Document Identity

Title:	Description and analysis of techniques for disseminating physical impairment info across the network
Subject:	Deliverable
Number:	D2.3
File name:	DICONET_D2.3_WP2_UPDATED30SEPT09
Registration Date:	4/18/2009 10:23:00 PM
Last Update:	9/30/2009 7:11:00 PM

Revision History

No.	Version	Edition	Author(s)	Date
1	0	1	Chava and Sergio	
	Comments:	Initial Release		
2	0	2	Dimitri and Didier	
	Comments:	PCEP extensions and simulation results		
3	0	3	Yabin and Ming	
	Comments:	PCE approach OSPF-TE approach		
4	0	4	Siamak and Yvan	
	Comments:	Integration issues of various architectures with Q-tool and architectural diagrams		
5	0	5	Chava	
	Comments:	Simulation results for signalling and hybrid approaches, executive summary		
6		6	Chava	
	Comments:	Revisions based on Yvan's comments		
7		7	Dimitri	
	Comments:	Revisions based on Yvan's comments		
8		8	Yabin	
	Comments:	Revisions based on Yvan's comments		
9		9	Chava	
	Comments:	Final version release for comments from Davide		
10		10	Dimitri, IBBT	
	Comments:	Update received from Dimitri		
11		11	Chava, Create-Net	
	Comments:	Final update and integration with results from real Q-tool		
12				
	Comments:			
13				
	Comments:			
14				
	Comments:			
15				
	Comments:			

Description and analysis of techniques for disseminating physical impairment info across the network

Deliverable

'D2.3'

Version: -1.0

For internal use of DICONET consortium members.

Table of Contents

List of Acronyms

AS	=	Autonomous System
CSPF	=	Constrained Shortest Path First
D-OCP	=	Distributed Optical Control Plane
ERO	=	Explicit Route Object
FA	=	Forwarding Adjacency
GMPLS	=	Generalized Multiprotocol Label Switching
IETF	=	Internet Engineering Task Force
LI	=	Linear Impairments
LMP	=	Link Management Protocol
LSA	=	Link State Advertisement
LSP	=	Label Switch Path
LSR	=	Label Switch Router
OCP	=	Optical Control Plane
ONIC	=	Optical Network Interface Card
OSPF-TE	=	Open Shortest Path First with TE extensions
RSVP-TE	=	Resource reSerVation Protocol with TE extensions
PCE	=	Path Computation Element
PCEP	=	Path Computation Element Protocol
TE	=	Traffic Engineering
TED	=	Traffic Engineering Database
PLID	=	Physical Layer Impairment Database
TLV	=	Type/Length/Value
RWA	=	Routing and Wavelength Assignment
PLI	=	Physical Layer Impairments
NLI	=	Non-Linear Impairments
PLIAR	=	Physical Layer Impairment Aware Routing
QoT	=	Quality of Transmission
QoS	=	Quality of Service
PSB	=	Path State Block
RSB	=	Reserve State Block
PPD	=	Physical Parameters Database
IA-RWA	=	Impairment-Aware Routing and Wavelength Assignment
PLIA-RWA	=	Physical Layer Impairment-Aware Routing and Wavelength Assignment

List of Figures

List of Tables

Executive Summary

In transparent optical networks, physical layer impairments (PLIs) incurred by non-ideal optical transmission media accumulate along an optical path, and the overall effect determines the feasibility of the lightpaths. If the received signal quality is not within the receiver sensitivity threshold, the receiver may not be able to correctly detect the optical signal and this may result in high bit-error rates. Introducing optical transparency in the physical layer on one hand leads to a dynamic, flexible optical layer with the possibility of adding intelligence such as optical performance monitoring, fault management, etc. On the other hand, transparency reduces the possibility of client layer interaction with the optical layer at intermediate nodes along the path. The standard GMPLS protocols used for dynamic establishment of lightpaths in transparent optical networks suffer from lack of PLIs information and optical component characteristics. Hence, there is a strong need for development of techniques that provide PLIs information to GMPLS protocols and mechanisms that use this information efficiently to evaluate optical feasibility of lightpaths, without which it would be impossible to automatically initiate a lightpath establishment from client layers, e.g., switch or an IP router or a label switched router (LSR). In the DICONET project various GMPLS protocols are extended to carry PLIs information. In this deliverable report, we briefly describe and compare various impairment-aware control plane approaches, namely, signalling-based approach, routing-based approach, hybrid approach, and PCE-based approach. The properties of these control plane architectures are qualitatively studied. Furthermore, the performance study of these various impairment-aware control plane approaches studied through preliminary results from simulation experiments. Finally, we compare various approaches w.r.t. engineering and performance metrics, with clear concluding remarks on which architectures will be further studied in DICONET project through emulation studies in WP5. As it is important to understand the advantages and disadvantages of both centralized and distributed impairment aware control plane approaches, DICONET plans to implement both centralized PCE based approach and distributed hybrid approach in which both RSVP-TE and OSPF-TE are modified to carry PLI information. As discussed later in the document, hybrid approach seems to perform better compare to other approaches at increased control plane overhead.

