mechanism that is defined for fast fault detection for point-to-point connections. BFD defines a simple packet that may be transmitted over any protocol. The BFD session mechanism requires an additional external mechanism (LSP Ping) to bootstrap and bind the session to a particular LSP or FEC.

- **PW VCCV:** PW VCCV provides end-to-end fault detection and diagnostics for PWs regardless of the underlying tunneling technology. The VCCV switching function provides a control channel associated with each PW. VCCV currently supports the following OAM mechanisms: ICMP Ping, LSP Ping, and BFD. VCCV consists of two components: (1) signaled component to communicate VCCV capabilities as part of the VC label, and (2) switching component to cause the PW payload to be treated as a control packet.

- **OWAMP:** One-Way Active Measurement Protocol, as defined in RFC4656, enables measurement of unidirectional characteristics of IP networks, such as packet loss and one-way delay. For its proper operation OWAMP requires accurate time of day setting at its end points.

- **TWAMP:** Two-Way Active Measurement Protocol, as defined in RFC5357 is a protocol similar to OWAMP that enables measurement of two-way (round trip) characteristics. TWAMP does not require accurate time of day.

- **Tools specified by ITU Recommendation Y.1731:** [Y.1731] specifies a set of OAM procedures and related packet data unit (PDU) formats that meet the transport network requirements for OAM. The PDU and procedures defined in [Y.1731] are described for an Ethernet environment, with the appropriate encapsulation for that environment. However, the actual PDU formats are technology agnostic and could be supported by MPLS-TP nodes just as they are supported by Ethernet nodes. [Y.1731] describes procedures to support the following OAM functions: (1) Connectivity and Continuity Monitoring for pro-active mode end-to-end checking; (2) Loopback functionality to verify connectivity to intermediate nodes in an on-demand mode; (3) Link Trace (activated in an on-demand mode) which provides information on the intermediate nodes of the path being monitored and may be used for fault localization; (4) Alarm Indication Signaling for alarm suppression in case of faults that are detected at the server layer (activated pro-actively). (5) Remote Defect Indication; (6) Locked Signal for alarm suppression; (7) Performance monitoring, which includes measurement of packet delays both uni- and bi-directional (on-demand), measurement of the ratio of lost packets (pro-active), measurement of the effective bandwidth that is supported without packet loss, and throughput measurement.

The quoted document [oam-analysis] also includes a comprehensive analysis of the main MPLS-TP OAM functions supported by the aforementioned IETF and ITU-T tools as well as a discussion on their current limitations and drawbacks. For example, LSP Ping is considered to be computational intensive. In addition, LSP Ping uses the loopback address range to protect against leakage outside the LSP; however, this implies that all of the intermediate nodes support some IP functionality. For what concerns BFD, discriminator values are used to identify the connection at both ends of the path. However, these discriminator values are set by each end-node to be unique only in the context of that node. This limited scope of uniqueness would not identify a misconnection of crossing paths that could assign the same discriminators to the different sessions.

3.1.2 OAM for WSON

The standardization of WSON specifications is still at an early stage, and limited OAM functionality has been discussed so far within the standardization bodies. In [oam-conf-fmk], OAM functions are discussed within the context of GMPLS-based networks, i.e., these functions have been typically designed to operate in an out-of-band control plane, supporting dynamic connection provisioning for any suitable data plane technology.

In terms of requirements, [oam-conf-fmk] specifies that OAM functions have to be activated/deactivated in sync with connection commissioning/decommissioning, avoiding spurious alarms and ensuring consistent operation. In general, it is required that the management plane and control plane mechanisms performing the connection set-up are synchronized with OAM establishment and activation. In particular, if the GMPLS control plane is employed, it is desirable to bind the OAM set-up and configuration to the signaling used for connection establishment, in order to avoid two separate management/configuration steps (connection setup followed by OAM configuration), which increases delay, processing extent, and more important, may be prone to configuration errors. Once OAM entities have been set-up and configured, pro-active as well as on-demand OAM functions can be activated via the management plane. On the other hand, it should be possible to activate/deactivate pro-active OAM functions via the GMPLS control plane, as well.

To this extent, extensions of the RSVP-TE protocol have been proposed in [oam-conf-fmk] to provide a framework for configuration and control of OAM entities, along with the capability to carry technology-specific information. In addition, [oam-conf-fmk] proposes an option for bootstrapping OAM in environments where RSVP-TE signaling is already in use to set up the LSPs that have to be monitored using OAM.

In terms of OAM tools and devices specifically designed for WSON, preliminary solutions performing impairment detection have been released to the market, to enable rapid fault detection and performance monitoring. These devices provide real-time and in-band monitoring of parameters that are critical for channel performance, including optical signal to noise ratio (OSNR) and physical impairments like polarization mode dispersion (PMD) and differential group delay (DGD). Additional features of these devices include bit-rate independency and compatibility with both OOK and PSK modulation schemes.

To effectively implement OAM in a WSON, additional management tools and solutions are required to retrieve and elaborate the WSON information obtained by impairment detectors and monitoring devices. Protocols, algorithms and mechanisms are under investigations to fulfill this gap.

Within the European DICONET project, a comprehensive analysis of the available literature in this field has been provided together with some preliminary solutions to address the issues related to accurate impairment measurement and estimation, fault localization and performance monitoring.

Some references relevant to these topics, are: [Tsuritani08], [Chung08], [Lee09], [Stanic10], [Wu10], [Sambo09], [Pinart09].

3.2 Definition of OAM requirements

3.2.1 OAM of multi-domain and multi-carrier transport networks

This subsection deals with multi-domain and multi-carrier issues in the existing standards (the multi-technology case has not been seriously considered until now by standardization bodies). The goal is to identify gaps and to propose new requirements to fill these gaps.

In many cases network services traverse several domains, and in long distance services this is the most probable case. Different administrative domains can be created and operated by a single carrier in order to simplify the network design and to increase the network scalability (multi-domain scenario); in other cases different administrative domains are operated by different carriers (multi-carrier scenario).

The multi-domain and multi-carrier scenarios pose special technical and commercial issues that should be defined and addressed.

In particular, OAM in multi-carrier networks has commercial aspects that do not exist in single carrier networks. Indeed, in case of failure or out-of-SLA service delivery, the violative carrier should compensate its partner carriers or the end customer. Based on the information made available by the OAM tools, the carriers should agree on the root cause.

Unfortunately, at present no reliable means to carry out this OAM based compensation procedure are available in existing standards.

Furthermore, the out-of-service duration is a significant factor when calculating the compensation/penalty in case of failure. Yet, currently, each service provider measures the out-of-service duration independently; as a result, it is difficult to agree on the out-of-service duration and, as a consequence, on the amount of compensation.

The existing standards for OAM in transport networks do not cope with the above mentioned problems; therefore, in a multi-carrier environment, the following additional requirements must be considered:

- The OAM system should support MEs that are handled by different administrative domains; a network service can be provided by several domains that are operated by different carriers (multi-carrier scenario).

- The OAM system should provide reliable means to the service providers to prove, in case of failure, which is the failed segment.

- The OAM system should provide reliable means to measure out-of-service duration; such measurement should be accepted by all parties.

3.2.2 OAM of multi-region transport networks: MPLS-TP / WSON interworking

Two main scenarios can be identified for MPLS-TP and WSON OAM interworking. The first scenario refers to vertical interworking, where a multi-layer network is considered, having the WSON technology as the lower layer, and MPLS-TP as the upper one. The second scenario refers to horizontal interworking, where a single-layer network is considered, having at least two adjacent domains of WSON and MPLS-TP technologies.

So far, in both standardization documents and scientific literature, no specific studies have been provided in terms of OAM functions for both scenarios.

According to the specific OAM function, OAM tools originally defined to operate in a specific technology layer might be applied to both scenarios and technologies; they might coexist or they might be suppressed in favor of other solutions.

The STRONGEST project will provide guidelines and solutions to enable, within both scenarios, effective end-to-end OAM solutions, taking into account the presence of technology-specific OAM solutions (particularly in the case of WSON), as well as technology-independent mechanisms.

3.3 STRONGEST OAM framework

STRONGEST is deeply involved in studying OAM aspects, as required by the contractual documents, where this theme is clearly quoted among the Project goals: "...to pursue end-to-end services delivery crossing domains that are heterogeneous in terms of technologies (circuit transport networks and connection-oriented packet transport networks), control plane models (e.g. multi-layer/multi-region), OAM mechanisms, vendors and operators."

The Project is pursuing particularly QoS monitoring of network services in a complex network scenario, fast fault notice for triggering recovery mechanisms and individuation of degrades for accurate maintenance

The following sections describe the framework of this activity on OAM. In particular, the scope of the work is detailed and the specific terminology adopted inside the Project is established.

3.3.1 OAM activity scope

One of the main Project goals is the design of a network architecture guaranteeing end-to-end performance and survivability. In fact, efficient OAM mechanisms play an essential role in STRONGEST architectures, in particular for:

- Monitoring QoS for offered network services, also in a complex (multi-technology, multi-operator, multi-domain) network scenarios
- Fast fault notice for triggering recovery mechanisms (protection, restoration, …)
- Individuation of degrades for accurate maintenance

A thorough analysis of OAM issues is essential for the Project, in order to share:

- A general scenario for OAM (compatible with the reference scenarios outlined for the control plane studies)
- Terminology
- State of the art and progress in standard bodies

Figure 4 depicts the STRONGEST general network architecture for the medium term, focusing on the basic OAM structure.

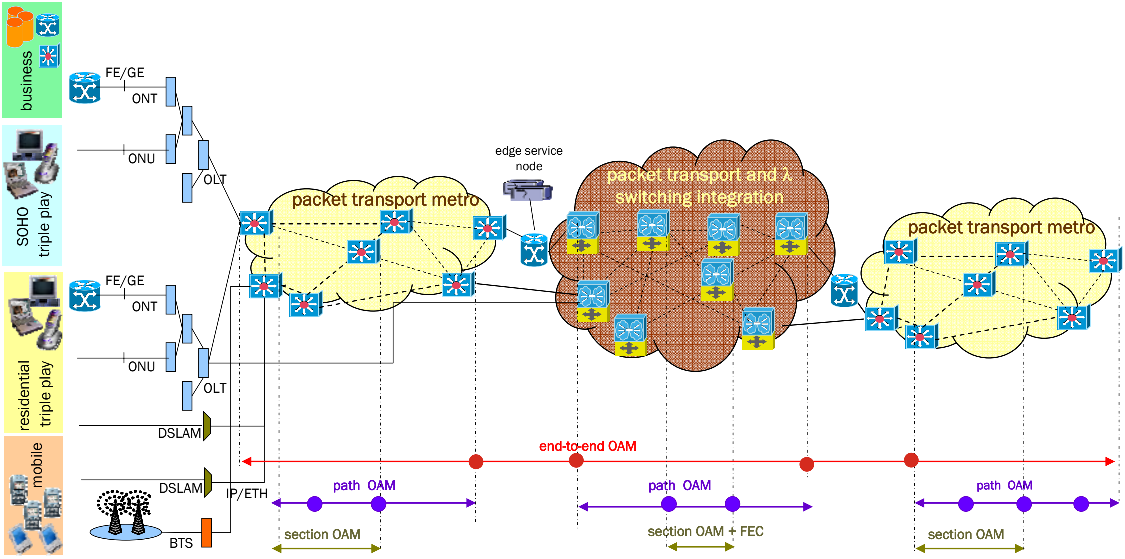


Figure 4 – STRONGEST general architecture, with focus on OAM

In the overall architecture, in particular, it is possible to identify metro-regional and core-backbone segments. Independently of the adopted scenarios, in a real network it is necessary to perform:

- End-to-end OAM
- Path OAM
- Section OAM

The three main scenarios, defined in Chapter 2:

- Single-Domain / Multi-Region / Single-Carrier
- Multi-Domain / Multi-Region / Single-Carrier
- Multi-Domain / Multi-Region / Multi-Carrier

shall be considered.

It is worth noting that, until now, very few attention has been paid to the multi-region configuration while the multi-domain and multi-carrier cases are already being considered by standardization bodies. STRONGEST will consider and study the main scenarios in details.

3.3.2 OAM terminology

STRONGEST inherits the general terminology, regarding OAM basic elements and functions, adopted by ITU-T in Y.1737/G.8114 [G8114] (consented but not approved by ITU-T), Y.1711 [Y1711] and Y.1731 [Y1731].

The following list reports the most important terms and definitions, modified to be suitable for the general STRONGEST scenarios. These definitions might be related to point-to-point and point-to-multipoint connection-oriented packet network services.

- Maintenance Entity (ME) – it represents an entity that requires management, and is a relationship between two Maintenance Entity Group End Points.
- ME Group (MEG) – it includes different MEs that satisfy the following conditions:
    - MEs in a MEG exist in the same administrative boundary,
    - MEs in a MEG operate at the same MEG Level (see below), and
    - MEs in a MEG belong to the same point-to-point connection or to the same point-to-multipoint connection1.
- MEG End Point (MEP) – it marks the end point of a MEG that is capable of initiating and terminating OAM packets for fault management and performance monitoring.
- Server MEP – it represents the compound function of the Server layer termination function and Server/(upper layer) adaptation function which is used to notify the upper layer MEPs upon failure detection by the Server layer termination function or Server/(upper layer) adaptation function, where the Server layer termination function is expected to run OAM mechanisms specific to the Server layer.
- MEG Intermediate Point (MIP) – it is an intermediate point in a MEG that is capable of reacting to some OAM packets. A MIP does not initiate OAM packets. A MIP takes no action on the data-plane connection.
- MEG Level (MEL) – in case MEGs are nested, the OAM packets of each MEG have to be clearly identifiable and separable from the OAM packets of the other MEGs[2].

[1] For point-to-point connections, a MEG usually contains a single ME. For point-to-multipoint connections containing n end-points, a MEG contains (n-1) MEs.

[2] Latest achievements inside ITU-T and IETF (July 2010) do not consider the possibility of having MEG levels that implement Tandem Connection Monitoring (TCM) adopting nested Label Switch Paths.

- OAM transparency – this term refers to the ability to allow transparent carrying of OAM packets belonging to higher level MEGs across other lower level MEGs when the MEGs are nested.

- In-service OAM – it refers to OAM actions which are carried out while the data traffic is not interrupted with an expectation that data traffic remains transparent to OAM actions.

- On-demand OAM – it refers to OAM actions which are initiated via manual intervention for a limited time to carry out diagnostics.

- Proactive OAM – it refers to OAM actions which are carried out continuously to permit proactive reporting of fault and/or performance results.

- Out-of-service OAM - it refers to OAM actions which are carried out while the data traffic is interrupted.

3.3.3 Future work

OAM is a crucial point in telecommunications network. The use of OAM is dictated by:

- The necessity of fault localization, to provide efficient and fast resilience mechanisms and to trigger maintenance operations.

- The need to control the observance of Service Level Agreements.

In a multi-domain/multi-region/multi-layer network scenario, end-to-end OAM is a really important task, not yet fully developed by research projects and standardization bodies. This importance is due to the need to provide an end-to-end network service (i.e. connections) crossing several domains and using 2 (or more) switching layer technologies. One of the most important STRONGEST challenges is to provide a network capable of controlling and guaranteeing the quality of end-to-end services.

The standardization actions should limit, at least, the problem of lack of compatibility among different vendors. In any case, different implementations of the same standard often result in some incompatibilities, due to inadequate or superficial standardization, inevitably resulting in different "flavors" of the same technology. These problems are more frequent and greater for the new technologies, like those addressed by STRONGEST.

The compatibility of OAM among different technologies is an essential task to guarantee the end-to-end monitoring and, unfortunately, it has not been treated in depth by standardization bodies, until now, also because different technologies are often developed and standardized under the control of different bodies (IEEE, ITU-T, IETF).

STRONGEST, regarding OAM, will provide studies about:

- OAM requirements.

  - What it is important to monitor?
  - Which is the frequency of monitoring?
- Quantification of OAM parameters and, as consequence, understanding of the correct trade-off between OAM efficiency and functions to be monitored.
- Application of OAM concepts for STRONGEST scenarios:
  - What to monitor in case of multi-domain single carrier (cf. Section 2.2) and multi-domain multi carrier (cf. Section 2.3)?
  - To monitor the e2e connection.
  - To understand which information can be exchanged among different domain and among different Carriers.
- Application of OAM concepts for two technological domains:
  - OAM for packet connection oriented (i.e. packet transport) technologies are considered.
  - Circuit and optical domains, adopting OTN and photonic switched networks. The issues to consider are:
    - OAM for OTN (already and completely defined in ITU-T)
    - OAM for photonic switching and in particular
      - What it is needed? (only fault detection or also some performance evaluations?)
      - What is necessary at the boundary between OTN or photonic and PTN, regarding OAM?
    - In the longer-term scenario, sub-wavelength switching will be considered. STRONGEST will define also OAM requirements and define some solutions; for this activity STRONGEST will feed discussion in standardization bodies.
    - Performance parameters and functions for multipoint connectivity

4 Control plane reference architecture

4.1 State of the art

Protocols

With the growing size of networks and the dynamics of Label Switched Path (LSP) establishments and tear downs, automatic provisioning coupled with traffic engineering (TE) has become essential to operate a network cost-efficiently. To increase capacity and improve traffic engineering, Core networks controlled by Generalized Multi-Protocol Label Switching (GMPLS) were introduced [RFC3945].

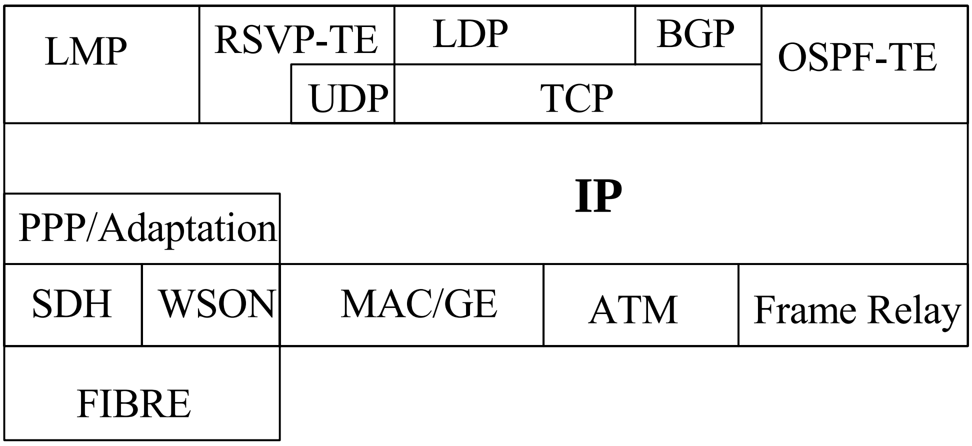


Figure 5 – GMPLS protocol stack

The basis of GMPLS involves enhancements to MPLS-TE protocols: Resource Reservation Protocol (RSVP-TE) signaling for setting up End-to-End (E2E), quality-enabled connections; Open Shortest Path First (OSPF-TE) routing for automatic topology and network resource dissemination, together with the Link Management Protocol (LMP) for link discovery and verification. The relevant protocol stack is shown in Figure 5 above.

The cost savings enabled by Control Plane (CP) technologies have encouraged service providers to plan for all-optical islands using nodes running GMPLS to provide control and data management. Accordingly, the GMPLS protocol suite has been enriched by new functionalities equivalent to those defined for packet-based MPLS, with key components of reachability and TE information, path computation and LSP signaling.

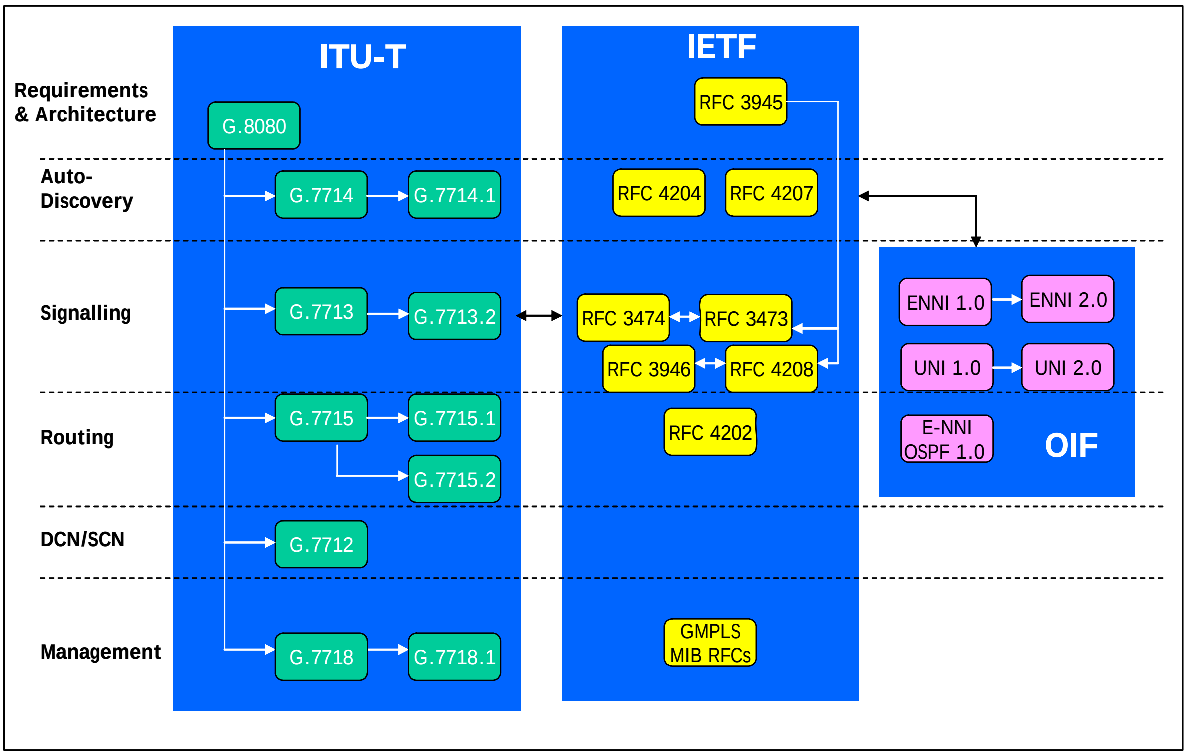


Figure 6 – Protocol standards

Standardization Bodies

Of the various standards development organizations, the Internet Engineering Task Force (IETF) has been concerned with optical routing the longest. The IETF has specified CP standards for a variety of data planes and switching types. The IETF also coordinates work on data plane specific CPs between IETF and other standards organizations. Multi-domain considerations have motivated several activities within the IETF, the aim being to standardize a model to distribute computation of TE LSPs among different areas or within a small group of domains.

The International Telecommunication Union (ITU-T) has simultaneously been working on the architecture, functional models and protocol specifics of the Automatically Switched Optical Network (ASON), a client-server architecture with well-defined interfaces that allows clients to request services from the optical network (server). The ITU-T and IETF recently agreed on an (optical) routing protocol that should support:

- Partitioning of transport networks that may be controlled by using multiple instances of different routing protocols at different levels
- Separation of control and data (i.e. transport) planes such that data and control entities are not required to be co-located and the scope of the data plane resources seen by a routing function is not limited

- Distribution, abstraction and filtering of information within hierarchical and partitioning relationships

- Independence of inter-domain and intra-domain routing protocols

The OIF is a consortium of optical networking vendors and service providers (carriers) whose goal is to expeditiously develop interoperability implementation agreements. Regarding the CP, the OIF has mostly followed the functional models developed by the ITU-T in identifying control interfaces. As per these models, the OIF work has focused on the interface between the user and the optical networks (UNI) and the interface between optical control domains within a single carrier network (NNI).

While the ITU has worked on the requirements and architecture of ASON based on the requirements of its members, it is explicitly aiming to avoid the development of new protocols whenever existing ones will work fine. The IETF, on the other hand, has been tasked with the development of new protocols in response to general industry needs. The relationship between protocols developed by the various standards organizations is shown in Figure 6.

4.1.1 Standardization and reference standards

This section discusses recent work on protocols particularly relevant to CP needs, primarily regarding network signaling functionality. In this context, “recent” normally refers to activity subsequent to the comprehensive review in [NOB2-D44], as shown by latest published RFCs.

The main enhancements introduced by GMPLS in provisioning functionalities may be summarized as:

- Minimization of setup delay in optical networks; GMPLS permits an upstream node to suggest a label (e.g. a wavelength in a WSON) to a downstream node to start reserving and configuring its hardware with the proposed wavelength from source to destination.

- Support for different switching technologies and capabilities.

- Separation of control and data plane.

- Support for a GMPLS hierarchy, nesting LSPs based on the switching technology.

- Support for bidirectional connection requests; MPLS LSPs are unidirectional; to establish a bidirectional connection, two unidirectional LSPs in opposite directions must be established independently.

Standardization efforts have also been made to inter-connect ASON and GMPLS CPs. A paper of Okamoto et al. [OKA1] and the IST NOBEL phase 2 deliverable D14 [NOB2-D14] describe the challenges in doing so, to allow automatic E2E path setup. They propose the use of an interworking function, which can translate the signaling at the border nodes to facilitate interworking between the GMPLS and ASON frameworks, supporting different inter-domain connection request scenarios. A field trial of the proposed function is also presented in their work.

With regard to open issues, single domain work has largely involved relatively simple extensions to support TE:

- [RFC5561] recognizes that a number of enhancements to the Label Distribution Protocol (LDP) have been proposed. Some have been implemented and others are advancing toward standardization, while it is likely that additional enhancements will be proposed in the future. The document defines a mechanism for advertising LDP enhancements at session initialization time, as well as a mechanism to enable and disable enhancements after LDP session establishment.

- [RFC5420] defines a new object for RSVP-TE messages that allows signaling of further attribute bits and also carriage of arbitrary attribute parameters to make RSVP-TE easily extensible to support new requirements. Additionally, this document defines a way to record the attributes applied to the LSP on a hop-by-hop basis. Mechanisms defined in this document are equally applicable to GMPLS Packet Switch Capable (PSC) LSPs and to GMPLS non-PSC LSPs.