1. Introduction

In WDM optical networks, the physical layer impairments (PLIs) and their significance depend on network type—opaque, translucent, or transparent; the optical reach—longer lightpaths without regeneration are subject to impairments accumulation; the number and type of network elements—fiber, wavelengths, amplifiers, switching elements, etc.; and the type of applications—real-time, non-real time, mission critical, etc. In transparent optical networks, PLIs incurred by non-ideal optical transmission media accumulate along an optical path, and the overall effect determines the feasibility of the lightpaths [10-15]. If the received signal quality is not within the receiver sensitivity threshold, the receiver may not be able to correctly detect the optical signal and this may result in high bit-error rates. Hence, it is important to understand various PLIs and their effect on optical feasibility, analytical models, and monitoring and mitigation techniques. Introducing optical transparency in the physical layer on one hand leads to a dynamic, flexible optical layer with the possibility of adding intelligence such as optical performance monitoring, fault management, etc. On the other hand, transparency reduces the possibility of client layer interaction with the optical layer at intermediate nodes along the path. This has an impact on network design, planning, control, and management which is addressed in the EU funded DICONET project. The purpose of this deliverable is to discuss and explore the issues related to control and management plane protocols with specific emphasis on interaction between physical layer and control and management plane protocols.
It is important to understand the techniques that provide PLI information to the control plane protocols and that use this information efficiently to compute feasible routes and wavelengths. A recent approach to network control and management using the GMPLS framework developed by Internet engineering task force (IETF) seems to be emerging as the winning control plane solution for the next-generation optical network. One of the main applications of GMPLS [5, 14] in the context of optical networks is the dynamic establishment of lightpaths. However, it suffers from a lack of physical layer details such as PLI, transponder and other physical layer component characteristics. The availability of this information would make a GMPLS-capable node to evaluate the effects of PLIs and to decide whether a proposed path is feasible in the optical domain. In addition, GMPLS also suffers from the lack of good techniques to disseminate and utilize physical layer details. Hence, there is a strong need for the development of efficient techniques to address these issues, without which it would be impossible to automatically initiate a lightpath establishment from client layers, for example, by a switch or IP router. In addition, control plane protocols need several extensions to make them aware of PLIs. In the DICONET project various GMPLS protocols are extended to carry physical layer impairment information. This information is used by RWA protocols to compute impairment aware routes and makes sure that the established path is feasible in the optical domain. The various approaches considered in DICONET project are 1) signaling based approach, 2) routing based approach, 3) hybrid approach, and 4) PCE based approaches; which are described in detail in the following sections. This document provides a summary of the extensions required in various architectures with block diagrams.

The goal of this deliverable is to discuss various control plane architectures. The properties of these control plane architectures are qualitatively studied. We finally evaluate selected architectures (based on Milestone M2.3 [3] and qualitative study) and study the performance through simulation experiments. At the end of the document we compare various architectures w.r.t. engineering and performance metrics, with clear concluding remarks on which architectures will be further studied in DICONET project through emulation studies in WP5.

2. PLI-aware control plane architectures and integration

Two main approaches to introduce impairment-awareness in GMPLS-based transparent optical networks are: centralized and distributed [11, 15]. Centralized approaches assume availability of a centralized server that is reachable by all network elements (NEs) and aware of complete network topology, resource availability and physical layer parameters through traffic engineering database (TED), which are used during routing and wavelength assignment (RWA). Hence, centralized approaches are able to guarantee and satisfy a set of lightpath specific requirements (e.g. bandwidth, diversity, QoT, latency, etc.). Two different mechanisms can be used to implement *centralized approach*:

- Network management system (NMS): For a connection request, NMS computes the route and selects a wavelength considering both the current TED and the requirements of the connection request. Then, it configures various optical components (e.g. optical cross-connects) along the route in parallel using the network management interface to setup an optically feasible lightpath.

- Path computation element (PCE): For a connection request, the source node sends a request to the central PCE which computes the route taking into account the TED and the connection requirements. PCE returns the computed route to the source node which establishes the lightpath using distributed signalling protocol such as RSVP-TE.