- [RFC5710] describes how RSVP PathErr messages may be used to trigger rerouting of MPLS and GMPLS point-to-point TE LSPs without first removing LSP state or resources. Such LSP rerouting may be desirable in a number of cases (e.g. soft pre-emption and graceful shutdown). It relies on mechanisms already defined as part of RSVP-TE and simply describes a sequence of actions to be executed. While existing protocol definitions can be used to support reroute applications, this document also defines a new reroute-specific error code to allow for future definition of reroute-application-specific error values.

In the context of the STRONGEST Project, the end-to-end control plane should manage heterogeneous networks in terms of: different domains, different technologies and also different carriers. In order to face all the related issues, all the methodologies and solutions that will be found and proposed within STRONGEST should take as a starting point the procedures and techniques defined by the main standardization bodies.

To support MPLS Inter-domain TE, IETF has defined procedures and extensions to (i) signaling protocols, (ii) routing protocols and (iii) the PCE Architecture.

Three different signaling methods have been identified for the setup and maintenance of TE-LSPs that span multiple domains [RFC4726]. They include:

- Contiguous – A single contiguous LSP is established from ingress to egress in a single signaling exchange, using the procedures of [RFC3209] and [RFC3473]. No further LSPs are required to be established to support this LSP so that hierarchical or stitched LSPs are not needed.

- Hierarchical (Nesting) – According to the procedures described in [RFC4206], a hierarchical LSP may provide a TE LSP tunnel to transport (i.e., nest) multiple TE

LSPs along a common part of their paths. Alternatively, a TE LSP may carry (i.e., nest) a single LSP in a one-to-one mapping.

- Stitching – According to the procedures described in [RFC5150], separate LSPs (referred to as a TE LSP segments) are established and are "stitched" together in the data plane so that a single end-to-end Label Switched Path is achieved. Unlike the hierarchical technique, the component LSP segments are signaled as distinct TE LSPs, with different source and destination in the control plane.

As stated in [RFC5151], an end-to-end inter-domain TE LSP may be achieved using any combination of the above described signaling techniques. That is, the options should be considered as per-domain transit options. The choice is a matter of policy for the node requesting LSP setup (the ingress node) and policy for each successive domain border node.

In terms of routing protocol enhancements, extensions have been proposed to advertise inter-domain TE information within a domain/Area (Inter-AS-TE-v2 LSA should have Area flooding scope). However, no support for flooding information from within one domain to another one is proposed. As stated in [RFC5392], when TE is enabled on an inter-domain link and the link is up, the ASBR should advertise this link using the normal procedures for OSPF-TE [RFC3630]. The information advertised comes from the ASBR's knowledge of the TE capabilities of the link, the ASBR's knowledge of the current status and usage of the link, and configuration at the ASBR of the remote domain number and remote ASBR TE Router ID.

In order to perform Inter-domain TE, IETF currently refer to the PCE Architecture. Two methods are proposed. The first method is called Per-Domain Path Computation [RFC5152]. It applies where the full path of an inter-domain TE LSP cannot be or is not determined at the ingress node of the TE LSP, and is not signaled across domain boundaries. When a boundary node LSR (e.g., ASBR) receives a Path message with an ERO that contains a loose hop, then it performs ERO expansions (with or without external PCE cooperation). Per-domain path computation can be used regardless of the nature of the inter-domain TE LSP (contiguous, stitched, or nested). However, optimality is not guaranteed. The second method is called Backward Recursive PCE-based Computation (BRPC) [RFC5441]. BRPC specifies a procedure relying on the use of multiple PCEs to compute inter-domain shortest constrained paths across a predetermined sequence of domains, using a backward recursive path computation technique. BRPC provides the optimal path computation. In addition, the BRPC specification claims that this technique, combined with the use of Path Key ([RFC5553]) guarantees the confidentiality of TE information across different domains.

With the purpose of Inter-domain TE (intra-carrier), OIF proposed the OIF E-NNI routing specifications, supporting a hierarchy of routing instances where each routing layer operates independently. Information among routing layers are exchanged by feeding up/down the adjacent level (different from the relationship between Areas in IGPs). OIF E-NNI routing allows the advertisement of intra-domain virtual links, inter-domain links and topology abstraction information (full mesh of virtual links between border nodes vs. abstract node). It takes no position on the detailed way to represent Virtual/Abstract TE information (e.g., how to deal with link attributes in case of multiple protection schemes).

Heterogeneous networks, such as the ones considered within the STRONGEST Project, are divided not only in terms of domains, but also in terms of regions and layers, that can be controlled by the GMPLS protocol suite. According to IETF [RFC4202] and [RFC4206], the information carried in the Interface Switching Capabilities (ISCD) is used to construct LSP regions and to determine regions' boundaries, that can belong to different network layers (e.g. Ethernet over SDH) or not (e.g. SDH hierarchy). A set of IETF RFCs and drafts are currently under deployment in order to provide extensions to GMPLS protocols for Multi-Region Networks (MRN) and Multi-Layer Networks (MLN).

In the context of the STRONGEST Project, an integration of GMPLS protocol suite and the hierarchical view of both PCE and RACS architectures will be investigated, exploiting the advantages of the different approaches. Particularly a possible integration between PCE and RACS functional components will be proposed, as well as an under-deployment summarization method that considers several general and administrative parameters. PCEP possible extensions are also under study in order to fulfill the requirements of the hybrid architectures under study in the context of reference scenarios defined in Chapter 2.

TISPAN and 3GPP

Both TISPAN (Telecommunications and Internet converged Services and Protocols for Advanced Networking) and 3GPP (3rd Generation Partnership Project) specify a control layer architecture.

TISPAN has recently published the release 3 of the Functional Architecture of the Resource and Admission Control Sub-System (RACS) [RACS], the NGN Subsystem responsible for elements of policy control, resource reservation and admission control. More details on RACS will be provided in Section 4.1.3.

3GPP specifies a control layer architecture known as PCC (Policy and Charging Control) here referred to as PCCA [PCCA]. PCCA includes the PCRF (Policy and Charging Rules Function) and the PCEF (Policy and Charging Enforcement Function) functional components. The PCRF, which plays for mobile networks a similar role of RACS for fixed networks, is a functional element that encompasses policy control decision and flow based charging control functionalities. The PCEF is the functional element that encompasses policy enforcement and flow based charging functionalities. This functional entity is located at the Gateway (e.g. GGSN in the GPRS case).

Lately, in the more general context of Fixed Mobile Convergence (FMC), different fora including 3GPP, TISPAN and BFF (Broadband Forum) are jointly working on a possible unification of End-to-End QoS and Policy Control architecture.

In the context of the STRONGEST Project the RACS framework will be mostly investigated because it best suits to fixed networks also offering a broader set of functionalities, e.g. admission control, (currently) missing in the 3GPP PCRF framework. However, the aim is also to look with attention at possible developments of a unified control layer architecture for both fixed and mobile networks.

4.1.2 PCE architecture

This section highlights recent work on Path Computation Element (PCE)-related protocols. Once again, the focus is on latest published RFCs, with reference to [NOB2-D44]. In general, inter-domain path computation, or the ability to compute optimal E2E paths across multiple domains, is the next step toward wide deployment of a distributed CP with support for traffic engineering. A key enabler to achieve this goal is the introduction of a dedicated PCE in each domain, which handles the topology information and is polled by various nodes to determine the path from a source to a destination.

Network state information is gathered into a database typically fed by intra-domain routing protocols (e.g. OSPF-TE) in combination with BGP routing information. The PCE uses domain hop information from topology information dissemination mechanisms to calculate optimal E2E inter-domain paths. The use of a path computation element eliminates the need for every node within the network to compute the path and all link state information is sent only to the relevant PCE. A single domain can have multiple PCEs to facilitate load sharing and avoid any single point of failure.

Studies suggest that a centralized PCE approach is preferable for intra-domain path calculations, as this approach computes optimal paths with no additional delay due to inter-PCE communication. However, the recommended PCE approach for multi-domain path computation depends on the required features (e.g. topology hiding vs. optimal path computation) and already deployed strategies inside different domains. PCE-based routing architectures for multi-domain networks can be classified [CHA] into two major groups:

- Peer-to-peer
- Hierarchical

In a peer-to-peer model, PCEs of neighboring domains create peering relationships and interact with each other to exchange routing information via mechanisms of the CP. To establish a multi-domain route, PCEs are probed sequentially to determine availability of the path. If available, the return message of the signal is the next hop information between different domains [GOM]. PCE peering mechanisms can also be used [BAL] to set up inter-domain paths supporting Optical Burst Switching networks.

4.1.2.1 Routing

Recent work in this area is almost exclusively focused on mechanisms to enable and enhance PCE routing, in terms of both information needs and quality (optimality) of the calculated paths.

Information needs may be exemplified by:

- [RFC5088] relates to circumstances where it is highly desirable for a Path Computation Client to be able to dynamically and automatically discover a set of PCEs, along with information that can be used for PCE selection. The document defines extensions to the OSPF routing protocol for the advertisement of PCE

Discovery information within an OSPF area or within the entire OSPF routing domain.

- [RFC5392] considers extensions to OSPF version 2 and 3 protocols to support MPLS/GMPLS TE for multiple ASes. Specifically, the focus is on flooding of TE information about inter-AS links that can be used to perform inter-AS TE path computation (i.e. no support for flooding information from within one AS to another AS is proposed).

- [RFC5440] specifies the PCE Communication Protocol (PCEP) for communications between a Path Computation Client and a PCE, or between two PCEs. Such interactions include path computation requests and path computation replies as well as notifications of specific states related to the use of a PCE in the context of MPLS/GMPLS TE. PCEP is designed to be flexible and extensible so as to easily allow for the addition of further messages and objects, should further requirements be expressed in the future.

Additional optimization-related RFCs reflect any constraints on paths, computational efficiency and possible differences between optimal E2E paths and those calculated step-by-step:

- [RFC5521] presents PCEP extensions for “route exclusions”, representing constraints on the path computation, nodes, resources and Shared Risk Link Groups that are to be explicitly excluded from the computed route. These exclusions form part of the route request submitted by the Path Computation Client to the PCE for a route.

- [RFC5557] provides PCEP extensions to support bulk Global Concurrent Optimization (GCO). When computing the routes for a set of TE LSPs through a network, it may be advantageous to perform bulk path computations to avoid blocking problems and achieve more optimal network-wide solutions. A GCO simultaneously considers the entire topology of the network and the complete set of existing TE LSPs with their respective constraints, looking to optimize the network for all TE LSPs. The GCO application is primarily a Network Management System solution.

- [RFC5441] proposes the Backward Recursive Path Computation (BRPC) protocol to compute optimal inter-domain paths in a multi-domain network. BRPC assumes that an inter-domain sequence is provided from source to destination before the actual path computation. Using this sequence, the PCE protocol is used to contact contiguous domains in a sequential manner. On reaching the PCE of the destination domain, the destination PCE returns a tree of possible paths to the destination from all possible ingress nodes. Each PCE extends this tree to the ingress nodes inside the respective domains and sends it to the previous PCE. On reaching the source PCE, the optimal E2E path from the received set of paths is selected for inter-domain path setup. It should be noted that only optimal paths to the ingress border nodes are considered in the tree, ensuring that the proposed mechanism is scalable.

4.1.2.2 Signalling

Typical multi-domain RFCs relating to TE and resilience may be summarized as:

- [RFC5150] considers scenarios where there may be a need to combine several GMPLS LSPs such that a single E2E LSP is realized and all traffic from one constituent LSP is switched onto the next LSP. This is termed "LSP stitching", the key requirement being that a constituent LSP not be allocated to more than one E2E LSP. Constituent LSPs are then referred to as "LSP Segments" (S-LSPs). The document describes extensions to the existing RSVP-TE signaling protocol to establish e2e LSPs from S-LSPs and describes how the LSPs can be managed using GMPLS signaling and routing protocols.

- [RFC5151] describes procedures and protocol extensions for the use of RSVP-TE signaling in MPLS/GMPLS-TE packet networks to support establishment and maintenance of LSPs that cross domain boundaries. A domain is considered to be any collection of network elements within a common realm of address space or path computation responsibility (e.g. Autonomous Systems, Interior Gateway Protocol (IGP) routing areas and GMPLS overlay networks).

- [RFC5298] proposes different methods to set up recovery LSPs for increased reliability, using per domain path computation and establishing traffic engineered LSPs across multiple GMPLS domains. Path computation can take place at the ingress Label Switched Router (LSR) for an E2E LSP, or a per-domain LSP or at a separate PCE in the domain [RFC5152]. LSP-signaling can be performed using contiguous E2E LSPs, domain-wise tunneling with nested LSPs [RFC4206] or segment-based LSPs [RFC5150].

Finally, work is in progress on multi-domain TE limitations due to both lack of visibility across domain boundaries and confidentiality concerns:

- [RFC5152] specifies a per-domain path computation technique for establishing inter-domain TE MPLS/GMPLS LSPs (where a domain refers to a collection of network elements within a common sphere of address management or path computational responsibility such as ASes. Per-domain computation applies where the full path of an inter-domain TE LSP is not determined at the ingress node and is not signaled across domain boundaries. This is most likely to arise owing to TE visibility limitations. The signaling message indicates the destination and nodes up to the next domain boundary. It may also indicate further domain boundaries or domain identifiers. The path through each domain, possibly including the choice of exit point from the domain, must be determined within the domain.

- [RFC5553] proposes a path-key mechanism to hide the topology inside domains. Each path computed by a PCE for an inter-domain path setup is mapped to a key. The key is sent in the PCE message and stored by RSVP during path setup. At the ingress border node of each domain, the path key is sent to the PCE to determine the path inside the domain, maintaining topology confidentiality.

4.1.2.3 Open issues review

The PCE approach helps scalability but computing by segments means the resulting paths are likely to be far from optimal. Finding high-quality paths remains an open issue. In particular, each segment of an inter-domain LSP is derived from very limited visibility of the state and topology of the network. An alternative approach is to work towards a TE information exchange model between domains, perhaps involving a small group of neighboring domains exchanging highly aggregated state and topology information (to respect confidentiality of ISP networks). One such option is cooperative path computation [TOR], a scheme where PCEs exchange path information in the context of a specific E2E path computation instance, often prior to signaling the path. Issues such as how TE information is to be distributed and updated then need to be carefully investigated.

Use of PCE opens up the possibility of pre-computation, generating solutions a priori for a large set of possible parameters. These parameters are then used during path computation on arrival of a request, increasing response time and scalability and reducing computational load on the PCE. A pre-computation scheme has been proposed [ORD] including both additive QoS metrics (delay) as well as bottleneck weight metrics (bandwidth). However, a more detailed analysis highlighting parameters such as pre-computation overhead, frequency of computation cycles and signaling load is required to analyze the benefit of such ideas.

For GMPLS-based VPNs, one of the main goals is to exploit the GMPLS multi-layer control applicability and hierarchy, allowing nesting of LSPs in the context of VPN services over coarse granular switching of large traffic streams in the core. Proposals have been made to extend the PCE framework to facilitate multi-layer routing and a few basic schemes are suggested in [RFC5623]. Various metrics are used to compare [DAS] performance of the PCE framework against per-domain signaling with crankback (where, if a path setup request fails at a given node, RSVP asks the upstream node to use an alternate path). Results show that the PCE framework can set up LSPs with lower costs compared to the signaling framework, while also reducing the possibility of failure during path computation.

4.1.2.4 Architecture specifics

The Path Computation Element (PCE) Architecture [RFC4655] has been introduced to provide effective Traffic Engineering solutions, i.e., to efficiently cope with complex constraint-based path computations. The main motivations that drove the introduction of the PCE Architecture included the need to perform CPU-intensive path computations and to deal with several scenarios where the node responsible for path computation has limited visibility of the network topology and resources (e.g., multi-domain and multi-layer networks).

The standardization process of the PCE Architecture began in 2005 and it has now reached a high level of stability. This section reports a brief overview of the main definitions, components and features of the PCE Architecture as defined by the IETF PCE Working Group. The detailed list of RFC and DRAFT can be found in [PCE-WG].

The PCE Architecture relies on two functional components: the PCE and the Path Computation Client (PCC). The PCE is defined as an entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints. The PCC is defined as any client application requesting a path computation to be performed by a PCE.

The PCE, possibly implemented on a dedicated server, is responsible for performing constraint-based path computation requested by a PCC, which is typically implemented on a Network Management System (NMS) or a network node. Particularly in the case of inter-domain scenarios, a PCE may also behave as PCC requesting path computations to a different PCE. Communication between PCC and PCE is guaranteed by the PCE Communication Protocol (PCEP).

To perform path computations, PCC and PCE first open a PCEP session within a TCP session. A path computation request is then included within a PCReq message specifying all the requested parameters and constraints. Reply (i.e., PCRep message) is provided by the PCE specifying either the positive result (e.g., explicit path route) or negative result (no path found). Additional messages are also defined to close the PCEP session, to handle specific events and communication errors (e.g., Error (PCErr) and Notification (PCNtf) messages) and to verify the operative status (e.g., liveness, overload) of a remote PCE.

To perform inter-domain path computations, the PCE Architecture defines two main procedures: the PCE-based Per-Domain (PPD) and the Backward Recursive PCE-based Computation (BRPC). They both exploit a backward recursive technique. The path computation request is first forwarded between PCEs, domain-by-domain, until the PCE responsible for the domain containing the destination node is reached. The PCE in the destination domain then computes either a single sub-path (as in PPD) or a tree of virtual sub-paths (as in BRPC) to the destination. The result is passed back to the previous PCE which in turn expands the sub-path(s) and passes the result back until the source domain PCE completes the entire path computation. In the case of BRPC, the source PCE also selects the shortest path among those included in the final tree. In multi-carrier scenarios, PCE and PCC will not exchange strict explicit list of traversed intra-domain hops and paths will be expressed in the form of an encrypted key.

The main path computation parameters defined in RFC4655 are summarized, together with the related PCEP objects, in Table 1. These objects allow one to specify the main path computation parameters including end-points (source and destination), connection bi-directionality and requested bandwidth. Other important parameters include: diversity (link, node and/or Shared Risk Link Group (SRLG) disjointness), the need for local protection (i.e., Fast ReRoute), and application of the BRPC procedure. In addition, PCEP specifications allow to provide information about failure in the path computation (i.e., NO-PATH information), to specify strict/loose sequences of hops to traverse or avoid, computed metric values, priority in the path computation, and information to perform re-optimization.

Additional PCEP extensions have been also introduced to request specific objective functions, point-to-multipoint path computations, global re-optimization, DiffServ-aware parameters, and policy-based path computations.

Still at the stage of requirements, PCEP extensions have been proposed in the context of inter-area, inter-Autonomous-System and inter-layer networks. For example, inter-area PCEP extensions will allow a PCC to require path computations indicating whether or not area crossing is permitted and to know whether the returned path is an inter-area path. Additional PCEP extensions will have to permit area inclusion/exclusion and inter-area diverse path computation.

The PCE Architecture encompasses discovery mechanisms to enable PCCs to dynamically and automatically discover the presence of PCEs. At this purpose, protocol extensions to Interior Gateway Protocols have been defined for the advertisement of PCE discovery information within a routing area or within the entire routing domain.

Table 1 – Main PCEP parameters and objects

PCEP Object	Parameter
Requested Parameter (RP)	Computation priority
	Re-optimization
	Bi-directionality
	Strict/Loose
	VSPT Flag (BRPC)
	Request ID
NO PATH	Nature of Issue
	Options
End Points	Source
	Destination
Bandwidth	Bandwidth
Metric	Type
	Bound
	Computed Metric
	Metric value
Explicit Route Object	Explicit Route Object
Reported Route Object	Reported Route Object
LSPA Object	Exclude/Include
	Setup/Holding priority
	Local Protection
Include Route Object	Include Route Object
SVEC	L (Link diverse)
	N (Node diverse)
	S (SRLG diverse)
	Request ID Numbers

The applicability of the PCE Architecture in the context of WSON networks is still in an early stage of discussion. So far, extensions to the PCEP protocol have been proposed within the PCE WG only to require routing and wavelength assignment either as a combined or separated process.

4.1.3 TISPAN RACS architecture

As it will be deeply investigated in Section 4.3.1, the proposed control layer/control plane architecture will be based on the integration and harmonization of RACS / PCE and GMPLS control plane functional components.

Purpose of the section is to briefly summarize main functionalities of release 3 of RACS as specified by TISPAN. RACS (whose reference architecture is depicted in Figure 7) is the NGN Subsystem [NGN] responsible for elements of policy control, resource reservation and admission control. RACS provides policy based transport control services to applications. This enables the request and reservation of transport resources from access and core transport networks, and the points of interconnection between them, in order to support e2e QoS.

Moreover, by hiding the interaction between applications and transport resources, RACS also ensures that applications do not need to be aware of the underlying transport networks.

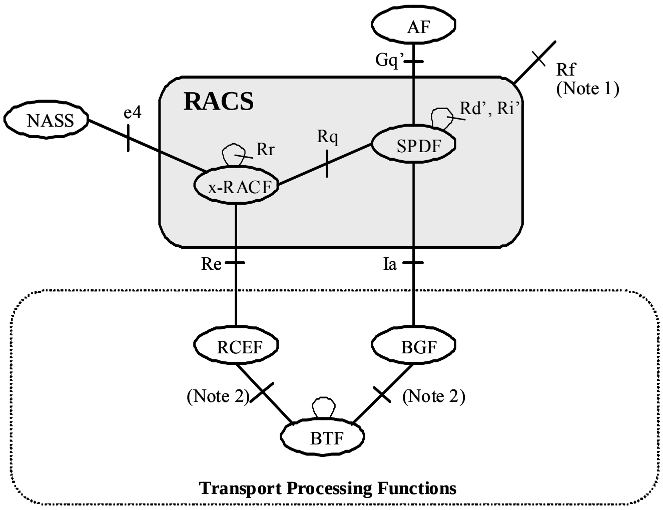


Figure 7 – ETSI TISPAN RACS functional architecture (Release 3)

RACS is built around two main functional entities: the service Policy Decision Function (SPDF) and the Generic Resource Admission Control Function (x-RACF). Main functions of SPDF are the following:

- it represents the single contact point for the applications (through Gq' reference point in Figure 7);
- it provides network topology and technologies abstraction to the applications;
- it is a first policy decision function, based on local services and operator policies;
- it coordinates and interconnects several x-RACFs (through Rq reference point);
- it can operate in a multi-domain and multi-carrier scenario.

Two functional specializations of the generic x-RACF exist: Access-RACF (A-RACF) and Core-RACF (C-RACF): the A-RACF controls both access and aggregation segment (A-RACF), the C-RACF the core segment. Summarizing, main functions of x-RACF are:

- It provides the mapping of the parameters requested at application layer into the parameters needed at network layer;
- It provides Admission Control based on subscriber policies (only A-RACF) and available network resources;
- It provides Resource Reservation;
- It can be interconnected, in a tree structure, to other x-RACF controlling the same network resources;
- It can be centralized and/or distributed in the network nodes;
- It can only operate in a single-domain and single-carrier scenario;
- It can operate in both policy-push and policy-pull mode: an example of RACS support and operation of a "proxy QoS reservation request with policy-push" is shown in Figure 8. In this case, when the User Equipment (UE) invokes a specific service of an Application Function (AF), the AF will issue a request to the RACS for QoS authorization (policy control) and resource reservation. RACS policy decisions are "pushed" to the policy enforcement point in the NGN access and or core;
- It can support both unicast and multicast services.

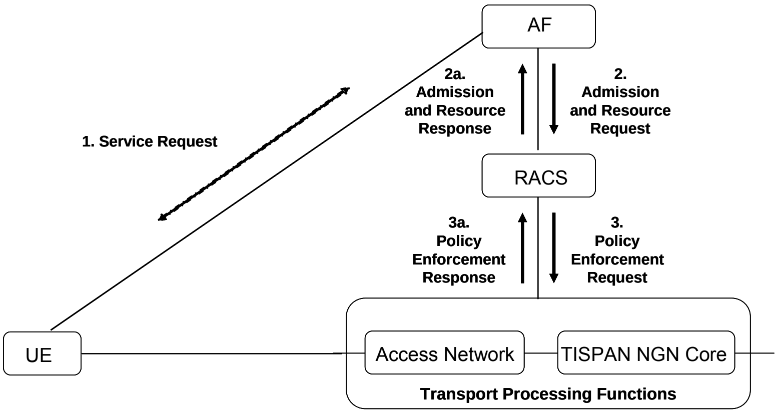

AF
2a. Admission and Resource Response
2. Admission and Resource Request
1. Service Request
RACS
3a. Policy Enforcement Response
3. Policy Enforcement Request
UE
Access Network
TISPAN NGN Core
Transport Processing Functions

Figure 8 – Example of user requested QoS with policy-push

Concerning with existing RACS limitations, in particular looking at STRONGEST context and requirements, a possible list is herewith provided:

- RACS only operates in a IP (routed) network and doesn't support L2 transport;
- RACS does not support traffic engineering (PWs, TE-Tunnels, LSPs…);
- RACS doesn't support path computation and GMPLS Control Plane interaction;

- RACS only supports routed data flows and L3 (5-uple) description of data flows. RACS doesn't support other parameters to identify a data flow or an aggregate (e.g. vlan id, PW id, Tunnel id…);

- RACS only supports Diameter as communication protocol for all the interfaces either at Application Layer, at Network Layer and at Control Layer.

4.1.4 RACF in IST-IP NOBEL2

Within IST IP NOBEL2 "Next generation Optical network for Broadband European Leadership" [NOB2-D43] a similar idea, to the one expressed in the previous section, to extend [ITU-T-RACF] to be adapted to GMPLS networks requirements was already considered. In particular, an extension to the RACF functional component, called RACF-G (G stays form GMPLS), was proposed with the aim to enable the NGN to directly request GMPLS-based connectivity to support NGN-based services. The extended features of the RACF included:

- New traffic engineering capabilities for the provisioning of NGN-based services in a multi-layer intra-provider network;

- New mechanisms that allow the extended RACF to discover and to control GMPLS resources;

- New mechanisms that allow the extended RACF to establish and to modify GMPLS connectivity at the boundary of GMPLS networks.

In addition, the RACF capability to directly interact or to cooperate with the PCE and the possibility of the RACF to include path computation capability was also investigated.

4.2 Definition of control plane requirements

4.2.1 Routing requirements for reference scenarios

Control plane reference scenarios described in chapter 2 are composed by heterogeneous networks spanning different domains and technologies (i.e., regions). In order to address the main issues concerning multi-domain and multi-region networks, two solutions have been identified:

- **OIF E-NNI OSPF-based routing solution** (GMPLS protocol based and in conformance with ITU-T G.7715 architecture). The aim is to allow the exchange of Traffic Engineering (TE) information between domains using an OSPF hierarchical architecture on top of the different domain (called Routing Control Domain or RCD).

- **IETF PCE solution**. (focused on centralized Path Computation Elements, Communication Protocol (PCEP) to allow Path Computation Client (e.g. ingress node) to request a Path Computation to a PCE. Discovery protocols to discover PCE location and capability are also considered in current IETF standards ([RFC5088 and RFC 5089]).

Their integration would lead to a hierarchical PCE hybrid solution, exploiting the main advantages of both, where the path computation is performed according to the routing information of summarized topologies, as shown in Figure 9.

Considering WSON regions of the reference scenarios described in chapter 2, the Routing and Wavelength Assignment (RWA) issue must also be considered. The following architecture options for RWA will be explored:

- **Combined RWA (R&WA)** - Here path selection and wavelength assignment are performed as a single process.

- **Separate Routing and WA (R+WA)** - In this case a first entity performs routing, while a second performs wavelength assignment. The first entity furnishes one or more paths to the second entity that will perform wavelength assignment and possibly final path selection.

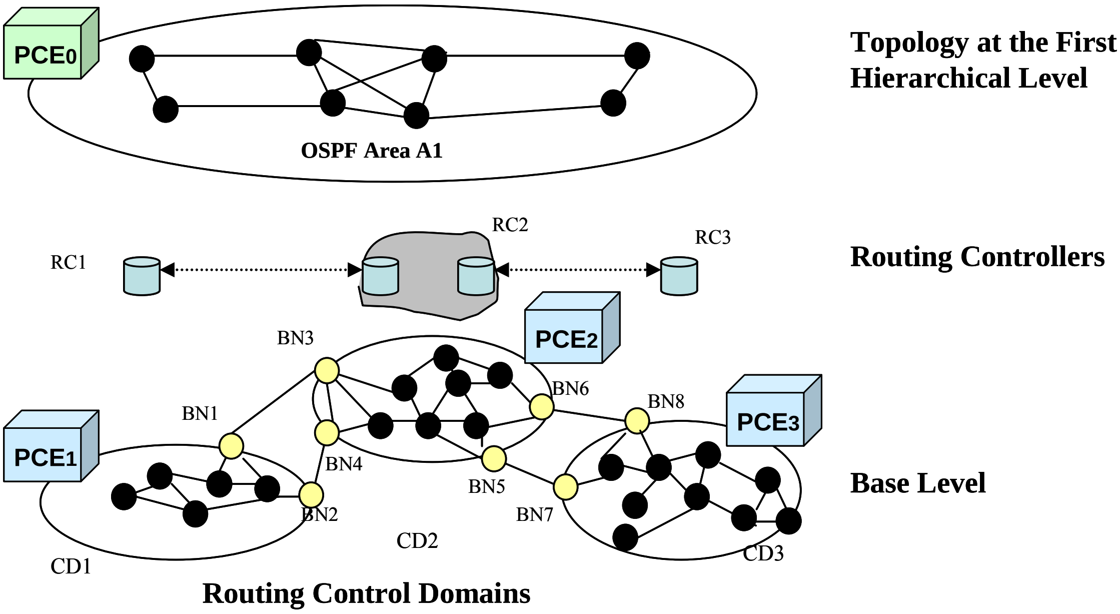


Figure 9 – Hierarchical PCE hybrid architecture

- **Routing with Distributed WA (R+DWA)** - In this case a first entity performs routing, while wavelength assignment is performed on a hop-by-hop manner along the previously computed route via signaling (RSVP-TE). This mechanism relies on updating of a list of potential wavelengths used to ensure the wavelength continuity constraint.

The considered routing architecture should also satisfy the following routing requirements:

- An OSPF area is built on top of the multi-domains network by the instantiation of at least one OSPF router per Routing Control Domain.

- An abstract topology is created by means of a topology summarization method that is actually under deployment.

- The routing information exchanged within the network should include all technology-related parameters that would be needed for an efficient path computation.

- Path computation in multi-region context should prefer grooming of resources (i.e. higher layer LSPs with lower Bandwidth granularities should be groomed as much as it is possible when nesting into lower layer LSPs with higher Bandwidth).

- The routing information exchanged within the network should be detailed in order to achieve better path computation.

- The amount of routing information exchanged within the network should be kept limited in order to avoid scalability issues.

- Referring to reference scenario 3, where different carriers are considered, confidentiality should be preserved (i.e. domains topology information exchange should be avoided).

- Forwarding Adjacencies (FAs) should be created in lower layers and advertised in upper layers topologies as virtual TE links (i.e. resources not reserved) or TE links (i.e. reserved resources).

- Topology summarization should be performed according to an efficient summarization criterion, as described in Section 4.2.2.

4.2.2 Topology summarization

The routing architecture considered within STRONGEST Project is a hierarchical PCE-based routing architecture, where each routing level deals with different topology views and scopes. Higher routing levels correspond to higher levels of topology abstraction, by means of resource summarization, performed according to specific criteria.

Performing a topology summarization reduces the amount of information carried by the routing protocol within inter-domain contexts and it can also impact routing performance, depending on the conditions within the Routing Area and the use of tools that provide additional routing information.

Moreover, in a multi-domain hierarchical E-NNI model, externally advertised topology can be a transformed view of the actual internal topology of a contained Routing Area to provide information for computation of paths crossing the Routing Area, represented by advertisements of links and associated costs.

One of the very key issues concerning the topology summarization is the range of abstraction to be performed, looking for a good trade-off between a too detailed information and a too poor one. The following abstraction models are defined:

- **Full topology advertisement** – In this model E-NNI will include all domain nodes and physical links in its topology database, and will compute paths based on a full knowledge of link resource availability within the domain.

- **Abstract Link Model** – In this model the domain is advertised as a set of border nodes connected by a full mesh of abstract links (assuming that full connectivity is being advertised). Bandwidth and costs can be associated with each link, but the links may not reflect the actual topology within the domain, only the supported connectivity.

- **Abstract node model –** In this model the domain is advertised as a single node, so no internal domain topology is visible to the outside, and E-NNI links appear from the advertisements to terminate on different ports of the same abstract node. Advertisement of minimum information is desired for policy or scalability reasons.

- **More Complex Models -** In more complex models, a domain can be advertised with a combination of abstract links and pseudo-nodes, physical links and interior nodes, to reveal a more complex topology.

The performances of the above mentioned models, shown in Figure 10, are summarized in Table 2..

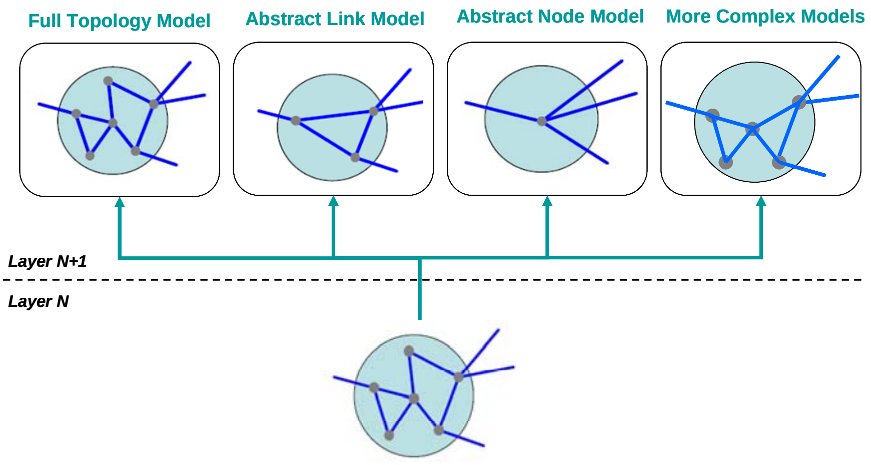


Figure 10 – Topology abstraction models

Table 2 – Performances of topology abstraction models

	TE efficiency	**Scalability**	**Confidentiality**
Full topology advertisement	High	Poor	Poor
Abstract node model	Low	High	High
Abstract link / more complex	Depends on the algorithm to	High (depends on the	High (depends on the

model	summarize the intra-domain topology	algorithm to summarize the intra-domain topology	algorithm to summarize the intra-domain topology

An efficient topology summarization criterion is being defined within STRONGEST, with the following advantages:

- optical and packet inter-working is allowed
- multi-domain routing should consider the aggregation of traffic request at different levels: at packet level and then at optical level
- multi-domain and intra-domain routing are separated
- multi-domain and intra-domain routing can be asynchronous
- each domain can adopt internal strategy (e.g. Optical domain can perform off-line path computing while packet domain can perform on-line routing)
- confidentiality and scalability is kept
- multi-domain advertisement due to topology change is reduced according to a thresholds mechanism

4.2.3 Signaling requirements for reference scenarios

As stated in [RFC5151], an end-to-end inter-domain TE LSP may be achieved using any combination of the signaling techniques described in [RFC4726]. That is, the options should be considered as per-domain transit options. The choice is a matter of policy for the node requesting LSP setup (the ingress node) and policy for each successive domain border node.

On receipt of an LSP setup request (RSVP-TE Path message) for an inter-domain TE LSP, the decision of whether to signal the LSP contiguously or whether to nest or stitch it to another TE LSP depends on the parameters signaled from the ingress node and on the configuration of the local node.

When a domain border node receives the RSVP Path message for an inter-domain TE LSP setup, it must carry out the following procedures before it can forward the Path message to the next node along the path:

- Apply policies for the domain and the domain border node.
- Determine the signaling method to be used to cross the domain.
- Carry out ERO procedures.

- Perform any path computations. In the case of nesting or stitching, either find an existing intra-domain TE LSP to carry the inter-domain TE LSP or signal a new one, depending on local policy.

Signaling requirements to be fulfilled within the STRONGEST Project are strictly related to the routing hierarchical architecture described in Section 4.2.1, in the context of reference scenarios described in chapter 2. Therefore, hierarchical LSP approach [RFC4206 and RFC4726] will be considered as the starting point for signaling architecture for all the three reference scenarios.

LSP hierarchy consists of a set of nested LSPs belonging to different hierarchical levels. In the context of MRN networks, different layers can correspond to LSPs of different technologies, considered in a hierarchical order, as shown in Figure 11.

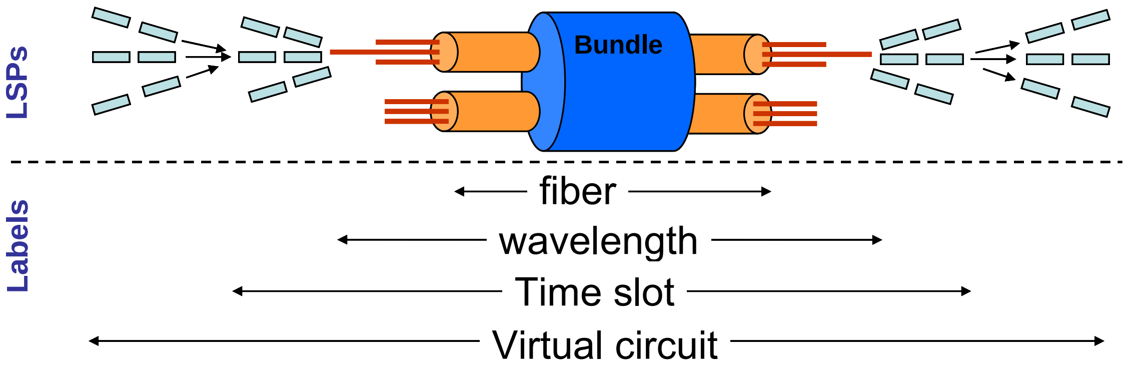


Figure 11 – LSP hierarchy

According to [RFC4206 and RFC4726], the signaling trigger for the establishment of a hierarchical LSP may be:

- Receipt of a signaling request for the TE LSP that it will carry
- Management action to pre-engineer a domain to be crossed by hierarchical LSPs
- Local policies at domain and/or region boundaries

Hierarchical LSPs may optionally be advertised as TE links within a domain and the mapping (inheritance rules) between attributes of the nested and the hierarchical LSPs (including bandwidth) may be either statically pre-configured or dynamically configured, according to the properties of the nested LSPs. Anyway it is worth noticing that inheritance from the properties of the nested LSP(s) can be also complemented by local or domain-wide policy rules.

The performed steps for creating a LSP hierarchy are the following ones:

1. A LSR creates a TE LSP that will result in a Forwarding Adjacency-LSP (FA-LSP)

2. The LSR forms a Forwarding Adjacency (FA) out of that LSP (by advertising this LSP as a TE link into the same instance of ISIS/OSPF as the one that was used to create the LSP)

3. Other LSRs are allowed to use FAs for their path computation

4. LSPs originated by other LSRs are nested into that LSP (by using the label stack construct)

5. The information carried in the ISCs is used to construct LSP regions (i.e. an ordering among interface switching capabilities, defined as follows: PSC-1 < PSC-2 < PSC-3 < PSC-4 < TDM < LSC < FSC) and to determine regions' boundaries

6. Path computation may take region boundaries into account when computing a path for an LSP

Figure 12 shows an example of nesting two higher layer LSPs into a lower layer LSP advertised as a FA.

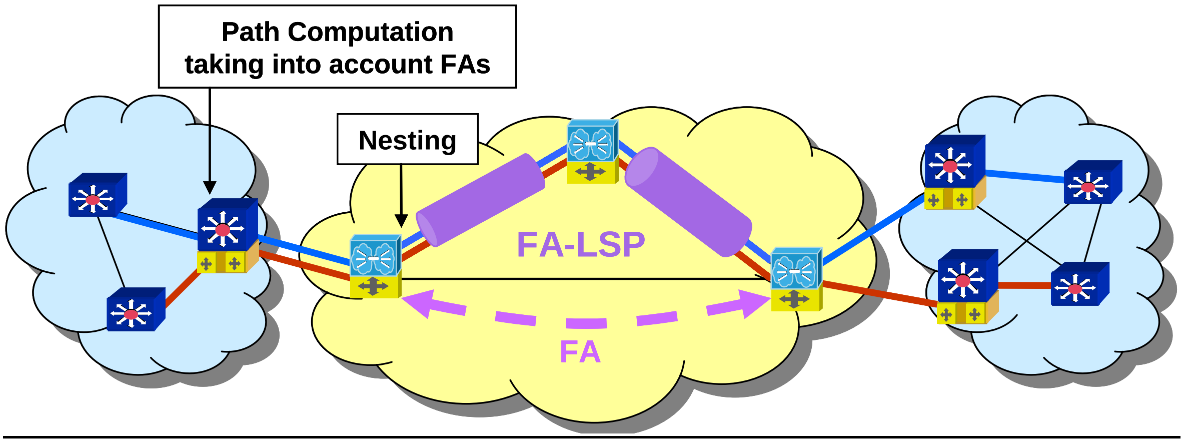


Figure 12 – Nesting with forwarding adjacency

The signaling procedure for LSP hierarchy, described in [RFC4206], states that for the purpose of processing the ERO in a Path/Request message of an LSP that is to be tunneled over an existing FA, an LSR at the head-end of the FA-LSP views the LSR at the tail of that FA-LSP as adjacent (one hop away).

At the edge of a region, the LSR receiving a Path/Request message must determine the other edge of the region (using ERO+IGP database), then it must compare the subsequence of hops toward the other end of the region with all existing FA-LSPs originated by the LSR. At that point the LSR should decide if using an existing FA-LSP or creating a new one. Finally, the Path/Request message for the original LSP is sent to the egress of the FA-LSP, with the PHOP set to the address of the LSR at the head-end of the FA-LSP.

In a hierarchical PCE architecture spanning different regions, the LSP hierarchy signaling procedures should be improved, when needed, in order to meet the following requirements:

- Grooming of different granularities should be performed on region border elements by means of local policies.
- Each domain should have its own policies concerning the dynamic or static creation of FAs.
- FAs should be advertised as virtual TE links (i.e. resources not reserved) or TE links (i.e. reserved resources), depending on the domain policies.
- If R+DWA is performed (see Section 4.2.1) then a set of possible wavelength should be carried in signaling protocol (e.g. the label set RSVP-TE object) in order to assign wavelength to the selected path.
- Signaling redundancy should be avoided.
- The choice between dynamically created or pre-computed lower layer FA-LSP (i.e. TE links on upper layers) should take into account the setup time (i.e. if several layers are involved each one of them should create dynamic FAs, then advertise them to an upper layer, a serialized process that is very time-consuming).
- The number of signaling sessions should be kept under control in order to avoid scalability issues.
- WSON elements such as transponders should be modeled, with impacts on signaling protocols also.

4.2.4 PCE/GMPLS requirements for WSON

The IETF PCE WG initial efforts were focused on improving path computation in MPLS networks, initially in intra-domain environments, and later extended to multi-domain networks. Thus, current set of published standards address the MPLS requirements, but do not address all the GMPLS network specific requirements. An example GMPLS network is shown in the next figures.

Each GMPLS domain can use several layers. A domain is defined as any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include Autonomous Systems [RFC4726]. A layer represents a given switching technology, for example, Packet, MPLS-TP, TDM (SDH/ODU) or Lambda switching (WDM). The interconnection of GMPLS domains can use, in addition, different technologies.

Figure 13 – Reference scenario 2 - multi-domain / multi-region

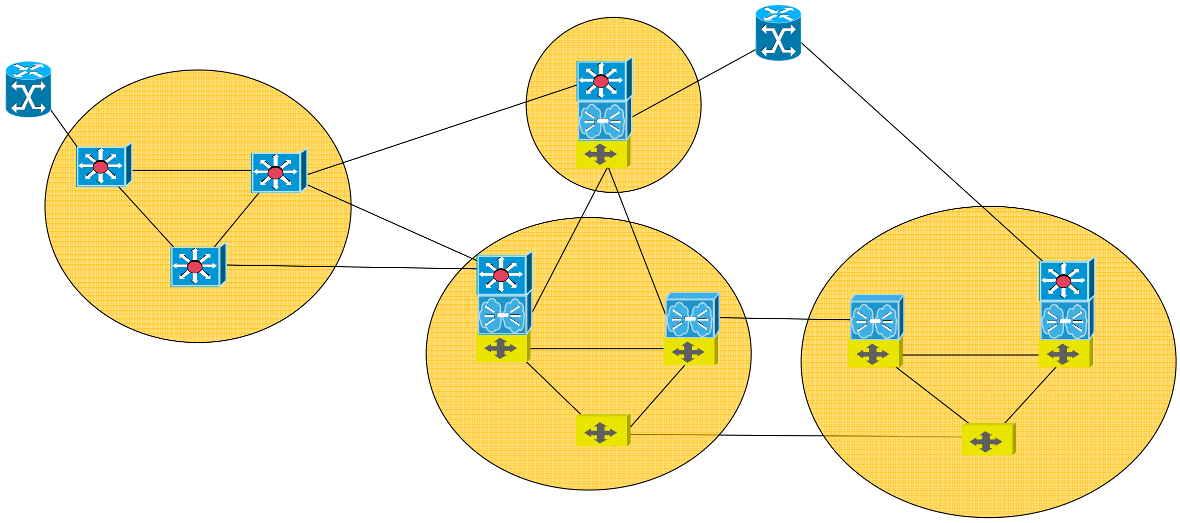


Figure 14 – GMPLS network example

4.2.4.1 Current GMPLS path computation requirements (non PCE)

For a connection spanning several domains or technologies, existing GMPLS path computation is designed to take into account the specific characteristics of those kind of networks, and, more specifically, the fact that:

- Data plane links can present different interface switching capabilities; connections can be requested for a given signal type and encoding, and may also need to make use of server layer connection(s) (using different signal type and encoding).
- Switching-technology specific traffic specifications cannot be mapped to a single bandwidth value. For instance :
    - Concatenation Type: In SDH/SONET and G.709 ODUk networks,
    - Concatenation Number: Indicates the number of signals that are requested to be contiguously or virtually concatenated. Also see [RFC4606] and [RFC4328].
    - Ethernet Traffic specification that match the Metro Ethernet Forum description [as described in draft-ietf-ccamp-ethernet-traffic-parameters-10.txt].
- Bandwidth priorities.
- Link protection.
- The network points (i.e. at which nodes) where adaptation (between layers) can be performed.
- Inter-layer adaptation capabilities.

In addition to the aforementioned requirements and restrictions, some data plane technologies may introduce additional specific constraints. For instance, DWDM has complex optical constraints to be considered in order to find a possible optical trail, and label switching (where the label corresponds to the assigned wavelength) is usually not possible due to the scarcity of wavelength converters (this is also known as the wavelength continuity constraint). It may also be possible that, on given nodes, label switching is restricted (for instance on 3R regeneration point or non-tunable transceivers) or optical signal constraints are present (Type of FEC supported).

GMPLS not only supports different data plane technologies but also supports extended TE-LSP functionality such as:

- Bidirectional LSPs.
- Asymmetric Bandwidth LSP (as introduced by RFC5467).
- End-to-end and segment protection.
- Unnumbered endpoint addressing.

The corresponding routing requirements are currently being considered in the IETF, namely in I-D. *ietf-pce-gmpls-aps-req* and I-D *ietf-pce-wson-routing-wavelength*.

Given the aforementioned points that enumerate and highlight core GMPLS functionalities, the existing PCE (as a functional architecture) along with the associated PCEP protocol (as a means to request path computation services conveying the required information objects, choices and constraints) need to be extended in order to allow the flexibility that the GMPLS umbrella already provides. This involves the identification of requirements and the subsequent proposal of PCEP extensions that cover most, if not all such requirements. The following section provides a summary of current ongoing work in that sense.

4.2.4.2 PCEP requirements for GMPLS

The PCE architecture does not mention the means by which it obtains the Traffic Engineering Database (TED) on top of which it is able to perform path computations, so the PCE should also consider that the information on constraints may be provided using the PCEP.

Based on those considerations, the additional requirements on PCE/PCEP for GMPLS/WSON are described in the next paragraph:

PCE and PCEP should support all the switching technology attributes supported by GMPLS. This includes the encoding, understanding and considering those attributes in the routing calculation. According to the RSVP-TE standards RFCs the set of attributes that need to be added to the PCE and PCEP are the following:

1. Switching capability : PSC1-1, L2SC, TSM; LSDC, FSC
2. Encoding type, as defined in RFC4202, RFC4203, e.g. Ethernet, SONET/SDH/Lambda, G709, Digital Hierarchy

3. Signal Type: Indicates the type of elementary signal that constitutes the requested TE-LSP. A lot of signal types with different granularity have been defined in SONET/SDH and G.709: ODUk, such as VC11, VC12, VC2, VC3 and VC4 in SDH, and ODU0, ODU1, ODU2, ODU2e, ODU3 ODU4 and ODUflex in G.709 ODUk [RFC4606] and [RFC4328]

4. Signal tolerance for G709 (work in progress, ID-draft-zhang-ccamp-gmpls-evolving-g709)

5. Concatenation Type: In SDH/SONET and G.709 ODUk networks, two kinds of concatenation modes are defined: contiguous concatenation which requires co-route for each member signal and requires all the interfaces along the path to support this capability, and virtual concatenation which allows diverse routes for the member signals and only requires the ingress and egress interfaces to support this capability. Note that for the virtual concatenation, it also may specify co-routed or separated-routed. See [RFC4606] and [RFC4328] about concatenation information

6. Concatenation Number: Indicates the number of signals that are requested to be contiguously or virtually concatenated. Also see [RFC4606] and [RFC4328].

The next aspect is requirement for routing a path over a given set of link protection types, as defined in RFC4203: in GMPLS the network trail(s) represented by the TE-link can be protected, currently GMPLS allow for a combination of unprotected, extra-traffic, 1:N, dedicated 1:1, 1+1, Enhanced (for example 4 fiber BLSR/MS-SPRING). This also applies when the server trail is a GMPLS service.

PCE/PCEP should also support the use of unnumbered interfaces as defined in [RFC3477]: in GMPLS, TE-Links can also be addressed by the couple (Node Id, interface ID) where NodeID is an IPv4 address and interface ID a 32 bit identifier. While this is already supported in the returned path (ERO) since the ERO includes all RSVP-TE route-sub-objects, the current PCEP ENDPOINTS object specification only covers IPv4 and IPv6 addresses and does not support this kind of addressing. In general terms, all GMPLS means to identify a given TE link or interface should be supported by the ENDPOINTS PCEP object. GMPLS and RSVP-TE also support asymmetric bandwidth requests [RFC5467]. PCEP protocol extensions are needed to support this, since a single IEEE 32 bit floating point value is used for bandwidth, regardless of whether it is a unidirectional or bidirectional path.

On some switching technology such as Lambda Switching Capable networks, the path computation engine should be able to perform not only routing calculation (routing as the selection of the physical path formed by links and nodes) but also resource allocation (such as label assignment) when needed. This is needed mainly for networks where constraints (label switching, physical constraints) are present and a distributed control plane cannot calculate optimally a path (or decide on the label). In addition the absence of resource allocation and constraint verification might lead to a high number of crankback operations during LSP setup and increased blocking. The routing request should include which operation the routing algorithm should perform (from I-D ietf-pce-wson-routing-wavelength). This translates in the routing request as an indication on which operation should the PCE perform and on the level of details present in the routing response (ERO) returned by the PCE:

1. A Routing calculation only (Node only, node and TE-Link, node and component link)
2. A Routing and Label assignment (in addition a label or a label set)
3. Re-optimize the path, the labels or both.