Centralized approaches have advantages due to the awareness of complete and detailed view of the whole network through the central TED and can find the optimal route. However, both NMS and PCE approaches have scalability problems. NMS has additional interoperability problem and PCE is lack of flexibility and can not realize rapid restoration in case of multiple failures. However, PCE based approaches can be extended to provide rapid restoration by computing protection paths for all failed lightpaths in single computation and using table and lookup kind of service. Finally, PCE standardization is still at an early stage and its applicability is mainly devoted to inter-domain path computation. Hence in this study we focus on distributed approaches.

Distributed approach (distributed optical control plane [D-OCP]) is performed via all NEs in the network using a common distributed GMPLS control plane which manages the required procedures for establishing lightpaths, i.e., routing and signalling. There are two different mechanisms under the distributed approach:

- Signalling-based optical control plane (S-OCP) [8, 10, 12, 17] extends the RSVP-TE protocol to include PLI information. Each node computes a route using standard OSPF-TE protocol without the knowledge of PLIs and then RSVP-TE carries the PLI information along the route till the destination node, which evaluates the optical feasibility and establishes the lightpath.
- Routing-based optical control plane (R-OCP) [11, 21] introduces PLI information into the routing protocol such as OSPF-TE, as suggested in IETF RFC4054 [21]. By flooding link state advertisements (LSAs) enhanced with PLI information, all nodes populate their TED, thereby allowing each node to compute a feasible route while standard RSVP-TE is used for lightpath establishment. In R-OCP each node has its own PCE, which means that it is functionally equal to the centralized PCE approach, though the coordination among different PCEs is missing.
- Hybrid approach extends both RSVP-TE and OSPF-TE is extended in order to carry and disseminate PLI information. Several variations of hybrid approach are possible to provide a trade-off between the information carried in OSPF-TE and RSVP-TE protocols, the complexity of protocol extensions, control overhead, and network performance. In general it is sufficient to extend to OSPF-TE to disseminate PLI information that is useful for intelligent RWA computation, while RSVP-TE carries full information to validate the optical feasibility during lightpath setup to accommodate the network status changes and inconsistent PLI information.

In this section, we present four architectural options in brief with schematic diagrams with interfaces between various modules. However, the complete definition of various interfaces is yet to be defined. The network planning and operation tool (NPOT) is the heart of the whole DICONET project, and encompasses the RWA module and the Q-Tool estimation module. The NPOT will provide interfaces to various kinds of architectures to support various functionalities. Some initial definition of interfaces and Q-tool implementation are available and are discussed in other deliverables (e.g., Q-Tool API in M3.2 [22], other APIs in their respective work packages, and the whole set of APIs will be summarized in D3.3 [23]).

Note that there is a difference between TE extensions and extensions required to carry PLIs information to compute the feasibility of lightpaths in optical domain. The usual TE extensions are provided to build traffic engineering database rather than earlier link state database. The TE extension provide capabilities such as extended link attributes required for local constraint-based routing and global traffic engineering database, etc. These extensions are based on opaque LSA extensions. For detailed TE extensions refer to RFC 3630 [26]. However, the aim of DICONET project is to provide GMPLS capable nodes with information related to PLIs and other optical component characteristics to enable GMPLS-capable switches to evaluate the feasibility of lightpaths in optical domain to realize dynamically reconfigurable next-generation optical networks, which can initiate lightpaths directly from client layers without knowing the physical layer details.

2.1. Signaling based (RSVP-TE) architecture

In the signaling based approach the routing component of the control plane (i.e. OSPF-TE) is NOT extended. However the signaling component (i.e. RSVP-TE) is extended to consider the PLIs and quality of transmission (QoT) metrics. The metrics are those that the Q-Tool takes as inputs, as described in milestone M3.2. In this approach the lightpath establishment procedure takes place as follows (see Figure 2.1):