In the context of Multi-Layer or Multi-Region networks, a GMPLS control plane instance might not have a full TED, so it should be possible in the path returned by the PCE to indicate details on the different boundaries to be considered in signaling. To achieve this, the control plane needs to know to which layer (and possible adaptation possibility) each of the interfaces in the returned route belongs to. As this information might not be provided by the IGP, it should be present in the ERO provided by the PCE. In summary, the following information can then be provided on a Path computation response:

1. Layer boundaries for multi-layer route
2. Label to be considered or used (explicit labels or a set of label to consider during path establishment and label allocation)
3. Has the route returned passed optical quality check (as it can influence the crankback mechanism)

The GMPLS control plane might also need to provide a set of restrictions that may apply to the route calculation (requested for the TE-LSP, policy based, depending on the availability of TED, etc.). Those new constraints to be supported by PCE/PCEP are:

1. Restrict the switching layers to be considered during routing.
2. Restrict the inter-layer path computation.
3. Support for mono-layer/multi-layer paths.
4. Support for inter-layer constraints.
5. Support for adaptation capability which layer adaptation are allowed and possibly where.
6. Support for inter-PCE communication: those restrictions can be passed from one PCE to another; those restrictions should not require having the TED knowledge´.
7. Support for inter-layer diverse path computation.
8. The set of labels to be considered for the routing calculation (end-to-end or on a specific trail).
9. Suggestion on the label to be assigned.
10. DWDM signal compatibility restriction.

Finally, if all those new parameters on the routing request might fail, corresponding error indication should be provided in case the any of the previous constraint failed.

Those requirements are currently considered in a number of drafts:

- I-D draft-ietf-pce-inter-layer-ext
- I-D draft-margaria-pce-gmpls-pcep-extensions
- I-D draft-zhang-pce-pcep-extensions-for-gmpls
- I-D draft-lee-pce-wson-signal-compatibility
- I-D draft-fuxh-pce-boundary-explicit-control-framework

4.3 STRONGEST control plane architecture

4.3.1 Proposed architecture

In the present section, a complete Control-Layer and Control-Plane architecture for the STRONGEST reference scenarios is presented and analyzed. In particular, at this stage, the only scenario under study is the multi-domain/multi-region/single-carrier scenario, described in Section 2.2, since it appears to be the most likely for an intra-carrier network deployment. Anyway, the other STRONGEST scenarios will be further analyzed. The proposed architecture aims at addressing the aforementioned requirements in terms of routing, signaling and path computation (cf. Section 4.2)

To design a *complete* Control-Layer and Control-Plane architecture for the STRONGEST scenarios means:

- To propose the functional, protocol and physical architecture components for the control layer and control plane. The functional aspect depends on functional elements and their relationship and communication interfaces. The protocol aspect depends on the corresponding control protocol family (e.g. GMPLS) used for the provisioning and management of network connectivity services. And the physical aspect depends on how the different entities are interconnected and the mapping in terms of data plane and control plane links, nodes and attributes.
- To deploy an *effective* PCC/PCE architecture, in order to perform traffic engineering and path computation in a multi-domain, multi-region (layer) and multi-carrier scenario, using different parameters such as routing information, and/or QoS, bandwidth and policy constrains defined either intra- or inter- operator (cf. [RFC5394 - PEPC]).
- To define the protocol extensions and related signaling procedures for the dynamic establishment, management and release of connections (label switched paths), covering the multi-domain and multi-layer aspects as detailed in the reference scenarios, including specifics for:

   - coordinated establishment of higher / lower layer LSPs (with dynamic and static Virtual Network Topologies) in inter-layer networks,

   - point-to-multipoint connections,

   - inter-domain.

- To build the previous TE-specific features into a wider NGN Control-Layer framework, as the ITU-T RACF, the ETSI TISPAN RACS, or the 3GPP PCC (more in a fixed-mobile integration perspective), supporting various value added services, both unicast and multicast. In the present document, only the ETSI TISPAN RACS architecture (cf. Section 4.1.3) will be considered. A similar study for the ITU-T RACF extension has been performed in IST NOBEL2, as depicted in Section 4.1.4.

An *effective* PCC/PCE deployment would provide an optimal path computation:

- Without distributing detailed topology and/or routing information between different domains (especially administrative), considering:

   - Other than the link state routing protocol that may be executed within a given TE / routing domain, a link state routing protocol with TE extensions could be allowed in a multi-domain/single carrier scenario (e.g. a separate OSPF inter-domain area can be configured), for the purpose of controlled topology, resource and reachability information dissemination at the domain boundaries.

   - A path vector routing protocol with TE extensions should be configured for inter-carrier routing.

   - In any case, the inter-domain/carrier summarization of routing and resource information is a critical point, and requires finding a good trade-off between granularity of the information, optimality and responsibility. It is worth noting that potential architectures may loosen slightly these two aforementioned requirements, provided that there is some mechanism that allows inter-domain links to be disseminated within the corresponding domains, thus not requiring neither a separated OSPF-TE instance/area nor a path vector routing protocol.

- Based on a detailed analysis of costs versus benefits when spreading detailed routing and resource information between different regions/layers, in order to:

   - Avoid scalability issues, for it is clear that the size and processor/memory requirements increase with the complexity of the Traffic Engineering Database (TED), which is also related, amongst other things, to the different switching capabilities, their hierarchical relationships and the eventual, technology specific, layer attributes and constraints.

   - Avoid redundant information (e.g. in a hierarchical architecture, replicating the same topology view at different layers).

- Avoid the spreading of useless information (e.g. available lambdas in a MPLS-TP region). One of the main limitations is that topology and resource information is disseminated by a protocol using information objects (e.g. Link State Advertisements) that include all traffic engineering attributes, so it is not straightforward to partition the attributes into those worth crossing region (layer) boundaries. This is an open issue in aggregation mechanisms that could be proposed to address this.

- Taking into account the inter-layer coordination, both for routing and signaling:
    - A path computation in different layers could assume PCE modules with full visibility.
    - A path computation in the upper layer could take advantage having an already active path in the lower layer. A path set-up in the lower layer modifies the topology and the available resources in the upper layer.
    - Signaling strongly depends on routing model (e.g. end-to-end, per-layer and/or per-domain path setup).
- Without having the need of slow and complicated crankback procedures, as in Per-Domain Path Computation.
- Easily working in a mesh of domains. Some optimal solutions in terms of path selection and routing/resource information spreading, such as Back-Recursive Path Computation, have problem navigating a mesh of domains. Other approaches may assume that the sequence of domains is known in advance, requiring some (supra) entity to provide such sequence. In any case, it is expected that the mesh of domains is of reasonable low size and a large mesh is not considered (e.g. the whole Internet).
- Using various information sources and repositories (apart from routing and resource constraints) for common (policy-enabled) control plane procedures, such as operator defined policies per domain/service/subscriber. Such information should be available by an external policy manager.

The previous remarks point out how the most critical issues in deploying an effective path provisioning architecture concern:

- Scalability
- Inter-domain/carrier selection and navigation
- Inter-layer coordination and cooperation
- Inter-domain and inter-layer path signaling, covering:
    - Requirements for signaling interfaces at domain boundaries (e.g. E-NNI)
    - Dynamic establishment of forwarding adjacencies

All the previous issues would be solved, or heavily limited, by control plane architectures using a hierarchical PCE (as detailed in Section 4.3.3). Moreover, implementing such hierarchical PCE architecture in a wider Control-Layer framework, like the ETSI TISPAN RACS would:

- Enable unicast and multicast NGN services, by means of well defined application interfaces, available either in *push-mode* (from *off-path* application servers) and in *pull-mode* (*on-path*, from the same network nodes involved in the data transport, as for example in the case of multicast IGMP Join requests).

- Offer an interface towards the offline charging and billing systems, easily allowing scenarios of tariff differentiation per service characteristics (e.g. the QoS actually provided). Provide a clear separation between equivalent operations (e.g. the admission control) performed at service layer or at network layer. Such separation would also allow strategic NGN scenarios, as for example the separation between the network provider and the service providers

- Use, in the path computation procedures, various operator defined policies and constraints, which could be configured per domain, per service, or per subscriber with a major granularity.

4.3.1.1 Control plane architecture summary

In summary, the Strongest Control Plane architecture targets:

- A protocol architecture based on extended GMPLS protocols: OSPF-TE and/or IS-IS as IGP routing protocol, (E) BGP with TE extensions where appropriate as EGP, RSVP-TE as signaling protocol, PCEP as the protocol to access path computation services and for inter-PCE communication, LMP as Link Management Protocol

- Flexible support for path computation in provisioning and restoration: from source based path computation to distributed path computation using collaborative PCEs in BRPC or hierarchical settings.

- Flexible support for inter-layer Traffic Engineering, allowing either overlay, augmented or peer deployment models, with decoupled or unified path computation.

- Extending Multi-Layer and Multi-Region control procedures and existing solutions (such as Forwarding Adjacencies or Virtual Network Topology), focusing on the specific case of MPLS-TP (as a packet switching capability layer) over WSON (as a lambda switching capability layer)

- Seamless operation in multi-domain settings, covering STRONGEST identified scenarios.

- Extended by means of a Control-Layer framework, like the ETSI TISPAN RACS G-RACS (the specific case of STRONGEST Scenario 2 is detailed in the next section), in order to:

- Have a middleware between application and transport functions
- Provide admission and policy control based on parameters defined by the service provider

4.3.1.2 Proposed architecture in STRONGEST scenario 2

In Figure 15, the proposed Control Layer and Control Plane architecture for STRONGEST Scenario 2 is reported.

Since Access Networks are outside the STRONGEST scope, they are modeled as Local Area Networks (LANs) directly connected to an ingress LSR. For the same reason, all the depicted x-RACF instances can be considered of C-RACF type, and the present study won't consider any specific A-RACF task, such as the admission and policy control, basing on the subscriber (end-user) presence, localization and profile information (derivable from the e4 interaction with the NASS).

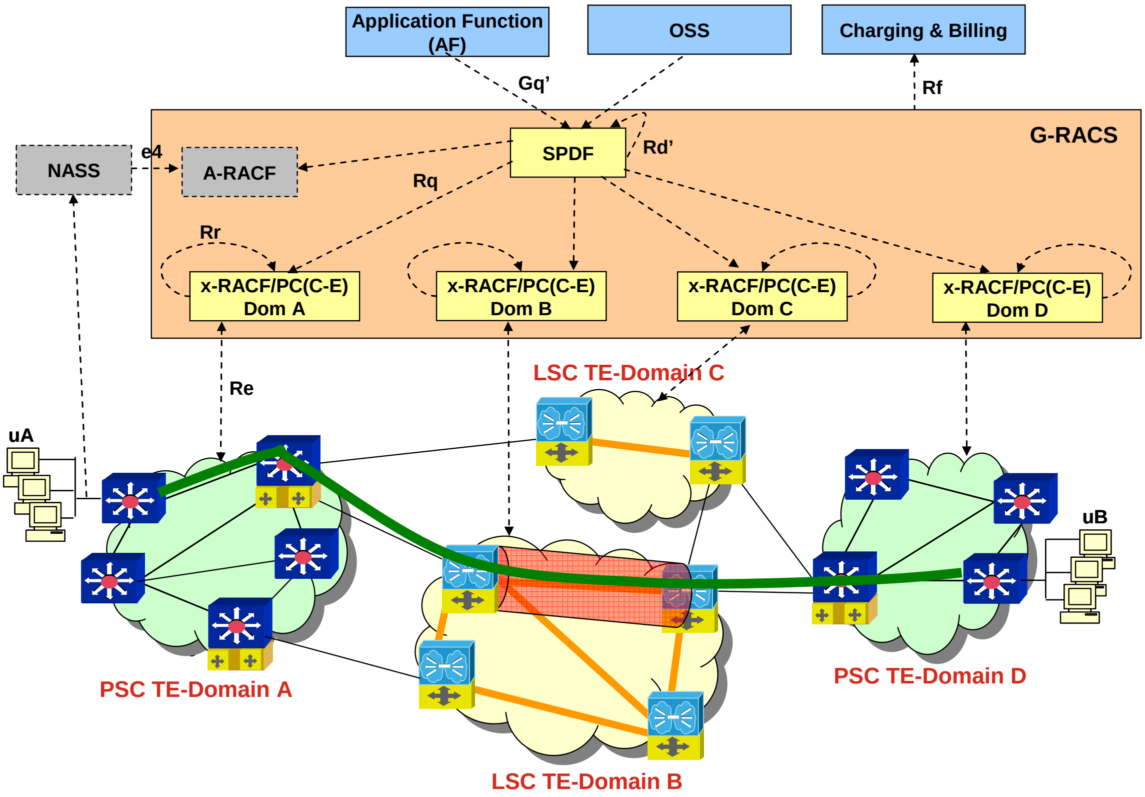


Figure 15 – G-RACS architecture

In Figure 15, the following functional elements are reported:

- SPDF, whose role is:
  - Single contact point with the Application Layer, in order to provide network topology and technology abstraction to the applications

    - SPDF shall be able to process AF requests in terms of an e2e data flow transport request (identified by a 5-uple) with resource constraints (in terms of peak and guaranteed bandwidth, maximum delay and jitter, media type, priority, security…)
    - SPDF should be able to process OSS requests, in terms of path provisioning requests between two end points (identified by the ingress/egress LSR and interface IDs), with some constraints (in connectivity, costs, resources…)
  - Policy Decision Function based on service policies and constraints defined by the operator, for example in terms of:
    - Inter-Domain connectivity (inclusions/exclusions), cost…
    - Resources (bandwidth, delay, jitter, class of service, priority…)
    - Paths (e.g. shared path selection, or dedicated path setup)
  - Coordination of multiple x-RACF/PCE modules, in order to satisfy e2e service requests
  - Interface towards the offline charging systems
- x-RACF/PCE, which role is:
  - Admission Control based on:
    - Subscriber profile (only in Access Network)
    - Intra-Domain resources and constraints
  - Intra-Domain Path Computation (stand-alone or in coordination with other x-RACF/PCEs)
  - Interface with the GMPLS Control Plane, in order to trigger
    - The path setup/update/delete signaling
    - The enforcement of L3 data flows within the paths

Network topology considerations

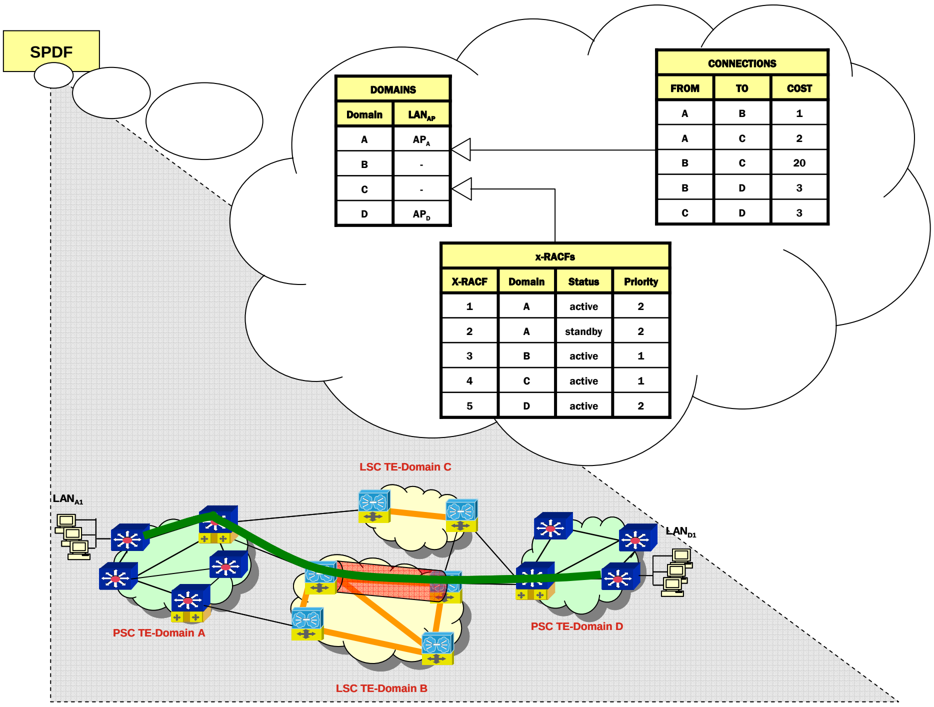
In a hierarchical architecture like the previously proposed G-RACS, each hierarchical layer should discover and virtualize the network topology in a different way, collecting essential information, and implementing a different view of the same network resource. Such operation of topology discovery and virtualization in different hierarchical layers should avoid as much as possible redundancies and duplicated information (cf. Section 4.3.3).

The SPDF, since its main role is to identify the list of the traversed domains, and the related list of the x-RACF/PCEs to call, for an e2e transport service:

- Should collect *rough* information about the domain interconnections. For example, as depicted in Figure 16:
    - Each domain could be modeled as a virtual router.
    - Each domain interconnection could be modeled as a link between two domains. Each link could have a cost configured/calculated, according to various operator-defined TE policies and parameters (e.g. involving requested QoS parameters, domain technologies, service privileges…).
- Should collect information about the x-RACF/PCEs controlling each domain. For example:
    - Multiple x-RACF/PCEs could be associated to the same domain for reliability reasons.
    - Each x-RACF/PCEs could have an assigned priority, to determine the exact sequence of x-RACFs to call in an e2e service request. For example, in a multi-layer scenario, the x-RACF/PCEs attending WSON domains should be called first, in order to establish the lower layer connectivity.

The x-RACF/PCEs could operate at different layers. For example, in a MPLS-TP domain, they could be divided in:

- x-RACF/PCEs operating at (lower) GMPLS layer could perform intra-domain computation, admission control and resource reservation for the new requested paths, using many routing and TE information. Hence, they can map both PCE and GMPLS Control Plane operations, such as:
    - To build a complete domain topology view, using protocols as OSPF-TE.
    - To use locally configured constraints, weights and load balancing policies.
    - To manage GMPLS sessions, in order to create, update and delete paths.
- x-RACF/PCEs operating at (upper) IP layer could perform admission control and resource reservation for the user data flows within the existing paths (as in the example of Figure 17). In particular:
    - They can trigger new path computations (acting as a PCC).
    - They can build a L3 topology view, using routing information and/or path computation results.
    - They can manage end user sessions, for a single e2e user data flow management.

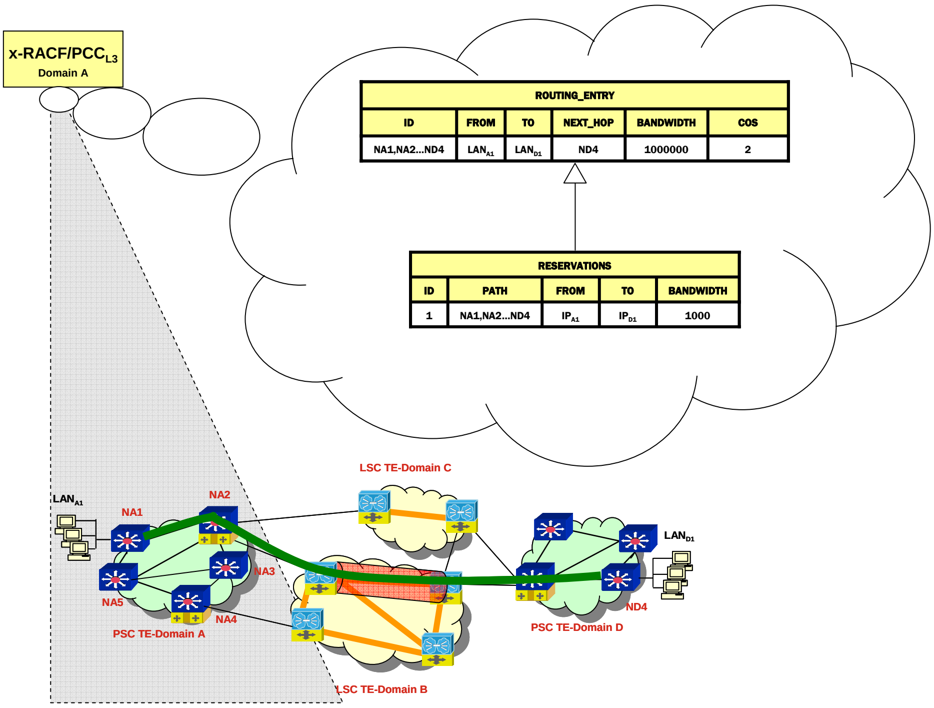


DOMAINS	
Domain	LAN_{AP}
A	AP_A
B	-
C	-
D	AP_D

CONNECTIONS		
FROM	TO	COST
A	B	1
A	C	2
B	C	20
B	D	3
C	D	3

x-RACFs			
X-RACF	Domain	Status	Priority
1	A	active	2
2	A	standby	2
3	B	active	1
4	C	active	1
5	D	active	2

Figure 16 – Example of SPDF topology virtualization



ROUTING_ENTRY					
ID	FROM	TO	NEXT_HOP	BANDWIDTH	COS
NA1,NA2…ND4	LAN_{A1}	LAN_{D1}	ND4	1000000	2

RESERVATIONS				
ID	PATH	FROM	TO	BANDWIDTH
1	NA1,NA2…ND4	IP_{A1}	IP_{D1}	1000

Figure 17 – Example of x-RACF/PCC L3 topology virtualization

x-RACF/PCE deployement considerations

In the ETSI TISPAN RACS model (cf. Section 4.1.3), the x-RACF can be either fully centralized, or fully distributed (in the network nodes), or partially distributed (with many x-RACF instances, organized in a tree structure, and controlling the same network resource. Only the root x-RACF is allowed to communicate with the SPDF). The same three options could be evaluated for the x-RACF/PCE modules.

The fully centralized x-RACF/PCEs option appears to be the trickiest and less optimized, since it:

- Requires to collect in a centralized server, all the routing and TE information already available in the LSRs.
- Could have scalability issues.
- Certainly is not optimized in terms of information duplication and redundancy.

The fully distributed x-RACF/PCEs option appears to be more practicable, since, as previously specified, x-RACF/PCEs:

- Use routing and TE information, available in the LSRs, for path computation.
- Cover GMPLS Control Plane functionalities, such as the admission control and resource reservation, or the session management at (GMPLS) path level.

However, to completely cover x-RACF/PCE functionalities, the GMPLS Control Plane should be extended to support admission control, resource reservation and session management of the end user IP data flows within the GMPLS paths (as, for example, in the “pure” RSVP IntServ scenario). Hence, such approach:

- Requires extensions in the GMPLS Control Plane.
- Could lead to scalability issues, due to the higher amount of session data to be stored in the LSRs.
- Will also (slightly) complicate the SPDF topology view, since the SPDF would need to directly address the x-RACF/PCE instance installed in the single (ingress/egress) LSR.

The partially distributed x-RACF/PCEs option appears to be the optimal option, due to the already discussed layered view of x-RACF/PCEs, and the lower impacts in the standalone RACS and PCE models. Within a domain, as depicted in Figure 18:

- A delegated x-RACF/PCC instance:
    - Would receive service requests from the SPDF.
    - Would address the x-RACF/PCE instances implemented in the (ingress/egress) LSRs.

- Would manage the end-user sessions at IP level, performing admission control and resource reservations within the existing paths.
- Could enforce traffic policies in the ingress LSRs.
- Would be technology independent, ignoring the particular transport (MPLS-TP, LSP…) implemented in its domain.

- The x-RACF/PCE instances implemented in the (ingress/egress) LSRs would implement PCE and GMPLS Control Plane functionalities, depending on the particular transport technology used in its domain.

Such solution would also need O&M connection, to permit the upper x-RACF/PCCs to be notified in case of path faults.

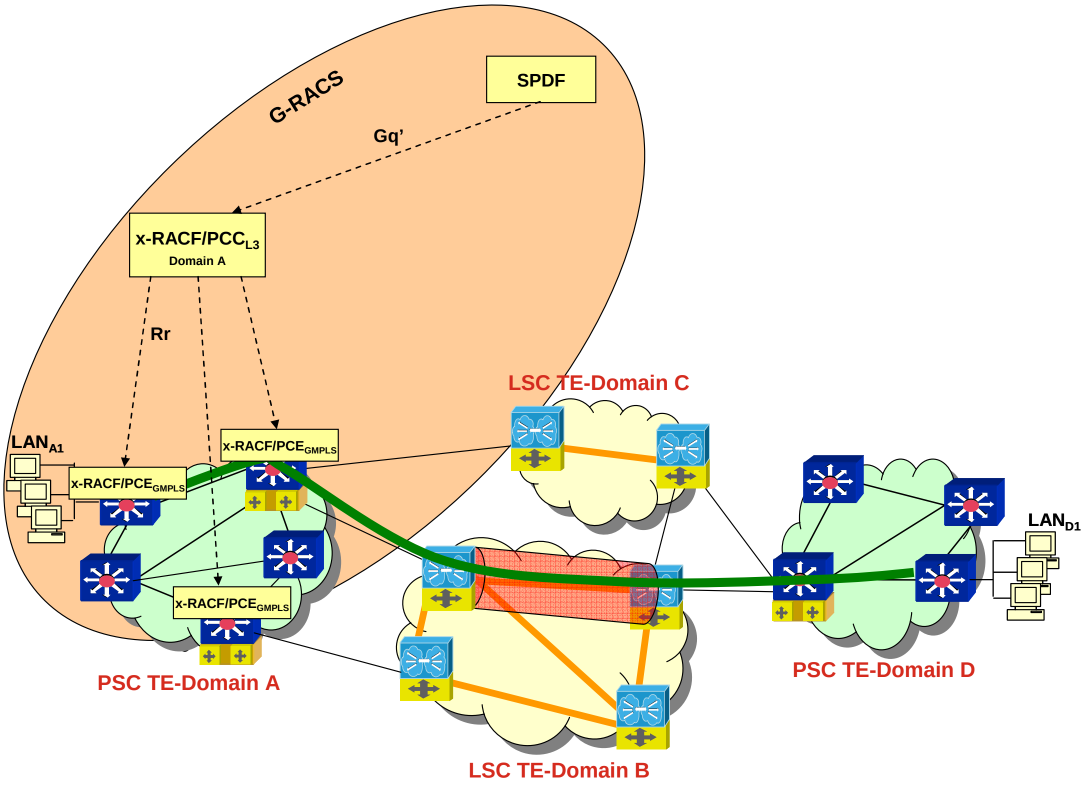


Figure 18 – Partially distributed x-RACF deployment

Protocol considerations

ETSI TISPAN defines Diameter as the only protocol used to interface every module at every level. RACS Diameter interfaces should be extended to support TE AVPs.

IETF standardized the PCEP protocol for the PC(C-E) communication. Further standardization work extending PCEP to support many TE parameters and transport technologies is ongoing, as detailed in Section 4.1.2.

The proposed G-RACS solution could initially consider PCEP as the protocol for the PC(C-E) communication, being an already existing solution, fulfilling all the needs. Protocol aspects for the G-RACS solution PC(C-E) communication will be though further investigated.

In the G-RACS solution, also the GMPLS Control Plane interaction protocol will need further investigation, since the UNI and NNI signaling would not be applicable from a control framework. In this case, the trigger solution would fit well.

4.3.2 Motivation for PCE in single domain scenarios

As detailed in the IETF PCE WG normative documents [PCE-WG], the motivations for the deployment of a single (or multiple) PCE(s) are several, and can be roughly divided into two main groups:

First, the motivations related to the capability to handle the computational complexity of advanced path computation algorithms when applied to different network topologies; complexity which can be increased when including additional technology-specific restrictions and constraints (for example, path computation in wavelength switched optical networks may require additional wavelength continuity or be constrained by the signal quality). In addition advanced path computation can help network operators to use the available resources more efficiently, enlarging the life of the deployed infrastructure.

Second, additional benefits can be identified when the PCE appears as a key functional component for use cases where several shortcomings have been identified. An example of the latter is the case of multi-layer and multi-domain networks. In such networks, constraints such as topology confidentiality or TED scalability apply, and having clearly defined interfaces, protocols and functional entities and roles render the proposed solution simpler.

Path Computation Elements can be deployed in single and multi-domain scenarios, where a domain refers to a collection of network elements within a common sphere of address management or path computational responsibility such as Interior Gateway Protocol (IGP) areas or Autonomous Systems (AS). The well defined interfaces (i.e. between a Path Computation Client or PCC and a PCE and between two collaborating PCEs) allow the specification of path computation models where end-to-end path computation can be decoupled into sub-problems, delegating specific path segments to one or more dedicated PCEs.

Although significant effort is being targeted at the later issues, PCE applicability in the single layer scenario is still significant. The purpose of the present section is to provide a list of topics in which, although restricted to a single domain, the PCE still provides a lot of value and justify one or more dedicated PCE servers.

Decoupled control of path computation algorithms

In common deployments, GMPLS-controlled transport networks are organized in islands of nodes, acquired from the same equipment vendor (they are often referred to as vendor domains). Such islands may be bound to a set of proprietary applications and management systems, and the nodes themselves are highly integrated in terms of bundling the control and transport planes. Although this approach has proven itself appropriate in most cases, it precludes network administrators to adapt and extend control plane aspects, which are seen as a “closed box”.

If we define a control plane as the set of functions, protocols and entities that allow the distributed establishment, management and release of Label Switched Paths (LSP) or connections, it is possible to identify several main control plane components: link discovery and link management, topology and network resource status dissemination, path computation and signaling. In this sense, by extracting the path computation function out of the GMPLS controller, and by defining an open interface, network operators may deploy their own path computation servers (or delegate them to third parties), deploying proprietary optimized path computation algorithms that are more under the control and policy of the network operator, for a greater flexibility and a finer control.

It must be noted that the network resources, e.g. wavelengths in WSON networks, are scarce, and if wasted, additional fiber links should be deployed and the equipment upgraded. The utilization of optimized algorithms can help to get a more efficient use of these resources. Moreover, the optimal routing algorithm (in terms of CAPEX reduction) may be different for every single network scenario and operator. By decoupling the path computation from the control plane, the operator could be able to implement customized algorithms to optimize the ROI. The algorithm customization option is almost non-existing where the path computation is clearly integrated in a closed vendor system.

Synchronized / dependent multiple path computation

The PCEP protocol has been defined to allow maximum flexibility in requesting path computation services from a PCE. Although the common case is the request for a path computation that refers to a single point to point route, the PCE and PCEP protocol allow bundling several requests, not only “physically” (as bundled in the same message) but also “logically”, to be used when network operators wish to be able to perform multiple path computations.

Such path computations may be independent or dependent. In both cases, having the possibility to synchronize them allows further optimizations. In the specific case of dependent path computations, the requests must be synchronized in order to meet specific objectives (such as being link or SRLG disjoint). The basic use case for synchronized connectivity requests comes from the need for disjoint routes for path protection purposes.

GMPLS based recovery

The usage of a PCE in GMPLS-based recovery scenarios is still open to evaluation. Roughly speaking, the applicability of a PCE strongly depends on the considered scenario: for example, for pre-planned restoration or the computation of 1+1 and 1:1 (SRLG)-disjoint paths, the PCE appears as an excellent candidate due to the dedicated computational resources. However, some concerns may be raised when the same PCE is applied, for example, to dynamic restoration: given the fact that a PCE deployment model is often “one

PCE per domain with optional redundancy", it is still unclear the impact of network latency, queuing and contention in those use cases, which may yield to unsatisfactory recovery times. On the other hand, a PCE can potentially provide a coordinated response to a failure event affecting multiple connections. In a typical GMPLS restoration scenario, the computation of the recovery path is performed independently at the sources of the affected LSPs, which may lead to a sub-optimal set of recovery LSPs. Moreover, as the computation is performed independently in the sources, the same resource could be chosen by different sources, leading to the later blocking of some of the LSPs setups. This situation can be avoided by a synchronized computation in the PCE.

In any case, a GMPLS controller may be able to signal paths with end-to-end or segment protection, but may lack the ability or required visibility for the constrained computation. Generally speaking, we can refer to these constrained path computations as "synchronized" or "dependent".

Re-optimization and re-configuration

The PCEP protocol allows a given path computation request to request, in fact, a re-optimization of an existing path. This is allowed by combining a request for a path computation with an existing Record Route Object (RRO) that details the resources that the path is currently using. With the RRO object, and using an optional bandwidth object, the PCE may re-optimize a path without incurring into double reservations or discarding links that do not have enough resources, which could happen without the knowledge of the existing path.

Policy based routing to support QoS

The PCE defines a framework to support differentiated routing for different service classes. Policy mechanisms can be implemented to select the proper algorithm for path requests with differentiated QoS requirements (e.g. real time services vs. Internet traffic). Additionally, a dedicated path computation engine can facilitate solving requests with additional constraints (e.g. QoS requirements) that need to be solved by complex algorithms. Note that this capability of the PCE could serve as a basis of integration with TISPAN architecture to provide end-to-end QoS management.

IA-RWA: impairment aware routing and wavelength assignment

As explained in 4.2.4, a particular case of notable importance in the context of the Strongest project, is the case of path computation in wavelength switched optical networks (WSON), which may require additional technological constraints, such as, for example, the wavelength continuity constraint (WCC). Moreover, since the signal propagates in the optical domain, it is sensible to the accumulation of physical impairments. To some extent, it is possible to extend the common control plane protocols such as RSVP-TE or OSPF-TE to take into account the physical parameters, such as the link Optical Signal to Noise Ratio (OSNR). However, a formal characterization and subsequent dissemination by the routing protocol of the node and link attributes may not be feasible or practical, requiring a centralized approach to do path computation or, more generically, IA-RWA: impairment-aware routing and wavelength assignment.

In this sense, the PCE is an appropriate solution to deploy advanced path computations when the standard GMPLS based traffic engineering attributes need to be extended. The "extended" TED is managed by the PCE, and used for path computation. Finally, the PCE can be used to compute either just spatial (i.e., links and nodes) or both spatial and spectral (i.e. wavelength) paths, named, respectively, R (Routing) and RWA (Routing and Wavelength Assignment).

Signal compatibility constraints

The signals used in wavelength switched optical networks are not always compatible with the network elements (regenerator, OEO switches, wavelength converters, etc.), as stated in draft-lee-pce-wson-signal-compatibility-01. Thus, PCE should be able to check the compatibility of the modulation format and FEC type of the transmitted signal through all the elements in the path. The WSON signals are characterized in draft-ietf-ccamp-rwa-wson-framework. Also, in line with the I-RWA calculation mentioned earlier, the PCE can indicate the special node processing required (e.g. regeneration points).

Point-to-multi-point path computation

As stated in draft-ietf-pce-pcep-p2mp-extensions, point to multipoint (P2MP) traffic engineered (TE) LSPs may be useful for several applications, such as are considered in support of various features, including layer 3 multicast virtual private networks [RFC4834]. The PCE has been identified as an appropriate technology to compute and re-optimize such paths, offloading such tasks from Label Switching Routers (LSRs) and GMPLS controllers. In this line, path computation requirements for point-to-multipoint (P2MP) MPLS TE LSPs are documented in [RFC4461], and signaling protocol extensions for setting up P2MP MPLS TE LSPs are defined in [RFC4875]. The applicability of the PCE-based path computation architecture to P2MP MPLS TE is described in [RFC5671].

The computation and network optimization of multiple P2MP TE LSPs requires considerable computational resources. As stated in the IETF draft, some of those PCEs might have the ability to satisfy certain objective functions (for example, least cost to destination), but lack support for other objective functions (for example, Steiner). Additionally, some PCEs might not be capable of the more complex P2MP re-optimization functionality. The PCE is suitable for the following reasons:

- Ability to compute and re-optimize P2MP paths with several constraints, with the addition/ removal of new/old trees given their RRO objects.

- The advanced determination and selection of branch points, constrained by the network topology and resources, and determined by the objective functions that may be applied to path computation.

- The need to take into account node capabilities regarding branching, control plane resources to accommodate extra signaling state, etc.

Multi-layer path computation

From the point of view of the involved network technology, a given network may be comprised of multiple layers, which can either represent:

- Separations of technologies such as packet switch capable (PSC), time division multiplexing (TDM), lambda switch capable (LSC) as stated in [RFC3945] (i.e., multi-layer),

- The separation of data plane switching granularity levels (e.g., PSC-1 and PSC-2, or VC4 and VC12) [RFC5212] (i.e., multi-region),

- A distinction between client and server networking roles (e.g., commercial or administrative separation of client and server networks), constructed from layered client/server networks.

In such a multi-layer network, LSPs in lower layers are used to carry upper-layer LSPs, e.g. by means of LSP nesting. The network topology formed by lower- layer LSPs advertised to the higher layer is called a Virtual Network Topology (VNT) [RFC5212]. To improve the network efficiency, it is reasonable to provide end-to-end traffic engineering across multiple network layers, comprising mechanisms to allow global optimization of the network resources, taking into account all layers, rather than optimizing resource utilization at each layer independently. This allows better network efficiency to be achieved (inter-layer traffic engineering). This includes:

- Mechanisms allowing computation of end-to-end paths across layers.

- Mechanisms for control and management of the VNT by setting up and releasing LSPs in the lower layers [RFC5212] and the underlying forwarding adjacencies. In some deployment scenarios the VNT may be static, or may change rarely. The PCE is able to seamlessly compute (and reconfigure) a VNT.

The PCE seems a good candidate for the multi-layer / inter-layer path computation problem with regard to the constraint that network operators may be reluctant to distribute routing information between layers. In this line, several methods have been defined for multi-layer computation:

- Single PCE with multi-layer visibility (each layer corresponds to a TE domain, with one PCE seeing them), which may present scalability issues.

- PCEs without cooperation, conceptually similar to the per-domain path computation.

- PCEs with cooperation, where some mechanism such as BRPC is used.

In summary, there is a clear and present need to address the applicability of PCEs in single domains.

4.3.3 Definition of multi-domain/multi-vendor architectures with hierarchical PCE

4.3.3.1 Introduction

The multi-domain network scenario is expected to be a fundamental part of new operators' network deployments. The main motivations for multi-domain architectures are, on the one hand, the fact that an operator's network may include several equipment vendors, and, on the other hand, the idea that segmenting the network into domains is a means to increase the overall scalability. In the particular case of optical (G.709 and WSON) networks, dividing the network in domains can ensure the all optical transmission without the need of regeneration, which is delegated to the domain's boundaries (insuring 3R signal regeneration). Also, wavelength continuity assurance is a challenge as the size of the network increases. Thus, wavelength assignment becomes easier when the network is divided in small areas. Finally, the control plane scalability is enhanced, since TE domains (such as OSPF-TE areas) do not grow extensively. This latter advantage can be noticed in performance criteria such as restoration performance, which is increased.

Taking into account the issues of multi-domain networking (not only at the optical layer), one of the key challenges is how to perform the multi-domain path computation. A very powerful tool for multi-domain path computation that includes constraints in the computation process is the Path Computation Element (PCE) [RFC 4655]. The PCE allows for several viable architectures to implement multi-domain path computation both in mono-vendor and multi-vendor networks.

The IETF has proposed several solutions to solve the multi-domain path computation using PCEs.

The first option is to perform a per-domain path computation. In this case, there is one PCE per domain, but there is no communication between PCEs of the neighbor domains. Following this approach, the sequence of domains and inter-domain links has to be known, or explicitly stated by the network operator. Thus, PCEs take care only of computing the intra-domain paths from/to the edge nodes of the involved domains.

The per-domain path computation is a simple solution, but due to the lack of global information, it is clearly sub-optimal. Thus, the next step, proposed by Farrel [PCE-WG] is the "simple cooperating PCEs" and involves some exchange of information between PCEs. Like in the previous case, the sequence of domains has to be known. However, the PCE of one domain asks its neighbor PCE for the best entry point to its domain. Then, the PCE computes the optimal path to its egress node. Although the results are slightly better, its use in a network involving more than 2 domains in a path is still sub-optimal.

The next solution, which involves more cooperation between the PCEs, is the Backward Recursive PCE-based Computation (BRPC), which has been standardized by the IETF in [RFC 5441]. This technique is also based on Inter-PCE cooperation, but works the opposite way as the previous one. The domain sequence may or may not be previously known, and the basis is to build a path tree from the destination domain instead of beginning path computation on the originating domain. Each domain PCE contributes to the tree with a set of optimal exit nodes from its domain. The tree is known as Virtual Shortest Path Tree (VSPT) and the BRPC mechanism is described in [RFC 5441].

However, most of the previous options are not able to obtain optimal paths and, in most of them, the sequence of domains has to be known in order to compute the path. Moreover, the solutions cannot achieve a good scalability, and are suitable only for a very small number of domains.

The hierarchical PCE is an alternative that aims at solving these challenges, and is aligned with ASON and OIF standards. In this scope, the notion of *hierarchy* has several complementary meanings:

- From the point of view of functional entities, a group of entities such as path computation elements may be interconnected in a hierarchical way, with well defined interfaces and functionalities at each level, defining, for example, a tree-like structure.

- From the point of view of the network topology, the concept of hierarchy is based on the idea that a multi-domain network can be seen, from a higher perspective, as a network graph where domains are the new nodes and the domain connectivity can be deduced since the inter-domain links become the new graph links. In fact, this concept is more generic and is based on the fact that graph nodes can be clustered at several hierarchy levels and the cluster topology can be abstracted and aggregated.

Both meanings are related in the sense that, for example, if it is assumed that there is a functional entity per network cluster/domain/group, the segregation and/or partitioning of the network defines the hierarchy of the entities.

The hierarchical PCE architecture is becoming an enabler for end-to-end path computation in the presence of multiple domains with different degrees of involvement: from domain selection exclusively to the actual path computation using some aggregated topology representing the domains.

4.3.3.2 Hierarchical PCE

The generic notion of hierarchical PCE refers to a family of functional architectures where collaborating PCEs are coupled by a hierarchy relationship (such as parent-child). A particular *hierarchical PCE* architecture is defined by the following factors, amongst others:

- The number of levels within the hierarchy. Common approaches just consider a 2-level hierarchy with children and parent nodes (PCEs), and the functionalities of each level (e.g. Traffic engineering and domain selection).

- Whether the children PCE are responsible for Path computation within their domain and/or topology aggregation. For example, a hierarchical PCE architecture may require that a path is computed by domain segments, where each child is responsible for the domain, and there is no topology aggregation. Alternatively, the top PCE may be responsible for path computation using some aggregated topology.

- Whether the parent PCE is responsible for just domain selection or the overall path computation.

- The methods (if any) by which the domains are abstracted: how their nodes and links are synthesized in order to reduce the number of topology elements and allow the architecture to scale with a large number of domains. Common approaches involve summarizing a domain as a set of virtual links (e.g. a mesh between all domain entry/exit node-pairs) or a virtual node. This is, in particular a complicated problem, since not only attributes in terms of bandwidth, delay etc. need to be considered, but also in terms of SRLG, protection capabilities, etc. There may be additional restrictions such as wavelength continuity in optical constraints.

- The methods (if any) by which inter-domain links are taken into account. At a bare minimum, their TE metrics should be taken into account during path selection. Other TE attributes may be considered or simply checked during provisioning.

- Whether parent PCEs are proposed to avoid direct communication between PCEs at the same hierarchy level. This maps any trust model that may be defined where a PCE only trusts other PCEs in a vertical setting.

Of particular interest is the hierarchical PCE architecture proposed by King [PCE-WG]. In this proposal, each domain has at least one PCE that is used for path computation within its domain. This PCE is known as “child PCE” and knows the identity of the neighbor domains. However, this child PCE has knowledge only about its own domain and the links connecting with the neighbor domains, but has no information about the other domains.

In addition, there is a parent PCE that has full visibility of its children PCEs. This visibility implies the topological map of children domains, summarized as domain vertices and links between neighbor domains. The hierarchical PCE has no visibility of the internal details of each domain, meaning it has no knowledge of the topology, internal resources usage or connections availability among domain borders.

The parent PCE builds an inter-domain topology map based on the information provided by the children PCEs, including traffic engineering information about inter-domain links. The parent PCE does not know the detailed internal topology of the domains.

This architecture works in the way shown in Figure 19, with communication between children and parent PCEs in order to compute the end-to-end multi-domain path.

Confidentiality issues

Children PCEs store only information about their own domains and the parent PCE only knows information of inter-domain connections, never knowing the internal details of each individual domain. However, the path response includes the whole nodes sequence involving all domains. Should a higher confidentiality level be required, this complete sequence may be replaced by a *path-key* [RFC 5520] that hides details of each domain segment.

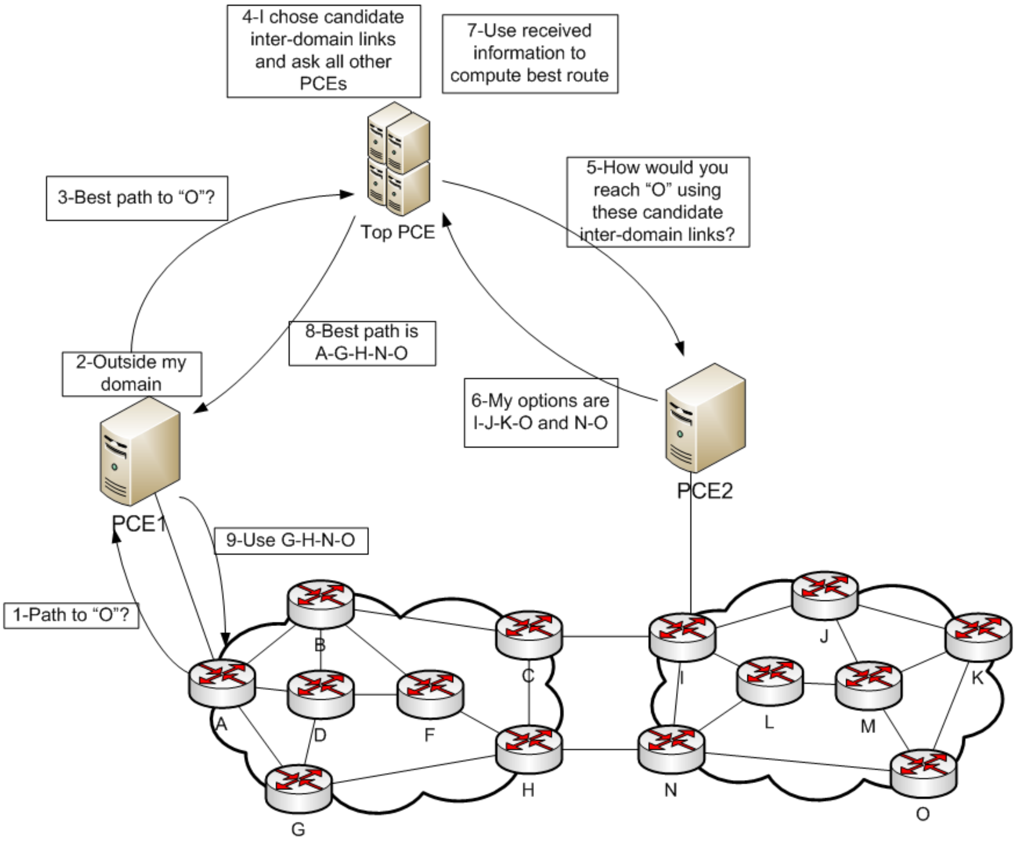

4-I chose candidate inter-domain links and ask all other PCEs
7-Use received information to compute best route
3-Best path to "O"?
Top PCE
5-How would you reach "O" using these candidate inter-domain links?
8-Best path is A-G-H-N-O
2-Outside my domain
6-My options are I-J-K-O and N-O
PCE1
PCE2
9-Use G-H-N-O
1-Path to "O"?
A
B
C
D
F
G
H
I
J
K
L
M
N
O

Figure 19 – Basic hierarchical PCE

4.3.3.3 Full-state umbrella PCE

So far, multi-domain PCE architectures proposed by IETF are based on the idea of limited information exchange among domains, and among parent/children PCEs, because of both confidentiality and scalability. However, in a network operator's scenario, composed of a small set of domains from different providers, these premises may be relaxed.

On the one side, an operator may find desirable the highest level of information exchange among its domains. On the other side, from the manufacturer's point of view, network internal details should not be disseminated. However, from the scalability point of view, having a dedicated umbrella PCE implemented on a high processing capacity machine would allow for path computation with an elevated number of nodes without scalability problems. Therefore, an umbrella PCE or "full state" PCE is a desirable option. It has enough flexibility to perform optimal path computations according to each network operator's needs, as well as perform restoration in a personalized way. Nevertheless, it has a number of unresolved technological issues in a multi-vendor scenario. An umbrella PCE in a mono-vendor environment is almost trivial, mostly reduced to computing scalability.

A definitive issue of the umbrella PCE is computation synchronization: if the umbrella PCE calculates an end-to-end path, it must ensure that each domain PCE's path computations do not interfere with a multi-domain path computation. If this is not taken into account, there could be resources assignment collisions.

This type of multi-domain PCE approach should be further studied, its benefits and scalability proofed and later proposed for standardization.

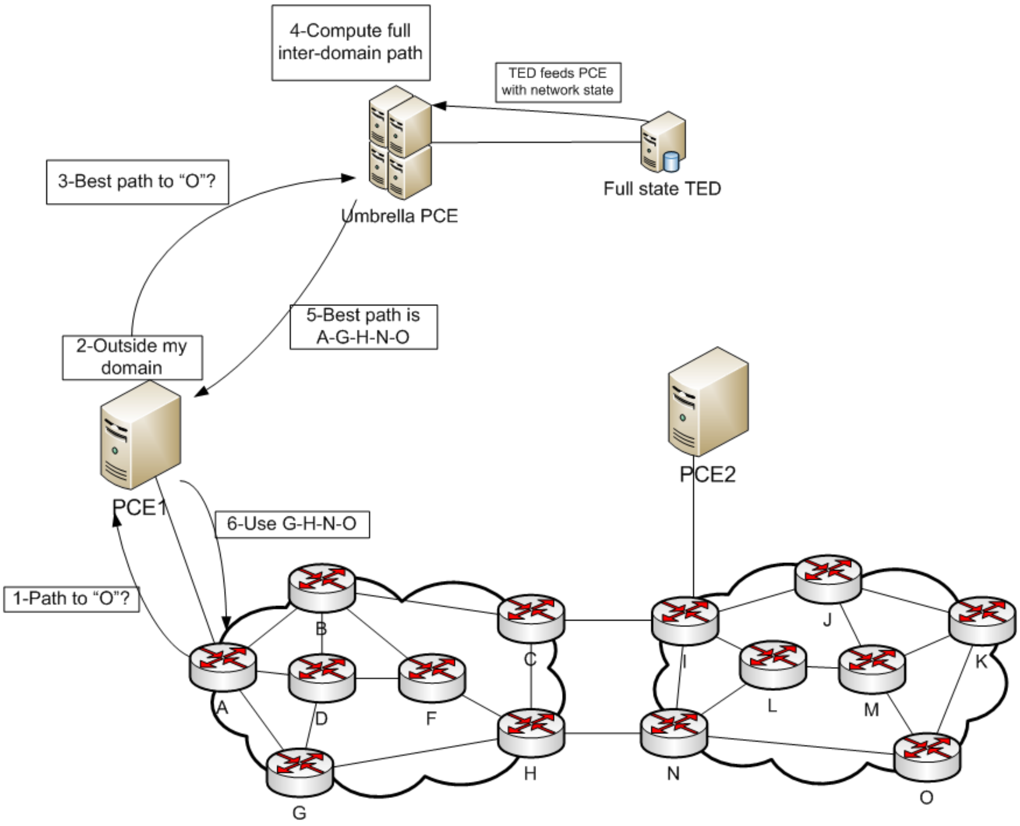


Figure 20 – Inter-domain path computation based on umbrella PCE

4.3.4 Future work

A considerable effort within WP3 will be spent in research work and activities related to the Path Computation Elements, as a differentiating and innovative aspect of STRONGEST. Although the basic functional architecture of the PCE is stable, there is room for improvement in scenarios and deployment models involving new applications of the PCE and new (or extended) collaborative processes, where different PCEs cooperate. This involves the following work items:

- Standardization work, at the IETF, regarding PCE and PCEP extensions for the application of PCE in GMPLS networks, with special emphasis on Wavelength Switched Optical Networks. This work will address the current shortcomings of the PCE/PCEP to address identified requirements.

- Application of the PCE for P2MP connections. Of notable importance in the context of the STRONGEST (in terms of optical lightpaths and P2MP Ethernet / MPLS-TP connections), PCE/PCEP extensions for P2MP connections will be studied, focusing on functional and protocol architectures (aspects such as the optimum tree-computation algorithms are somehow out of scope of WP3).

- PCE Monitoring will cover the mechanisms by which one (or a sequence of) PCE(s) will be monitored, covering both in-band and off-band methods.

- PCE Path Confidentiality and security. Network operators have stringent requirements on topology confidentiality. In this sense, two constraints need to be jointly addressed: path optimality and network topology confidentiality. Enabling Path Keys can indeed preserve confidentiality, but it may not cover all the requirements and be open to improvement.

- Hierarchical PCE and Multi-Domain and Multi-Layer applications of the PCE. The multi-domain aspect is a basic requirement in the considered STRONGEST scenarios. The hierarchical PCE targets the fundamental problem of path computation in such networks. It is a very hot topic currently. The IETF has recently (06/07/2010) published an updated draft on hierarchical PCE, with very rough macroscopic details and preliminary protocol guidelines.

- Hierarchical routing. In addition to specific studies on the Hierarchical PCE, other solutions need to be investigated in the context of either single or multi-carrier networks. They include hierarchical routing based on OIF E-NNI (abstract link, abstract node and star models) possibly implemented together with PCE-based forward or backward schemes for multi-domain path computations.

- TE Metric Abstraction. In the context of multi-domain multi-carrier networks, e.g. exploiting both PCE and hierarchical OIF E-NNI routing, studies will be provided to define the TE Metric to adopt and the related TE Metric Abstraction methods capable of guaranteeing control plane stability, confidentiality and effective network resource utilization.

- PCE TED Creation in GMPLS Multi-Layer Networks. In the context of large-scale multi-layer networks, where multiple PCEs cooperate to perform path computations, procedures and solutions will be investigated in support of the PCE TED creation (e.g., FA-LSP). The goal of the activity is to enable the implementation of a lightweight control plane while providing the PCE with adequate information on network resources (e.g., FA-LSPs) and thus guaranteeing effective TE performance.

Additional effort within WP3 will be dedicated on the performance evaluation and improvement of the GMPLS protocol suite, particularly in the context of large scale-networks. This will include:

- Studying implications for MPLS signaling, where the IGP may need to scale to the order of thousands of routers/sites in a single IGP area.

- Exploring feasibility of sub-second IGP convergence and efficient TE metric distribution.
- Identifying possible problems (e.g. scalability and synchronization) for LDP and RSVP-TE.

4.4 Standard compliance

4.4.1 Compliancy with reference standards

Although at this stage of the project an analysis of the standard compliance of the proposed solutions cannot be complete and exhaustive, some indications can be provided.

The STRONGEST control plane architecture refers to PCE hierarchy, taking as a starting point [RFC4655], [RFC5623] and all PCE-related RFCs. New issues are targeted to be addressed in the context of hierarchical PCE for multi-domain GMPLS networks, so extensions to the current standards are under evaluation.

Standard extensions should be needed especially for the TISPAN RACS integration and WSON impairments in order to perform an efficient multi-region path computation with grooming features.

The control plane architecture proposed within the STRONGEST Project aims to integrate the PCE plus GMPLS architecture from the IETF with the E-NNI hierarchical model from the OIF, taking the best from both worlds. Particularly, a WP3 PCE/PCEP internal task force was defined, with the aim of studying PCE architectures and PCEP protocol for their applicability to GMPLS networks. The studies are mainly focused on Multi-region, multi-layer and multi-domain path computations.

PCEP extensions are also under deployment. Particularly, a first proposal was submitted to the IETF PCE Working Group [PCEP_marg].

End-to-end OAM task and End-to-end services and traffic admittance solutions task will be also impacted by the control plane architecture (especially by the RACS/PCE integration proposal). Standards extensions are still under evaluation in this preliminary phase of the tasks. An OAM WP3 internal task force was defined, with the aim of facing new challenging issues and exploring new solutions for OAM for multi-domain and multi-region networks like the ones depicted in the referring to the reference scenarios.

4.4.2 Evaluation of E-NNI interfaces

External–Network to Network Interface (E-NNI) is the OIF interface interconnecting different domains. Interconnecting domains that could be based on different technologies have different metrics and parameters, etc. E-NNI must act as a kind of “translator point” where all the information that has to cross domains boundaries must be reported in a standard, universally understandable manner.

Due to the hierarchical nature of the interface, that reflects the OIF's hierarchical view of the network, E-NNI approach is considered within the STRONGEST Project as a standard interface for inter-domain communications.

[OIF E-NNI routing] specifies the requirements on and use of OSPF-TE as an E-NNI routing protocol among ASON domains. This routing architecture relies on functional elements for the control plane called Routing Controllers (RCs) that are responsible for the routing information sharing and the construction of virtual topologies.

Main features of this routing architecture are the following ones:

- E-NNI supports a hierarchy of routing instances, where each routing layer operates independently.
- Information among routing layers are exchanged by feeding up/down the adjacent level (different from the relationship between Areas in IGPs).
- It allows the advertisement of Intra-domain Virtual links, Inter-domain links and topology abstraction information.
- It takes no position on the detailed way to represent Virtual/Abstract TE information (e.g., how to deal with link attributes in case of multiple protection schemes).
- Definition of a set of abstract routing (i.e. OSPF-TE) messages and attributes in order to support all the above mentioned features.

Similarly, [OIF E-NNI Signaling] specifies the requirements on and use of RSVP-TE as an E-NNI signaling protocol among ASON domains. This signaling architecture relies on the following functional elements for the control plane:

- Connection Controller (CC): Connection Controller components cooperate to set up connections.
- Calling/Called Party Call Controller (CCC) and Network Call Controller (NCC): Call Controller components cooperate to control the setup, release, and modification of calls. They are relevant to service demarcation points (i.e., CCC is relevant to the client facing side of the UNI; NCCs are relevant to the network facing side of the UNI).
- Protocol Controller (PC): The Protocol Controller maps relevant control component (e.g. CC, CCC, NCC) parameters into messages, which are carried by an implemented protocol to support interconnection over a physical interface. (This includes support for implementations that handle, for example, multiple layers.).

Main features of this signaling architecture are the following ones:

- Support of calls and connections
- Call and connection separation across multiple domains

- Support of Permanent connections (PC) services
- Support of switched connections (SC) services
- Support of soft permanent connections (SPC) services
- Support of hybrid switched connections/ soft permanent connections services
- Definition of a set of abstract signaling (i.e. RSVP-TE) messages and attributes in order to support all the above mentioned features

5 End-to-end services definition

5.1 State of the art

The goal of any transport network is to constitute the infrastructure for network services. The definitions of these services have a major impact on the network technology and design, and they determine many network requirements.

The network services are:

1. **Point-to-point** service which connects two UNIs, e.g. for such service: two branches of an enterprise or a connection between customer and POP.
2. **Multipoint-to-multipoint** which connects many UNIs and allows traffic flow between of them, e.g. L2/L3 VPN.
3. **Rooted-multipoint** which enables a UNI to send information to many UNIs (aka Multicast/broadcast), e.g. broadcast TV.

The services can be L3 services (e.g. IP services, IP VPN), L2 services (e.g. Ethernet ELine/ELAN) or L1 services (e.g. SDH Private Line services, wavelength services).

The services are defined with a set of attributes such as:

1. **The end points** which are the UNIs that are involved in the service.
2. **Bandwidth profile** includes the total bandwidth with its CIR and EIR components.
3. The **QoS** attributes describe the required delay, delay variation and the BER.
4. **Multiplexing and bundling** defines the service delineation. E.g. in Ethernet services, when many services are provided on the same UNI, the VLAN ID is the service identifier.

There are more service attributes which are specific to the service type like Ethernet services, VPLS, emulated services (e.g. CES), wavelength services etc. These services are defined in the different standard bodies and forums (e.g. MEF, OIF, IETF, IEEE etc.).

The service definition should be revised according to the new applications and the customer's services. It is also important to evaluate the service definition according to the expected changes in the networks and especially according to the new STRONGEST architecture and to investigate how principles like router offload will affect the services (e.g. IP services). The service definition has also implications on the control plane (e.g. on signaling).

This chapter raises some of the above mentioned topics and proposes next steps for analysis of service definition in STRONGEST.

5.2 Definition of new end-to-end services and their requirements

According to recent analyses [Forecast], global IP traffic is expected to increase 4.3 times from 2009 to 2014, reaching 63.9 exabytes per month in 2014. Advanced video traffic, including 3D and high-definition TV, will increase 13 times between 2009 and 2014. By 2014, 3D and HD video is forecast to comprise 42 percent of total consumer Internet video traffic.

Such remarkable growth of internet traffic poses a number of significant issues to service providers to provision reliable consumer and business services with the adequate bandwidth and SLA. Advanced technologies and solutions are required to support end-to-end QoS-guaranteed services. Today many service providers prefer to have over capacity rather than having sophisticated QoS mechanism. However, to provide more cost-effective solutions, QoS-guaranteed services will probably have to coexist together with bandwidth-greedy applications on the same network infrastructure.

5.2.1 End-to-end services with strict delay constraints

Advanced network technologies and solutions are particularly required to support end-to-end services with strict delay constraints.

Among business services, trading, financial applications, mobile backhaul and cloud computing represent typical examples of services with strict QoS requirements in terms of reliability and delay-guaranteed content delivery. Traditionally, these services have been implemented on dedicated legacy TDM technologies. However, such technologies are now facing significant scalability issues and new solutions are required.

Within both business and consumer applications, a number of services requiring strict delay constraints are rapidly emerging. They can be considered as the evolution of the currently available low-quality videoconferencing services which so far have not completely satisfied customer expectations. On the other hand, services exploiting interactive high-definition video with stereophonic and surround sounds are expected to improve the sense of reality and significantly increase their utilization. Examples of such services include 3D-telepresence, telemedicine (e.g., remote surgery and diagnosis), education, applications in hazardous situations (e.g., military operations, toxic atmospheres) and services for entertainment (e.g., gaming).

Other delay-critical applications include the delivery of information and contents. In fact, great competition between content service providers is already in place to provide the best performance also from a technical point of view. Gambling and “googling” represent two different types of high-revenue services which strongly require minimum end-to-end delay.

The provisioning of these high-quality services is expected to require not only high bandwidth, availability, reliability and null loss rate, but also guaranteed and deterministic end-to-end delay and jitter which have to be minimized across the network infrastructure.

This opens the debate on the introduction of additional features within the network nodes and the control plane. For example, delay parameters might be considered as possible multi-domain TE metrics or constraints to be used in path computations. PCE

Architecture might require specific additional extensions such as novel objective functions and extensions in the PCEP protocol specifications. Routing protocols might be enhanced both in intra-domain and in inter-domain advertisement (i.e., OSPF-TE Opaque LSA and OIF E-NNI respectively). Signaling procedures might also require specific extensions (e.g., within the SESSION ATTRIBUTE object). The possible evolution and configuration of routing and signaling protocols has to be provided while preserving the adequate level of security and confidentiality across domains controlled by different carriers.

Additional procedures, tools and enhanced OAM functionalities might also be required to retrieve and manage detailed information, for example to precisely measure one-way and two-way delay between pairs of end-points. This could be required to verify the provided SLA during service setup and operation as part of an integrated set of delay-based OAM functionalities and management functions. Additional details on OAM functions and open issues are addressed in Chapter 3.

5.2.2 Multi-domain end-to-end service definition

In some cases, network services traverse locations that are not all served by a single carrier. In order to run the service, multiple carriers will need to support the service. It is also possible that the carriers that have the UNIs do not connect each other and a transit carrier is needed also to build the service end-to-end. This situation poses further potential complication to support the service definitions as basically were defined for single carrier scenario.

The goal of this subsection is to identify the multi domain and multi carrier aspects in service definition.

The service provider who provides the service to the end user selects and contracts with various carriers to deliver the UNI-to-UNI services. It is the responsibility of the service provider to ensure that the appropriate service and interface attribute values from each Operator are such that the UNI to UNI service features specified in the SLA will be fulfilled.

Multi domain networking requires detailed definition of the services over external network-to-network interface (E-NNI) and the Operator Services Attributes. Such attributes can exist between UNIs, between E-NNIs, and between a UNI and an E-NNI, this subsection will be focused on the E-NNI attributes. Like in UNI where the attributes are the SLA between the subscriber and the service provider, the E-NNI attributes are the SLA between the carriers.

It is required that the UNI to UNI service observed by the subscriber will be the same where a single carrier is involved or where multiple carriers are involved.

The MEF service definition is widely adopted by the industry for Ethernet service definition. This definition can be relevant also for all packet based services (e.g. IP, MPLS). This subsection is based on the MEF service definition.

5.2.2.1 Multi-domain and multi-carrier players

The multi-domain and multi-carrier players are:

- **Customer:** The customer is the consumer of a service provided by a service provider.

- **Service provider:** The service provider offers services to the customers. The service provider is responsible for composition and setup of the service, while managing all aspects of the service, which may involve one or more operators (i.e. carriers). For a given Service, the Customer contracts with a service provider that is responsible for delivering the service bounded by the UNIs associated with the service.

- **Operator (carrier):** An operator is a network operator who participates in the delivery of a service by providing a segment (or segments) of the service. An operator operates one or several domains in multi domain services.

5.2.2.2 The service attributes

As mentioned above, the MEF service definition can be taken as a reference. [MEF 6.1] defines the Ethernet Service Types (EPL, EVPL, E-Line, E-LAN, E-Tree, etc) and MEF 10.2 defines the service attributes and parameters required to offer the services.

The main service attribute are:

- Delivery type: broadcast, multicast and unicast (unlearned and learned).

- Policy for delivery of control protocols (e.g. BPDU).

- Service identification and delineation (e.g. VLAN delineation UNI delineation, service multiplexing, service bundling etc.).

- Bandwidth and QoS attributes (e.g. committed information rate (CIR), excessive information rate (EIR), rate enforcement - shaping and policing, burst size, latency, delay variation, frame-loss).

5.2.2.3 The E-NNI service attributes

In case of Inter domain service, the attributes of the service should align with the domains’ interconnections (E-NNIs). Following are the E-NNI service attributes as defined by [MEF 26]:

- Frame format (format of the PDUs at the E-NNI).

- E-NNI maximum transmission unit size (maximum length of E-NNI frame in bytes allowed at the E-NNI).

5.2.2.4 Multi-domain service attributes

In multi domain scenario, the customer purchases a service from the service provider and specifies the service according to the attributes that are mentioned in 5.2.2.2. The service provider builds the service from segments that he buys from other operators, each segment being provided in a different domain (the service provider can communicate with the other operators directly or via neutral facilities like the emerging Ethernet exchanges). The service segments that are sold by the service provider have service attributes that are described by service templates. The operators publish these templates to the potential service providers (can be through market places like Ethernet exchange).

Table 3 – Example for simple connectivity service template

Parameter	**Description**
Template number	Template number is globally unique and is consisting of a prefix (i.e. domain number) and a suffix allocated by the domain manager
Template expiration	Template's expiration date
Connection end points	The template offers service between connection points (E-NNIs)
Offer #1 *[The provider may offer several different transit services with different attributes]*	
Max delay	Important in real time services
Max delay variation	Important in real time services
Maximum packet loss	
Max PBS	Maximum allowed burst size for peak information rate
Max CBS	Maximum allowed burst size for committed information rate
Overbooking factor (OBF)	Over booking factor describes the amount of EIR BW that can be sold to the customer above the links capacity EIR≤(link –CIR)*OBF
Service availability	
Mean time to repair	
Minimum duration	Minimum duration for an active service
Statistic reporting - periodical	
Statistic reporting – on demand	

CIR BW Price per 1Mbps EIR BW Price per 1Mbps	Price per Mbps The price depends among others on: • BW • Protection o Protected service price o Unprotected service price • Service Duration (in some cases it affects the price)

Templates contain information about the service types that are provided by the domains; the templates describe the service types, their attributes and their prices. The templates allow very flexible service creation; an operator can publish the service attributes like delay, delay variation and frame loss. Following is an example of the basic template for basic connectivity service provided by an operator.

The attributes of the end-to-end service are calculated from the segments' attributes, e.g.:

- The end-to-end price that the service provider should pay for the service is the sum of the segments prices.
- The end-to-end delay is calculated from the segments delay.
- The end-to-end delay variation is calculated from the segments' delay variation.
- The service packet loss is calculated from the segments' packet loss.

5.2.3 End-to-end multicast and grooming

5.2.3.1 Data applications that employ end-to-end multicast

There are plenty of data application types which might require point to multi-point connections. Most obvious benefit of employing multicast can be found in following classes of applications:

- **GRID computing**: Grid computing is an emerging technology that provides seamless access to computing power and data storage capacity distributed over the globe. Grid computing (or the use of computational grids) is the application of several computers to a single problem at the same time -- usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data. Grid computing depends on software to divide and apportion pieces of a program among several computers, sometimes up to many thousands. Grid computing can also be thought of as distributed and large-scale cluster computing, as well as a form of network-distributed parallel processing.

- **Videoconferencing**: Videoconferencing is a set of interactive telecommunication technologies which allow two or more locations to interact via two-way video and audio transmissions simultaneously. Videoconferencing uses telecommunications of audio and video to bring people at different sites together for a meeting. This can be as simple as a conversation between two people in private offices (point-to-point) or involve several sites (multi-point) with more than one person in large rooms at different sites. Besides the audio and visual transmission of meeting activities, videoconferencing can be used to share documents, computer-displayed information, and whiteboards.

- **Video-on-demand (VoD)**: Video on demand systems enable TV viewers to access the programs that they want to see, when they want to see them. True video on demand systems usually store the programs or films on a server located on or near the premises. Near-video on demand systems work slightly differently, playing the same programs on a regular basis and enabling viewers to access them the next time they are on-line.

- **High-definition television (HDTV)**: High-definition television refers to video having resolution substantially higher than traditional television systems (standard-definition TV, or SDTV). HD has one or two million pixels per frame; roughly five times that of SD. Early HDTV broadcasting used analogue techniques, but today HDTV is digitally broadcast using video compression.

- **Multimedia document distribution**: Multimedia document distribution is the distribution of an electronic document that incorporates interactive material from a variety of different media such as text, video, sound, graphics, and animation. Such documents can be viewed on a multimedia computer or transmitted via the Internet Interactive distance learning. Multimedia document distribution services aim to overcome interoperability problem between different systems in order to exchange multimedia information.

- **Interactive distance learning**: Interactive Distance Learning entails transferring knowledge to multiple locations for purposes of education, then tracking and responding to the resulting activity. It often includes video for e-learning and distance learning. Educational institutions as well as businesses use Interactive distance learning.

- **Live auctions**: Live auctions are auctions performed over the Internet. The internet age has transformed auction into a truly open process in which thousands of goods (from books to ships) and services (from air travel to legal advice) may be offered for bidding by anyone from anywhere and at any time on websites such as eBay.com.

- **Optical storage area networks (OSAN)**: Optical Storage area networks (OSANs) are a promising technology to efficiently manage the ever-increasing amount of business data. Extending SANs over large distances becomes essential to facilitate data protection and sharing storage resources over large geographic distances. The optical storage area networks entail the extension of SAN into the synchronous optical Network (SONET) and over wavelength division multiplexing (WDM).

5.2.3.2 Issues and optimal multicasting of light-trees in WSON

Use of multicast data streams as opposed to unicast raises certain list of problems related both to control plane and data plane.

In order to realize all-optical multicast operations on the data plane, the light path concept has been generalized to that of light-tree, which consists of an all-optical point-to-multipoint channel originated at any source and has more than one destination node. Light-tree-based virtual topology requires fewer hops and fewer optoelectronic components than a light-path-based virtual topology does. Because of their ability to provide “point-to-multipoint” connections, as required by the multicast applications, light-trees are well suited for carrying multicast traffic. For example, if we wish to establish a multicast session from source node to a set of destination nodes, this will require a “point-to-multipoint” connection from source node to every destination node. Such a session can be established using a light-tree approach, with the source node as root and the destination nodes as leaves. In a network equipped with all-optical switches, an optical splitter is needed at node to replicate the incoming bit stream into multiple copies. In the absence of wavelength converters in a network, this light-tree based multicast session will exhibit the wavelength-continuity constraint.

The major concern from the control plane perspective is communication cost minimization. For this purpose it is important to establish a session along minimum weight links. The cost of a multicast session is the sum of the weights of the links occupied by the multicast session. It is desirable to minimize the cost of establishing a light-tree so that sufficient resources are available for other connections. Each multicast session might occupy a wavelength channel along the fiber links. The cost of carrying traffic between adjacent nodes – which may be the fiber distance or cost of amplifiers/regenerators on the link – is represented as weights. To minimize communication cost, it is important to establish a session along minimum weight links. The cost of a multicast session is the sum of the weights of the links occupied by the multicast session. It is desirable to minimize the cost of establishing a light-tree so that sufficient resources are available for other connections. Such a minimum-cost multicast tree is called a minimum-cost Steiner tree, and the problem of finding a Steiner minimum tree in a graph is NP-complete. A light-tree is a directed Steiner tree and the problem of finding a directed Steiner minimum tree is NP-complete, which follows from the fact that its special-case Steiner minimum tree is NP-complete. Unfortunately, no multicast routing protocol today is able to maintain such an optimal tree. Different multicast protocols will therefore, in general, generate different trees.

5.2.3.3 Traffic grooming in WSON

Traffic grooming is the field of study that is concerned with the development of algorithms and protocols for the design, operation, and control of networks with multi-granular bandwidth demands. The objective of traffic grooming techniques is to ensure that sub-wavelength traffic components are transported over the network in an efficient and cost-effective manner. Traffic grooming research has, in general, followed one of two directions. In dynamic grooming it is assumed that the node grooming capabilities in terms of available electronic ports, level of wavelength conversion and switching capacity are fixed and known, and the goal is to develop online algorithms for grooming and routing of connection requests that arrive in real time. Typical solution approaches transform the grooming problem into a shortest path problem on a new layered graph modeling both the underlying physical topology and the capabilities of individual nodes.

In static grooming, the starting point is the set of (forecast) long-term traffic demands, and the objective is to provision the network nodes to carry all the demands while minimizing the overall network cost. The cost metric frequently considered in the literature is the total number of electronic ports required to originate and terminate the lightpaths created to carry the traffic components. As backbone networks migrate from ring to mesh topologies, traffic grooming in general topology networks is becoming the subject of an increasing number of studies. Most studies provide an integer linear programming (ILP) formulation as the basis for reasoning about and tackling the problem. Unfortunately, solving the ILP directly does not scale to instances with more than a handful of nodes, and consequently it cannot be applied to networks of practical size covering a national or international geographical area. Consequently, either the integer linear programming is tackled using standard relaxation techniques, or the problem is decomposed into sub-problems which are solved using heuristics.

5.2.4 Wholesale QoS domains

The concept of Wholesale QoS domains is complementing the paradigm of exclusive capacity reservation for QoS provisioning. We introduce the new concept due to the widespread perception that a reasonable part of the source traffic is too fragmented to be applicable for full featured end-to-end reservation mechanisms. Anyway we agree that many applications of that class are claiming better quality than simple best effort. Today, indeed there are network installations that are using the best effort IP technology but nevertheless do provide nearly guaranteed performance. These are special purpose networks, apart from the Internet, with controlled access by a dedicated set of applications. In the global network picture we are talking about sub-networks that are using dedicated and exclusively reserved resources from other networks. Most of the today’s bandwidth reservation business is dedicated to this kind of sub-networking: VPNs, digital leased line service, λ-services, etc.

The concept of a Wholesale QoS domain generalizes the sub-networking paradigm in several directions. It provides capacity dimensioning rules depending on the actual traffic. This way it allows for automated capacity reservation requests. And, it opens the concept of special purpose network domains to interconnected networks without losing the control over the traffic profile.

5.2.4.1 Co-existence of services on shared resources

The idea is simple: aggregated traffic that fluctuates well below a given capacity limit does not exhibit considerable losses. Buffers in network nodes are used to absorb contention between simultaneously arriving packets (on different interfaces), but not more. In particular, buffers should not resolve congestion beyond the timescale of a few milliseconds. Queues, if not empty at all, are short and the corresponding packet delay remains quite limited. In such circumstances the mixture of services does not matter. Everything is served on time up to a remaining statistical uncertainty. The crucial question in the sketched idea is an appropriate measure for “well below the capacity limit”. This measure depends not only on the traffic load but also on the volatility of traffic.

Figure 21 illustrates the case where the fluctuating traffic (1Gbps in average) sometimes hits the capacity limit of the link (1.2Gbps). In those cases the queue at input to the link is quickly filling up, which initially manifests itself in a sudden increase of the latency (queuing delay), and next in a burst of packet losses (queue overflow).

Obviously, in Figure 21, given a slightly higher capacity, the latency bursts and packet losses would disappear. The initial target of low jitter and nearly lossless co-existence of packet services on a shared link could be reached.

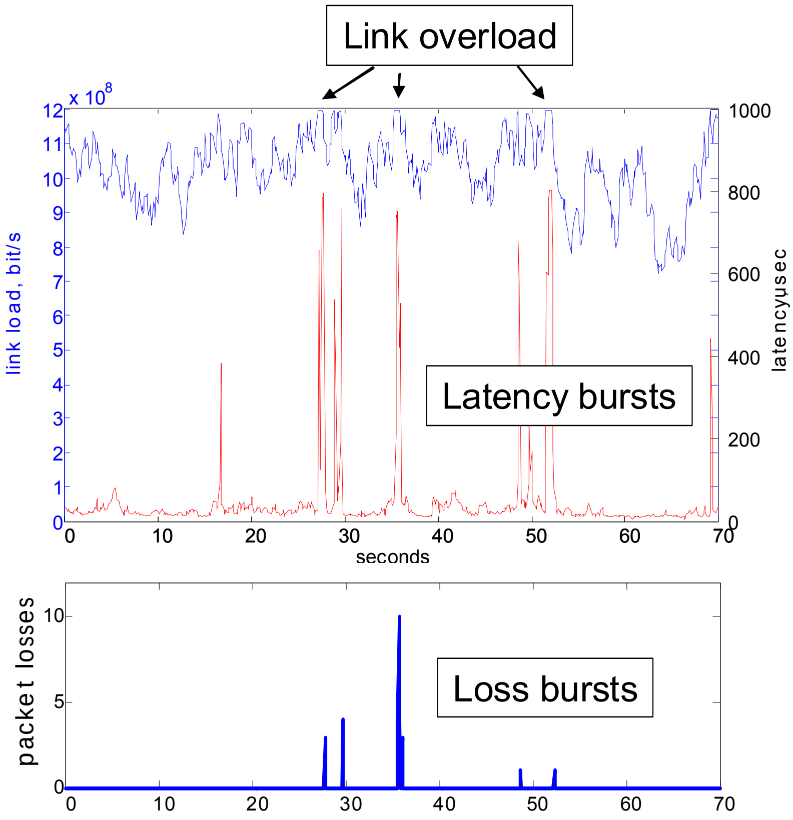


Figure 21 – Link overload, latency bursts and packet loss are tied together

5.2.4.2 Prediction of traffic and its volatility

In previous work [Lau08] we followed the assumption of aggregated traffic as an overlay of randomly arriving and terminating (short lived) application streams. In this model the average number of contributing application streams determines the volatility of the aggregated load. In short, aggregated traffic of only a few high bit rate streams fluctuates much more than similar traffic of many narrow band streams. The probability distribution of the fluctuations is well defined by just two parameters: The mean traffic load and the (mean) bit rate of contributing application streams. By knowing the both we can calculate an appropriate capacity that the traffic fluctuations will not exceed by any desired probability.

According to the model, both parameters are rather invariant, in opposite to the fluctuations themselves, which are unpredictable. The reason is the statistical assumption of a large population of independent end users, who are creating the aggregated traffic, each one only contributing a small fraction. It is unlikely that a large user population changes its mean activity all at once. Equally unlikely all members will change their favorite applications all at once, and hence their typical application stream bit rates.

Finally, traffic load and application stream bit rate are convenient for cascaded traffic aggregation. On the way from access to core the mean traffic load contributions simply sum up, whereas the application stream bit rates remain invariant, in particular independent of the actual transport bit rates.

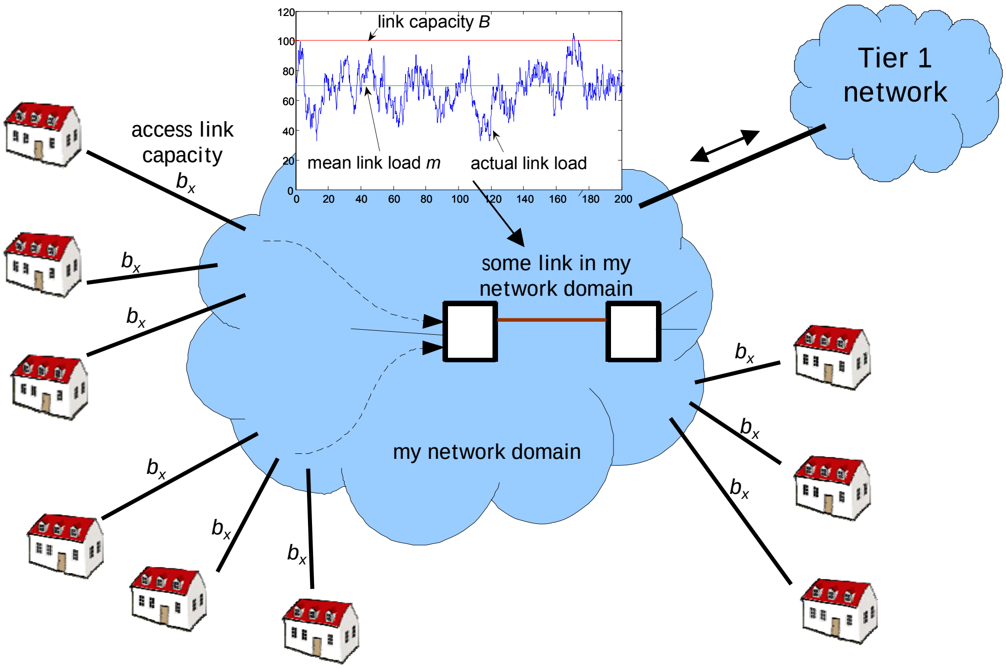


Figure 22 – Network without transit traffic – all application streams are shaped by the access link capacity b_x

The remaining open question is how to quantify the application stream bit rate. In plain packet networks the application streams are neither explicitly declared nor are they associated with a fixed bandwidth. In [Lau8] we argued that the end users access link capacity is a natural limit that application streams cannot exceed. A network operator, by knowing the installed base, can rely on this traffic shaping effect as a worst case limit. In the example of Figure 22, all application stream bit rates are not bigger than b_x. By monitoring the evolution of the mean load m on some network link, the operator can calculate an appropriate link capacity B, which the actual traffic fluctuations will not exceed. Sample graphs of the function $B=f(m,b_x,P_{loss})$ are given in Figure 23.

5.2.4.3 The “wholesale” generalization

The link dimensioning according to Figure 22 is still limited by a number of restrictions. (i) Access link capacities should be uniform throughout the network domain. (ii) Access link capacities should be sufficiently narrow band, so that they exhibit the bottleneck for the majority of application streams. Otherwise the worst case assumption still would be true but would be much too pessimistic. (iii) Connections to other network domains must be restricted to locally terminated application streams. The method is not applicable to transit traffic, where neither the sending nor the receiving end of connections is under control of the transit network.

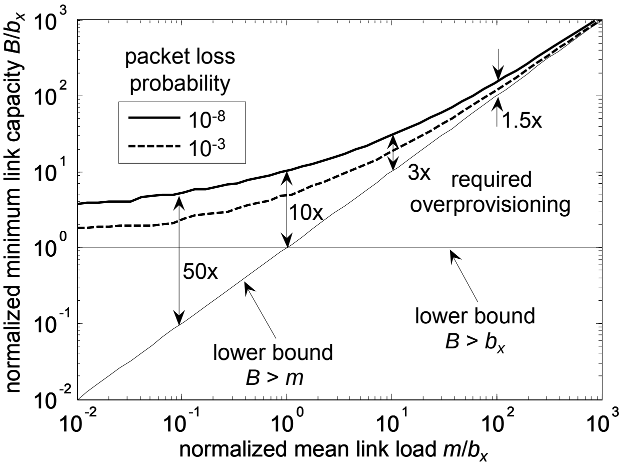


Figure 23 – Capacity *B* as function of load *m* and access stream bit rate b_x

In scope of STRONGEST following ideas will be investigated to overcome the mentioned restrictions:

1) Signaling of effective access link capacities along collection/distribution trees. In such cascaded traffic aggregation schemas it could be the case that traffic contributions with different b_x declarations are merged. The resulting aggregate would fluctuate roughly according to the cumulative load and according to the larger of the b_x values of the contributions. A more precise solution could take into account the relative weight of the contributions and calculate a resulting “effective” b_x of the aggregate traffic. However, this value would not be a declarative (constant) parameter anymore but a slowly changing value similarly as the traffic load itself. It should be investigated, if the required signaling overhead, to hand over the b_x values, is worth the achievable gain.

2) Measurement of the effective application stream bit rate (b_x). We could show in first experiments that by observation of the traffic load in short time intervals (millisecond range) the original application stream bit rate can be deduced.

3) Policing of traffic injections that fluctuate higher than declared and agreed between network domains.

The first proposal is intended for a more realistic treatment of a single provider aggregation network with a variety of different end-user contracts, but still in one homogeneous trust domain. The last two ideas open the way to an Internet of autonomous wholesale QoS domains. With the measurement and policing functions at the borders between administrative domains, the particular operators can be made liable for the fluctuations they inject into neighboring domains. Figure 24 shows a wholesale QoS domain that receives transit traffic from other domains (A and B).

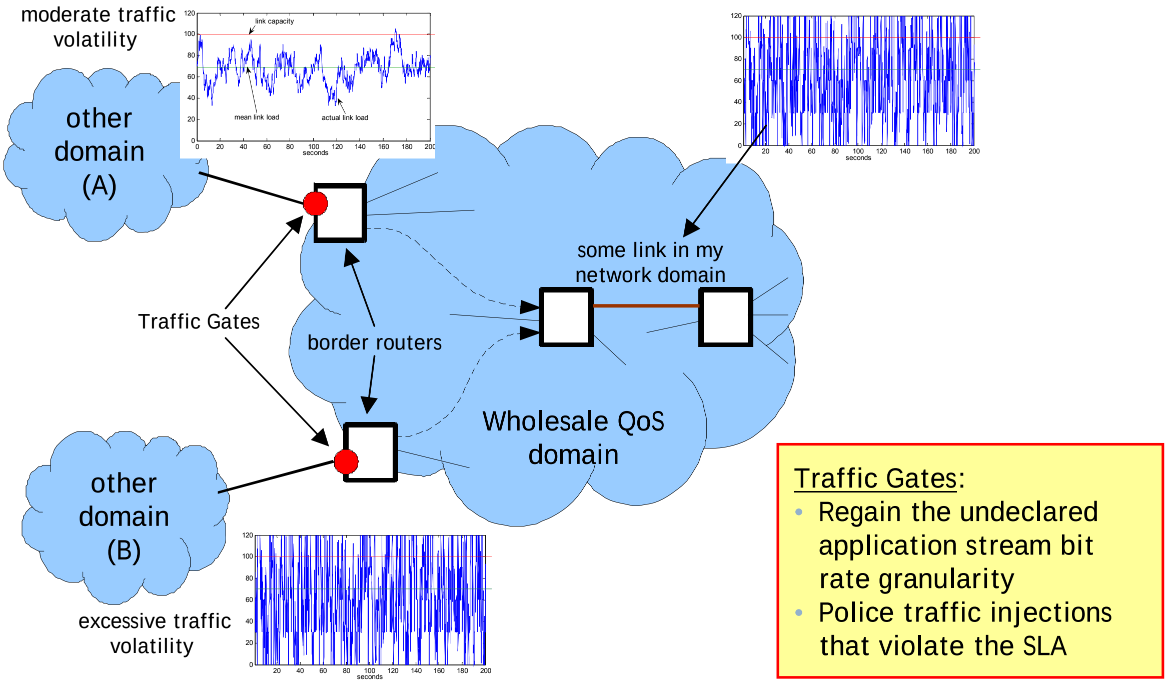

moderate traffic volatility
link capacity
mean link load
actual link load
seconds
other domain (A)
some link in my network domain
Traffic Gates
border routers
Wholesale QoS domain
other domain (B)
excessive traffic volatility
Traffic Gates:
Regain the undeclared application stream bit rate granularity
Police traffic injections that violate the SLA

Figure 24 – Generalized wholesale QoS domain: transit traffic volatility can be measured and policed according to mutual service level agreements (SLA)

Traffic gates at domain ingress measure the bit rate granularity. Due to the invariance of application stream bit rates along the path, the network operator is able to predict upper bounds of the traffic volatility on internal links and dimension them accordingly. Given there are Service Level Agreements (SLA) in place that regulate the granularity of border crossing traffic, the traffic gates could also police SLA violations in order to protect other users from the adversarial effects of excessively high volatile traffic injections.

In scope of WP3 we particularly investigate the management of similar wholesale QoS domains. Questions will be answered like propagation of SLA parameters to the border gates, collection and distribution of measurement data from the border gates, or adaptation of link capacities according to the measurements, appropriate signaling to lower (TE enabled) network layers, etc.

5.3 End-to-end services for STRONGEST

5.3.1 Mapping of end-to-end services into reference scenarios

In the context of STRONGEST Project, a separation between service provider Infrastructure and Network Provider Infrastructure was considered.

A service provider is an organization that provides some kind of communications service, storage service or processing service or any combination of the three. Examples are a local or long distance telephone company, Internet service provider (ISP), application service provider (ASP) and storage service provider (SSP).

A network provider is an organization that maintains and operates the network components, allows network functionality to be distributed flexibly and allows the architecture to be modified to control the services. A network provider may also take more than one role, e.g., also acting as service provider.

From the service provider side, a preliminary set of *Application Classes (ACs)* were identified, as described in Table 4.

Similarly, from the Network Provider side, a preliminary set of *Network Service Types (NSTs)* were also analyzed and discussed, as well as the mapping of ACs into NSTs.

When an end-to-end service request comes, its Application Class is determined, and then it is mapped into a Network Service Type. As a consequence, when the control plane serves the request, this is done by providing an end-to-end path satisfying the requirements of the corresponding Network Service Type. Figure 25 shows an example of this concepts referring to control plane reference scenario 2.

Table 4 – Application services

Type of application	Description	Examples	Typical performance requirements
Class #1 (interactive)	Apps interested in a fixed rate with tight requirements on latency and loss. Typically intolerant to any loss and high delay.	Voice, interactive media, circuit emulation, video conference, high quality storage services, broadcast TV	Guaranteed min rate, guaranteed max delay, very low jitter, no loss
Class #2 (best effort)	Apps interested in the shortest time to completion, but that can cope with any rate that achieves that (TCP-like rate adaptation, Internet traffic, web, email, file transfer, telnet)	Web browsing, Peer-to-peer file exchange	Minimum time-to-completion
Class #3 (guaranteed)	Apps interested in a small range of rates (typically min-max range) with some (possibly complex) requirements on latency and loss. Tolerant to some small amounts of delay or loss	Media streaming, business VPNs, massive business back-up	Low loss, high rate within traffic envelope

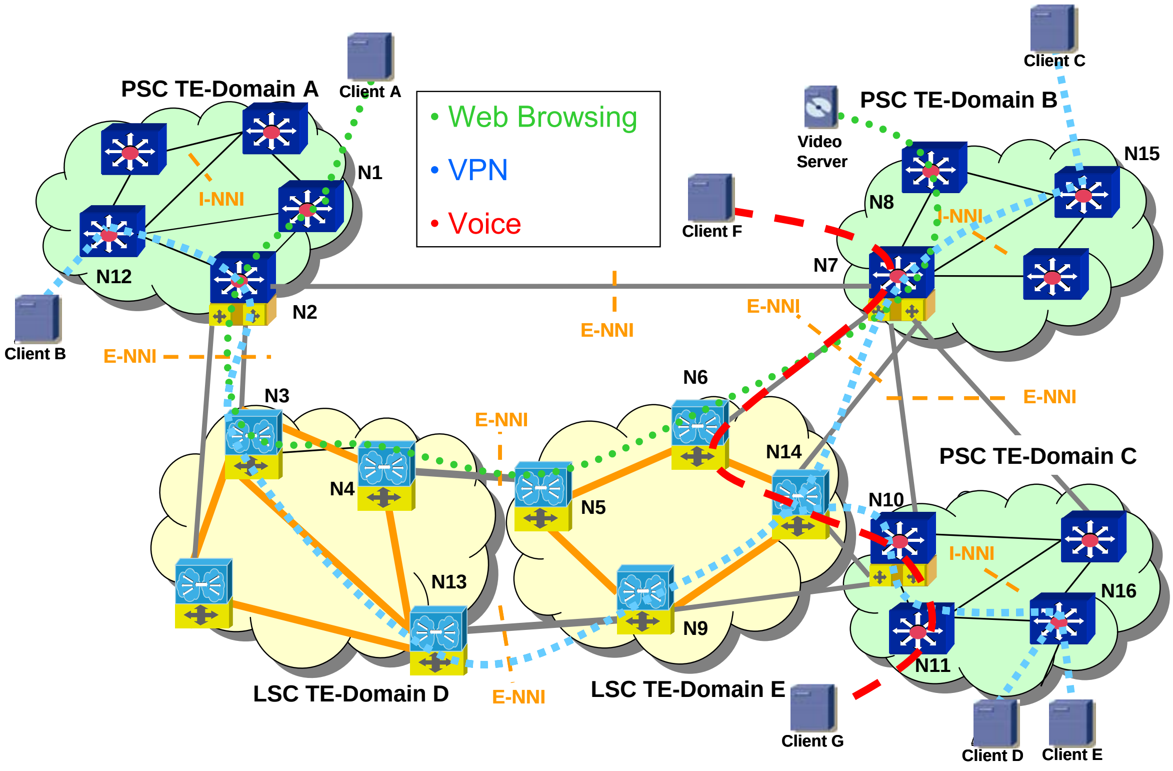


Figure 25 – End-to-end service mapping into reference scenario 2

A connection for a web browsing application is requested from a client in the PSC TE-Domain A to a video server in the PSC TE-Domain B. The request is mapped as a Class #2 application and then mapped into the corresponding Network Service Type. As a consequence, both optical and packet LSPs will be selected/set up according to the considered Network Service Type requirements (i.e. delay, bandwidth, etc.). In the example the following path is selected: N1-N2-N3-N4-N5-N6-N7-N8, spanning PSC TE-Domain A, LSC Te-Domain D, LSC TE-Domain E and PSC TE-Domain B.

A second connection is then requested, this time for a VPN connecting Client B in PSC TE-Domain A, Client C in PSC TE-Domain B and Client D and Client E in LSC TE-Domain C. The request is mapped as a Class #3 application and then mapped into the corresponding Network Service Type. As a consequence, both optical and packet LSPs will be selected/set up according to the considered Network Service Type requirements (i.e. delay, bandwidth, etc.). In the example Client B, Client C, Client D and Client E are connected through nodes N12, N2, N3, N13, N9, N14, N7, N15, N10, N11 and N16, spanning all the TE-Domains composing the network (both PSC and LSC ones). Network Service types are different, so LSPs with different characteristics are selected / set up. Anyway, if there is a portion of the networks (i.e. LSPs or segment LSPs) that satisfies both classes' requirements, grooming of higher level LSPs is possible (i.e. PSC LSPs corresponding to different services can be nested in the same LSC LSP, having the needed characteristics).

A third connection is then requested, in order to carry a voice application between Client F in the PSC TE-Domain B to Client G in the PSC TE-Domain C. The request is mapped as a Class #1 application and then mapped into the corresponding Network Service Type. As a consequence, both optical and packet LSPs will be selected/set up according to the considered Network Service Type requirements (i.e. delay, bandwidth, etc.). In the example the following path is selected: N7-N6-N14-N10-N11, spanning PSC TE-Domain B, LSC TE-Domain E and PSC TE-Domain C.

Again, if there is a portion of the networks (i.e. LSPs or segment LSPs) that satisfies both classes' requirements, grooming of higher level LSPs is possible (i.e. PSC LSPs corresponding to different services can be nested in the same LSC LSP, having the needed characteristics).

Grooming lower valued application classes into higher valued network types, could be not economical (high valued resources are provided to services that doesn't need them), but it could be needed/useful in several cases:

- Lower valued LSPs can be considered as pre-emptable (so that if higher valued resources are available for new requested higher valued LSPs), allowing lower valued LSPs to be used as backup paths also for higher valued LSPs.
- Lower valued end-to-end services (that can be a relevant percentage of the total ones) can be accommodated also in case of lack of corresponding network resources. That can happen also for a limited time, waiting that new lower valued optical LSPs are set up or that already existing ones will free some resources.

5.3.2 Preliminary considerations on signaling

Working on complex and heterogeneous networks, as the ones considered in the STRONGEST Reference Scenarios, means dealing with LSPs of different technologies (MRN/MLN networks) that are hierarchically structured and have also different bandwidth granularity.

Obviously, the considered end-to-end services, as well as the considered architecture (i.e. LSP hierarchy structure, as well as the bandwidth granularity) impacts the end-to-end services signaling. As a matter of fact, end-to-end services, divided in differently valued application classes, should be served setting up different sets of end-to-end connections, spanning different regions and domains, reflecting the corresponding network service types.

In this context, the following set of preliminary consideration on signaling should be made, in order to evaluate how the considered control plane architecture should perform a smart signaling for the considered end-to-end services. In order to avoid wasting of resources, control plane should be able to optimize their utilization, especially high value ones, such as wavelength on a fiber.

In WSON domains, as the ones considered in the STRONGEST Reference Scenarios, lightpaths are valuable resources. As a matter of fact, setting up a lightpath requires dealing with the Routing and Wavelength Assignment constraint and is also time expensive for the reconfigurations of optical cross-connections. Therefore, setting up/tearing down an optical LSP should be done carefully and possibly its resources (e.g. bandwidth) should be completely exploited. In order to achieve this target, end-to-end traffic should be routed considering the already set up optical LSPs and/or the pre-configured ones. The latter should be filled before creating new ones. In hierarchical architectures, such as the one considered for STRONGEST, this can be achieved by grooming of higher layer LSPs into lower layer ones, having more bandwidth.

Nodes with add/drop functionalities should be aware of the total load of the lower layer LSPs and path computation should evaluate the opportunity of aggregating higher layer LSPs into already set up and partially free lower layer LSPs as priority choice. In other words, the filling of an optical LSP should be a parameter to be reflected into path computation constraints. Also, the dynamicity of the traffic should be considered in order to minimize, where possible, the number of signaling information. Less dynamic traffic should be aggregated into the same optical LSPs that would result in more stable optical connections.

Moreover, small and frequent changes should not be advertised until some parameters will cross some threshold values. Bandwidth modifications should be also considered in a weighted way, without advertise small and frequent changes until some parameters will cross some threshold values. Analysis on the opportunity to re-optimize the network periodically should also be considered.

Analysis topics within STRONGEST project will cope with the following issues:

- Aggregate fluxes of different granularities in a single wavelength.
- Modeling WSON muxponders in order to perform grooming.
- Reducing the RSVP-TE sessions where possible in order to improve scalability.
- Find an optimal trade-off between too detailed and too poor signaling information.
- Analysis on setup time limitation due to the serialized procedures, especially where more layers are involved (i.e. if there is a need to create FAs for N layers, FA-LSP setup and corresponding TE-Link advertisement must be repeated N times for each of the N involved layers).
- Advertisement of different signaling sessions between hierarchical layers and attribute inheritance.

5.3.3 QoS and admission control for end-to-end services

QoS and admission control can be offered to end-to-end services through the control layer. As stated in Section 4.1.1, in the context of the STRONGEST Project an extension of the ETSI TISPAN RACS [RACS] and IETF PCE architecture (called G-RACS and fully described in Section 4.3.1) will be considered as the reference architecture for the control layer.

RACS based control layer allows applications to ask for a set of resources, without having to deal with transport network details (e.g. used transport technology, involved domains, etc.). Using control layer a logical separation between application services and transport layer services is achieved; a single application request can generate different network requests and can involve different kinds of networks and domains.

RACS control layer is able to handle resources both in the access (A-RACF) and core (C-RACF) network, enabling a real control of the whole end-to-end path and giving to the applications more guarantees about end-to-end QoS[3]. Moreover, control layer can optimize network resources, mapping new application requests into existing network paths with enough available resources. For example, if a 100 Mb network path is established in order to satisfy a 60 Mb application request, 40 Mb remains free and can be used for another subsequent 40 Mb application request without the need to setup a new network path. However, in the evaluation process of allocating resources other parameters could be considered. As an example, if a criteria is based on the price of the resource and the new service needs less expensive transport service (due to less QoS requirements), then maybe it could be better to setup another path. In general, many criteria could be used, based on either operator defined policies or network resources availability.

Control layer is therefore able to receive QoS application requests, to identify involved networks and domains, to interact with them in order to perform admission control and resource reservations. If necessary, new paths are created in order to satisfy the application requests.

In the next paragraph, an example explains how the control layer can guarantee end-to-end QoS to an application service.

5.3.3.1 End to end QoS using control layer: an example

In this paragraph an example of how control layer is able to handle an end-to-end QoS request coming from a service application is presented. In order to better fit the example in the STRONGEST project, scenario 2 (multi-domain / multi-region / single-carrier) has been selected as the reference scenario.

[3] The A-RACF is deployed in the access network domain and requires the provisioning of the transport resources on a per subscriber basis. The C-RACF is deployed in the core transport network domain, which doesn’t provision the transport resources on a per subscriber basis.

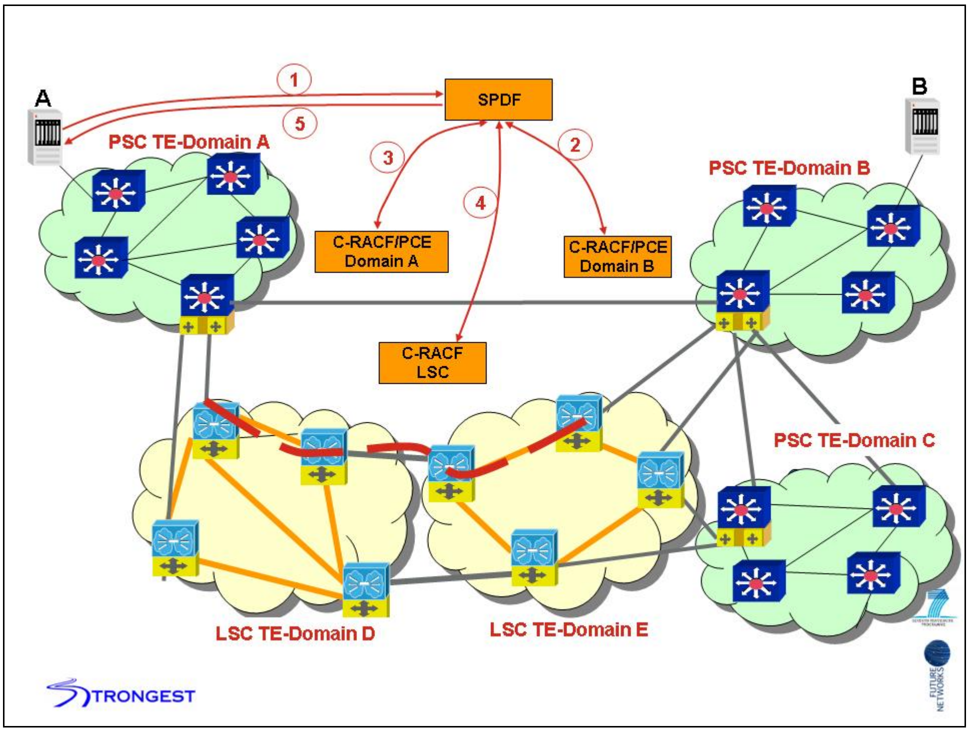


Figure 26 – End to end QoS handling using control layer

The example shows a service application that needs bandwidth reservation between server A and server B with a certain guaranteed QoS (see Figure 26). The depicted functional entities (FE) follow the G-RACS reference architecture as presented in Chapter 4.

- The application, through the G-RACS Application Function (AF-FE) functional entity, interacts with control layer asking for bandwidth between server A and B.
- SPDF (single contact point for the applications and capable of operating in a multi-domain and multi-carrier environment[4]) receives the request, performs admission control based on service policy and identifies the networks and domains involved in the request. The request in this example involves domain A, B, D and E.
- SPDF sends a QoS request to C-RACFs[5] handling LSC domains D and E, asking for QoS reservation, admission control, path computation and path creation.
    - o As stated above, C-RACF can reuse a previously created path if there is one already available to handle the requested amount of bandwidth.

[4] Either interface Gq' towards the AFs and interface Ri' towards another SPDF can be used inter-carrier.

[5] As indicated in section 4, since Access Networks are outside the STRONGEST scope, only C-RACF FEs will be considered. Only limited considerations will be devoted to A-RACF task.

- If SPDF receives positive responses, it sends a QoS request to C-RACFs handling PSC domains A and B, asking for QoS reservation, admission control, path computation and creation (if needed).

- If SPDF receives a positive response, it sends a positive response to the application.

- The application can start sending traffic from A to B.

In the following the message flow of the same scenario is assumed, but considering the case there are no resources available in one of the two (D or E) LSC domains:

- The application interacts with control layer asking for bandwidth between server A and B.

- SPDF receives the request, performs admission control based on service policy and identifies the networks and domains involved in the request. The request in this example involves Domain A, B, D and E.

- SPDF sends a QoS request to C-RACF handling LSC domains D and E asking for QoS reservation, admission control, path computation and path creation.

- If SPDF receives a negative response, it has to rollback the already reserved resources (e.g. in domain D).

- If other feasible paths do not exist (e.g. excluding domain D without available resources) SPDF sends to the application a negative response.

5.3.4 Impact of the STRONGEST architecture on the deployment of IP services

5.3.4.1 Adjacency explosion

Nowadays, network operators are concerned about the current network scenario in which traffic is increasing at an exponential rate. With this growth, the IP layer is being left behind for cheaper optical or packet transport solutions, and traffic bypass by means of IP Offloading is expected to be deployed. Transit traffic switched at the IP/MPLS layer increases costs in core networks and therefore optical bypasses are proposed as a plausible migration strategy reducing costs and the risk of reaching negative margins.

However, the IP Offloading strategy is not exempt from problems. In a large operator network, with hundreds of access nodes, a full mesh of the IP layer will lead to tens of thousands virtual links and big routing tables contained in each IP router in the network. Maintaining hundreds of adjacencies per router consumes a high amount of CPU and memory resources in each router. Those resources are not infinite, although current technology state could be sufficient.

The issue of the number of interfaces comes from the heavy increase of operation and maintenance of a very high number of links and adjacencies. In this context, the number of physical interfaces can be reduced by means of an aggregation layer such as OBS, OTN or MPLS-TP.

Besides, with current processing capacity of IP/MPLS routers, storage and management of huge routing tables should not be a big problem. However, operating, maintaining and troubleshooting such a network would become very complex, if automated process were not developed.

Nevertheless, the number of total adjacencies is not, in general, the most relevant problem that appears when IP Offloading techniques are applied to current networks. In the context of a full mesh network with all-to-all IP/MPLS adjacencies, the most important issue will appear when topology changes. That is, when a link or node failure occurs or when a new router is deployed, the link-state routing protocol will provoke an avalanche of updates, leading to routing storms, because of the need of flooding this information to hundreds of neighbors, which will thereafter flood it again to hundreds of hundreds of neighbors. This flooding process could lead to high CPU consumption, potential instabilities and reduction of overall network performance. The solution to this flooding problem is still to be found.

5.3.4.2 L2 VPN supporting end-user services

Not only is the IP layer strongly affected by the removal of IP processing in the core network and the phenomena that appears when a full mesh is established on the physical layer. When Layer-2 services are provided on an operator's network, if end-to-end connections are to be established, the Layer-2 VPNs could have some scalability issues as well. This problem is depicted in Figure 27.

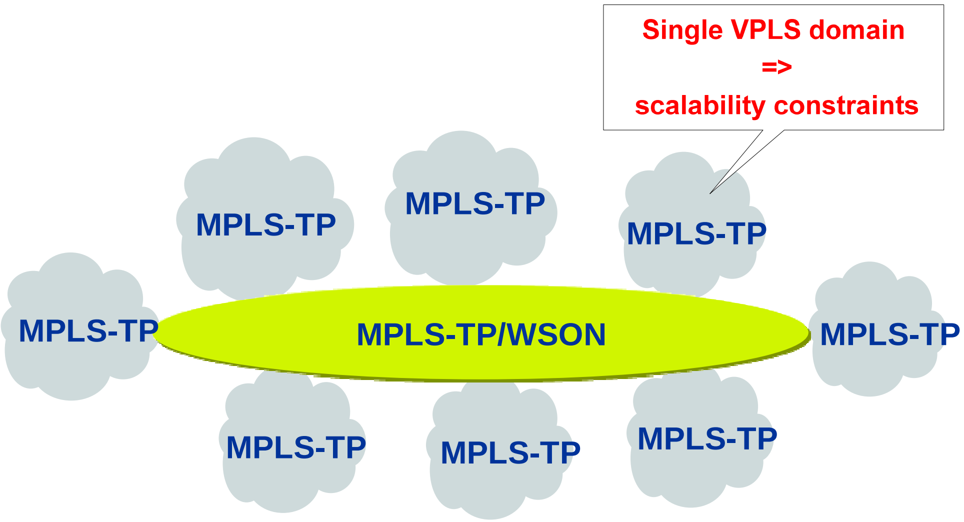


Figure 27 – Layer 2 VPN supporting end-to-end services

On one side, L2 VPN services are currently provided only for corporate clients, but in the future, operators could consider the possibility to provide these services to end-users. In this context, for single domain environments, VPNs and clients scalability is compromised, due to limitations on the number of nodes and tunnels.

On the other side, control plane scalability for L2 services is not guaranteed as the network grows. In current architectures, L2 services are usually provided over VPLS or H_VPLS architectures in a regional basis, and the provision of end-to-end L2 services is a challenge. If evolving towards an end-to-end scenario, there scalability issues could appear regarding the number of tunnels, pseudowires and clients that a single VPLS domain can support. Potential solutions could leverage on the development of multi-domain solutions or recent approaches studying “seamless MPLS”, which could potentially solve the scalability concerns.

Anyway, it must be noted that the STRONGEST architecture relegates IP processing to the network edge, so the applicability of multi-domain or seamless MPLS solutions on top of a MPLS-TP network would need to be further analyzed.

5.3.5 Future work

This section proposes directions for future work on end-to-end services in STRONGEST.

Evaluations of the application classes/network service types mapping procedures will be performed, as well as analysis on setup time limitations of end-to-end connections. Moreover, further study on end-to-end signaling will be carried out, with the aim of improving the exploitation of network resources. Particularly, smart grooming procedures (especially for WSON lightpaths), as well as methods for improving scalability of end-to-end services signaling, will be considered.

Research activities will be provided in support of the provisioning of high-quality end-to-end services with strict delay constraints. In particular, the introduction of additional features into network nodes, control plane and PCE architecture will be evaluated. For example, delay parameters will be considered as possible multi-domain TE metrics or constraints to be used in PCE-based path computations. Enhancements might then be required in support of both intra-domain and inter-domain advertisement (i.e., OSPF-TE Opaque LSA and OIF E-NNI respectively). The activity will be carried out considering both OAM aspects (e.g., how to guarantee and verify the provisioned QoS) and control plane aspects (e.g., scalability, confidentiality, TE performance).

Further investigation of multi-domain and multi-carrier scenarios, about the aspects of service definition (e.g. service templates) and the aspects of service setup including the implication on the control plane (e.g. PCE, OAM) will be future work.

6 Conclusions

The STRONGEST main objective is to design and demonstrate an evolutionary ultra-high capacity, multi-layer transport network that is compatible with Gbit/s access rates, based on optimized integration of optical and packet network nodes, and equipped with a multi-domain, multi-technology control plane. This network shall offer high scalability and flexibility, guaranteed end-to-end performance and survivability, increased energy efficiency, and reduced total cost of ownership.

The present deliverable has particularly analyzed the status of existing standards and technologies as far as OAM and control plane in complex transport networks are concerned. From this, we derived proposals for the STRONGEST network architecture that aims at providing high quality end-to-end services, focusing on medium-term network scenarios addressing the interworking between heterogeneous GMPLS-controlled networks, such as WSON and MPLS-TP.

Analyzing the state-of-the-art revealed that neither existing standards, nor available transport technologies are completely suitable to support the requirements set by the STRONGEST objectives. A number of key solutions are still missing to comply with the envisaged complex networks. For instance, the MPLS-TP / WSON interworking at OAM level is not defined, making a complete end-to-end monitoring impossible. Furthermore, the horizontal interworking in existing control plane paradigms is still immature and hampers the efficient control (set-up, tear-down, parameter modifications) of connections in a multi-region, multi-domain environment. In addition, smart grooming features and multi-domain signaling mechanisms still lack proper definitions which allow for an efficient end-to-end service delivery.

After identifying the lacks of existing standards and technologies, we recognized in this deliverable some firm starting points upon which we proceed with the next steps to further develop the STRONGEST project. In particular, the network reference scenarios have been identified together with the required, advanced OAM functions, and the preliminary control plane reference architecture. Furthermore, the guidelines for developing an automatic, control plane-governed linkage of applications needs and network services have been outlined.

7 References

[BAL]	I. Baldine, P. Mehrotra, G. Rouskas, A. Bragg, D. Stevenson, ”An intra and interdomain routing architecture for optical burst switched (OBS) networks”, International Conference on Broadband Networks 2005, vol. 2, pp. 1073 – 1082
[CHA]	Chamania, “A survey of interdomain peering and provisioning solutions for the next generation optical networks”, 2009
[Chung08]	Y. C. Chung, “Optical Performance Monitoring Techniques; Current Status and Future Challenges”, ECOC 2008, 21-25 Sept. 2008, Brussels, Belgium
[DAS]	S. Dasgupta, J. C. De Oliveira, J. P. Vasseur, ”Path computation element-based architecture for interdomain MPLS/GMPLS traffic engineering: overview and performance”, IEEE Network, pp.38-45, Jul. 2007
[draft-vigoureux]	M. Vigoureux et al, “Requirements for OAM in MPLS Transport Networks”, draft-vigoureux-mpls-tp-oam-requirements-00, Jul 2008
[Forecast]	http://www.lightreading.com/document.asp?doc_id=192711&
[G8114]	ITU-T Recommendation G.8114, Operation & Maintenance mechanisms for T-MPLS layer networks, 2007
[GOM]	L. Gommans, F. Dijkstra, C. de Laat, A. Taal, A. Wan, B. van Oudenaarde, T. Lavian, I. Monga and F. Travostino, ”Applications drive secure lightpath creation across heterogeneous domains”, IEEE Commun. Mag., vol. 44, no. 3, pp. 100-106, Mar. 2006
[ITU-T-RACF]	ITU-T Y.2111 - Resource and admission control functions in next generation networks
[Lau08]	W.Lautenschlaeger, W.Frohberg, Bandwidth Dimensioning in Packet-based Aggregation Networks, 13th International Telecommunications Network Strategy and Planning Symposium, Networks2008, Budapest, 2008
[Lee09]	J. H. Lee, H. Guo, T. Tsuritani, N. Yoshikane, and T. Otani, “Field Trial of All-Optical Networking Controlled by Intelligent Control Plane With Assistance of Optical Performance Monitors”, J. of Lightwave Technology, Vol. 27, No. 2, January 15, 2009
[MEF 26]	Technical Specification, MEF 26, External Network Network Interface (ENNI) – Phase 1 January 2010
[MEF 6.1]	Technical Specification, MEF 6.1, Ethernet Services Definitions - Phase 2 April, 2008
[NGN]	ETSI ES 282 001: TISPAN NGN Functional Architecture Release 3
[NOB2-D14]	“End-to-end experiments: results and analysis”, NOBEL phase 2, Deliverable D 1.4, Feb 2008
[NOB2-D43]	IST-IP Nobel2 Deliverable 4.3: Preliminary definition of ML TE solutions for long-term ML GMPLS networks

[NOB2-D44]	Nobel2 D4.4, “Final report on architectures, evolution scenarios and feasibility studies on innovative management and control issues in ML and multidomain GMPLS networks”, 2008
[oam-analysis]	N. Sprecher et al, “MPLS-TP OAM Analysis”, draft-sprecher-mpls-tp-oam-analysis-07.txt, Oct 2009
[oam-conf-fmk]	A. Tacaks et al, “OAM Configuration Framework and Requirements for GMPLS RSVP-TE”, draft-ietf-ccamp-oam-configuration-fwk-03, Jan 2010
[OIF E-NNI routing]	OIF E-NNI Signaling Specification 1.0
[OIF E-NNI signalling]	OIF E-NNI Signaling Specification 2.0
[OKA1]	S. Okamoto, H. Otsuki, T. Otani, ”Multi-ASON and GMPLS network domain interworking challenges”, IEEE Communications Magazine, vol. 46 pp. 88-93, Jun. 2008
[ORD]	A. Orda, A. Sprintson, ”Precomputation schemes for QoS routing”, IEEE/ACM Transactions on Networking, vol. 11, no. 4, pp. 578-591, Aug. 2003.
[PCCA]	3GPP TS 23.203 - Policy and Charging Control architecture
[PCEF]	Policy and Charging Enforcement Function
[PCEP_Marg]	draft-margaria-pce-gmpls-pcep-extensions
[PCE-WG]	Path Computation Element Working Group, http://tools.ietf.org/wg/pce/
[PCRF]	Policy and Charging Rules Function
[PEPC]	IETF RFC 5394 - Policy-Enabled Path Computation Framework
[Pinart09]	C. Pinart, “A multilayer fault localization framework for IP over all-optical multilayer networks”, IEEE Network, June 2009
[RACS]	ETSI ES 282 003 - TISPAN Resource and Admission Control Sub-system
[RFC3209]	RSVP-TE: Extensions to RSVP for LSP Tunnels
[RFC3473]	Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions
[RFC3477]	K. Kompella, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", January 2003
[RFC3630]	Traffic Engineering (TE) Extensions to OSPF Version 2
[RFC3945]	E. Mannie et al., IETF RFC 3945, “Generalized Multi-Protocol Label Switching (GMPLS) Architecture”, Oct. 2004.
[RFC4202]	Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)
[RFC4206]	K. Kompella, Y. Rekhter, ”Label switched paths (LSP) hierarchy with generalized multi-protocol label switching (GMPLS) traffic engineering (TE)”, IETF RFC 4206, Oct. 2005.
[RFC4328]	D. Papadimitriou, “Generalized Multi-Protocol Label Switching (GMPLS)

	Signaling Extensions for G.709 Optical Transport Networks Control", Jan 2006
[RFC4461]	S. Yasukawa, "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", April 2006
[RFC4606]	E. Mannie et al, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", August 2006
[RFC4655]	A. Farrel et al, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006
[RFC4726]	A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering
[RFC4834]	T. Morin, "Requirements for Multicast in Layer 3 Provider-Provisioned Virtual Private Networks (PPVPNs)", April 2007
[RFC4875]	R. Aggarwal et al, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", May 2007
[RFC5088]	J. Le Roux, "OSPF protocol extensions for path computation element (PCE) discovery", IETF RFC 5088, Jan. 2008
[RFC5089]	OSPF Protocol Extensions for Path Computation Element (PCE) Discovery
[RFC5150]	A. Ayyangar, K. Kompella, J-P. Vasseur, A. Farrel, "Label switched path stitching with generalized multi-protocol label switching traffic engineering (GMPLS TE)", IETF RFC 5150, Feb. 2008
[RFC5151]	A. Farrel, A. Ayyangar, J-P. Vasseur, "Inter domain MPLS and GMPLS traffic engineering - RSVP-TE extensions", IETF RFC 5151, Feb. 2008
[RFC5152]	J.P. Vasseur, "A per-domain path computation method for establishing interdomain traffic engineering (TE) label switched paths (LSPs)", IETF RFC 5152, Feb. 2008
[RFC5212]	K. Shiomoto et al, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", July 2008
[RFC5298]	T. Takeda, A. Farrel, Y. Ikejiri, JP. Vasseur, "Analysis of interdomain label switched path recovery", IETF RFC 5298, Aug. 2008
[RFC5392]	M. Chen, "OSPF extensions in support of inter-Autonomous System (AS) MPLS and GMPLS traffic engineering", IETF RFC 5392, Jan. 2009
[RFC5394]	I. Bryskin, "Policy-Enabled Path Computation Framework", December 2008
[RFC5420]	A. Farrel, "Encoding of attributes for MPLS LSP establishment using resource reservation protocol traffic engineering (RSVP-TE)", IETF RFC 5420, Feb. 2009
[RFC5440]	J.P. Vasseur, "Path computation element (PCE) communication protocol (PCEP)", IETF RFC 5440, Mar. 2009
[RFC5441]	JP. Vasseur, R. Zhang, N. Bitar, JL. Le Roux, "A backward recursive PCE-based computation (BRPC) procedure to compute shortest constrained interdomain traffic engineering label switched paths", IETF RFC 5441, Apr. 2009
[RFC5467]	L. Berger, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths

	(LSPs)", March 2009
[RFC5520]	R. Bradford, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", April 2009
[RFC5521]	E. Oki, "Extensions to the path computation element communication protocol (PCEP) for route exclusions", IETF RFC 5521, Apr. 2009
[RFC5553]	A. Farrel, "Resource reservation protocol (RSVP) extensions for path key support", IETF RFC 5553, May 2009
[RFC5557]	Y. Lee, "Path computation element communication protocol (PCEP) requirements and protocol extensions in support of global concurrent optimization", IETF RFC 5557, Jul. 2009
[RFC5561]	B. Thomas, "LDP capabilities", IETF RFC 5561, Jul. 2009
[RFC5623]	E. Oki, T. Takeda, J-L Le Roux, A. Farrel, "Framework for PCE-based inter-layer MPLS and GMPLS traffic engineering", IETF RFC 5623, Sep. 2009
[RFC5671]	S. Yasukawa, "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", October 2009
[RFC5710]	L. Berger, "PathErr message triggered MPLS and GMPLS LSP reroutes", IETF RFC 5710, Jan. 2010
[Sambo09]	N. Sambo, Y. Pointurier, F. Cugini, L. Valcarenghi, P. Castoldi, I. Tomkos, "Lightpath establishment in distributed transparent dynamic optical networks using network kriging", ECOC 2009, 20-24 September, 2009, Vienna, Austria
[Stanic10]	Stanic, S. ; Subramaniam, S. ; Sahin, G. ; Choi, H. ; Choi, H.-A.; "Active monitoring and alarm management for fault localization in transparent all-optical networks" IEEE Transactions on Network and Service Management, Vol. 7, Issue 2, Pages: 118 - 131, June 2010
[TOR]	Torab, "On cooperative interdomain path computation", 2006
[Y1711]	ITU-T Recommendation Y.1711, Operation & Maintenance mechanism for MPLS networks, 2004
[Y1731]	ITU-T Recommendation Y.1731, OAM functions and mechanisms for Ethernet based networks, 2008

8 List of acronyms

3GPP	3rd Generation Partnership Project
AC	Application Class
AF	Assured Forwarding
A-RACF	Access Resource Admission Control Function
ASBR	Autonomous System Border Router
ASON	Automatically Switched Optical Network
ASP	application service provider
ATM	Asynchronous Transfer Mode
BFF	Broadband Forum
BGP	Border Gateway Protocol
BM	Burst Mode
BPDU	Bridge PDU
BRPC	Backward Recursive Path Computation
BW	Bandwidth
CAPEX	Capital Expenditures
CBR	Constant Bit Rate
CC	Connection Controller
CCC	Calling/Called Party Call Controller
CIR	Committed Information Rate
CP	Control Plane
C-RACF	Core Resource Admission Control Function
E2E	End-to-end
EF	Expedited Forwarding
EIR	Excessive Information Rate
E-NNI	External NNI
E-NNI	External Network-to-Network Interface
ERO	Explicit Route Object
ETSI	European Telecommunications Standards Institute
GCO	Global Concurrent Optimization

GE	Gigabit Ethernet
GMPLS	Generalized Multi-Protocol Label Switching
HD	High Definition (TV)
HDTV	High Definition Television
H-VPLS	Hierarchical VPLS
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
ILP	Integer Linear Programming
IP	Internet Protocol
IPTV	Internet Protocol Television
ISP	Internet Service Provider
ITU-T	International Telecommunication Union
LAN	Local Area Network
LDP	Label Distribution Protocol
LMP	Link Management Protocol
LSA	Link State Advertisement
LSC	Light Switching Cluster
LSP	Label Switched Path
LSR	Label Switched Router
MAN	Metropolitan Area Network
MEN	Metro Ethernet Network
MLN	Multi-Layer Network
MPLS	Multiprotocol Label Switching
MPLS-TP	MPLS Transport Profile
MRN	Multi-Region Network
NGN	Next Generation Networks
NMS	Network management System
NNI	Network to Network Interface
OAM	Operation Administration Management
OBF	Overbooking Factor
OBS	Optical Burst Switching

OIF	Optical Interworking Forum
OPEX	Operational expenditures
OSAN	Optical Storage Area Networks
OSPF	Open Shortest Path First
OSPF-TE	OSPF Traffic Engineering
OSS	Operational Support Systems
OTN	Optical Transport Network
PC	Protocol Controller
PCC	Path Computation Client (IETF)
PCCA	Policy and Charging Control (3GPP)
PCE	Path Computation Element
PCEF	Policy and Charging Enforcement Function
PCEP	PCE Communication Protocol
PCRF	Policy and Charging Rules Function
PDU	Protocol Data Unit
PEPC	Policy-Enabled Path Computation
PSC	Packet Switching Cluster
PSC	Packet Switch Capable
PTN	Packet Transport Network
PTT	Packet Transport Technology
QoS	Quality of Service
RACS	Resource and Admission Control Subsystem
RC	Routing Controllers
RFC	Request For Comment
RSVP	Resource Reservation Protocol
SAN	Storage Area Network
SDTV	Standard Definition Television
SLA	Service Level Agreement
S-LSP	LSP Segment
SONET	Synchronous Optical Network
SPDF	Service Policy Decision Function

SRLG	Shared Risk Link Group
SSP	Storage Service Provider
TCO	Total cost of ownership
TDM	Time Division Multiplexing
TE	Traffic Engineering
TISPAN	Telecommunication and Internet converged Services and Protocols for Advanced Networking
UBR	Unspecified Bit Rate
UDP	User Datagram Protocol
UNI	User Network Interface
VBR	Variable Bit Rate
VLAN	virtual LAN
VoD	Video on Demand
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
WDM	Wavelength Division Multiplexing
WP	Work Package
WSON	Wavelength Switched Optical Networks
x-RACF	Generic Resource Admission Control Function

9 List of figures

10 List of tables

11 Document history

Version	Date	Authors	Comment
0.01	05/03/2010	Jens Milbrandt	D3.1 template distribution
0.05	14/04/2010	Jens Milbrandt	D3.1 ToC 1st proposal
0.06	22/04/2010	Jens Milbrandt, Cristiano Zema	D3.1 ToC 1st revised
0.07	05/05/2010	Jens Milbrandt, Cristiano Zema, Javier Jimenez	D3.1 ToC 2nd proposal
0.1	28/05/2010	Jens Milbrandt	D3.1 ToC agreed and editing guidelines
0.11	09/06/2010	Jens Milbrandt	D3.1 chapter editors appointed
0.12	22/07/2010	Jens Milbrandt, Paola Iovanna, Ulrich Broniecki, Filippo Cugini, David Berechya	D3.1 1st draft with all chapters included sent to chapter editors for review
0.13	29/07/2010	Jens Milbrandt	D3.1 2nd draft with all chapters included sent to chapter editors for review
0.14	17/08/2010	Jens Milbrandt, Emilio Vezzoni	D3.1 3rd draft with all chapters revised by quality manager (Emilio Vezzoni)
0.15	26/08/2010	Jens Milbrandt, Andrea Di Giglio, Emilio Vezzoni	D3.1 final draft with minor changes and a new conclusions section
1.0	30/08/2010	Andrea Di Giglio, Emilio Vezzoni	D3.1 final draft
2.0	31/08/2010	Paola Iovanno	D3.1 approved final version