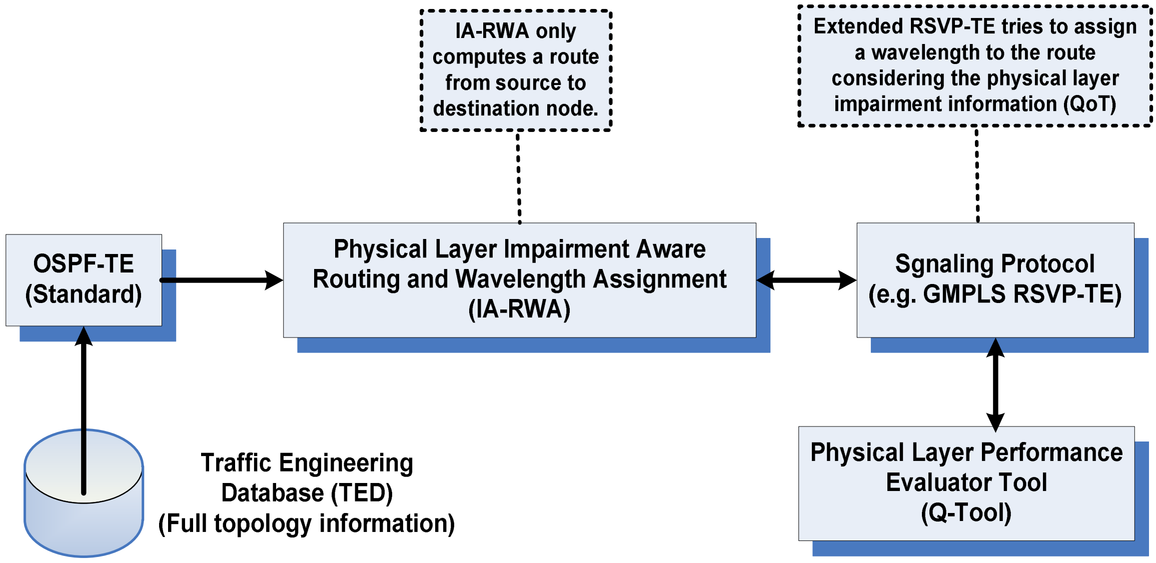

Figure 2-1 Signaling-based (RSVP-TE) architecture

1. The OSPF-TE component of the control plane provides the PLIA-RWA or IA-RWA with the full network topology information. Note that unlike in routing based approach, the topology information and active connection information are not sent to the IA-RWA component. Hence the path calculated is not aware of PLI information in signaling based approach.
2. The IA-RWA component in this approach ONLY computes a route (or K-routes) from source to destination without PLI information. The computed route is used in extended PATH message. When the extended PATH message reaches the destination node, it has all available wavelengths. The first available wavelength is checked for Q-factor value using Q-tool interface and local information.
3. If a wavelength is found for the given route, the QoT of the lightpath is checked at the destination node using the interfaces provided to Q-tool and the lightpath will be established. Otherwise the extended signaling protocol requests next possible from the (IA)-R(WA) component. If none of the candidate routes satisfies the QoT requirement, the demand is blocked.

In RSVP-TE based architecture, TED is built using the usual TE extensions to OSPF without PLI information and is available at all nodes in the network. As the lightpath feasibility is evaluated at the destination nodes, each node needs to run an instance of Q-tool. IA-RWA computes the route at source without using any impairment information and hence it needs to be implemented on all nodes in the network. The establishment of new lightpath may introduce excessive nonlinear crosstalk (XPM, FWM) on existing active connections and need to be checked before establishing the new lightpath and these details are discussed in detail in Section 3. Also note that there is several use cases defined in D5.1 [28], e.g., automatic rerouting in case of failures using protection/restoration techniques, etc. Here we have provided an example use case scenario of lightpath establishment procedure.

2.2. Routing based (OSPF-TE) architecture

The building blocks of this approach are depicted in Figure 2.2. Only the routing component (i.e. OSPF-TE) of the control plane (i.e. GMPLS) is extended. These building blocks have to be implemented in all nodes in the network. In this integration approach, the lightpath establishment process takes place as follows:

1. Since the OSPF-TE is extended, the full network topology, information about all active (already established) lightpaths (connections) and physical layer impairment information are at disposal of physical layer impairment aware routing and wavelength assignment (IA-RWA) component. Note that IA-RWA cannot directly access the physical layer database (PLD)/physical parameters database (PPD) or the traffic engineering database (TED). The proper interface to these two databases is provided through extended OSPF-TE component.

2. The IA-RWA component, utilizes its interface to the Q-Tool and computes a route and available wavelength (if any is available) for the requested demand.

The IA-RWA interfaces to the signaling part of the control plane (i.e. RSVP-TE for GMPLS) and establishes the lightpath. Note that the signaling protocol (i.e. RSVP-TE) is not extended in this approach. In this architecture, IA-RWA computes the explicit routes with the knowledge of impairments provided by PPD and TED through proper OSPF-TE interfaces; hence an instance of IA-RWA needs to be run on all nodes in the network. Also as the Q-tool verifies the feasibility of lightpath at the source nodes in conjunction with IA-RWA, it also need to be run on all the nodes in the network.

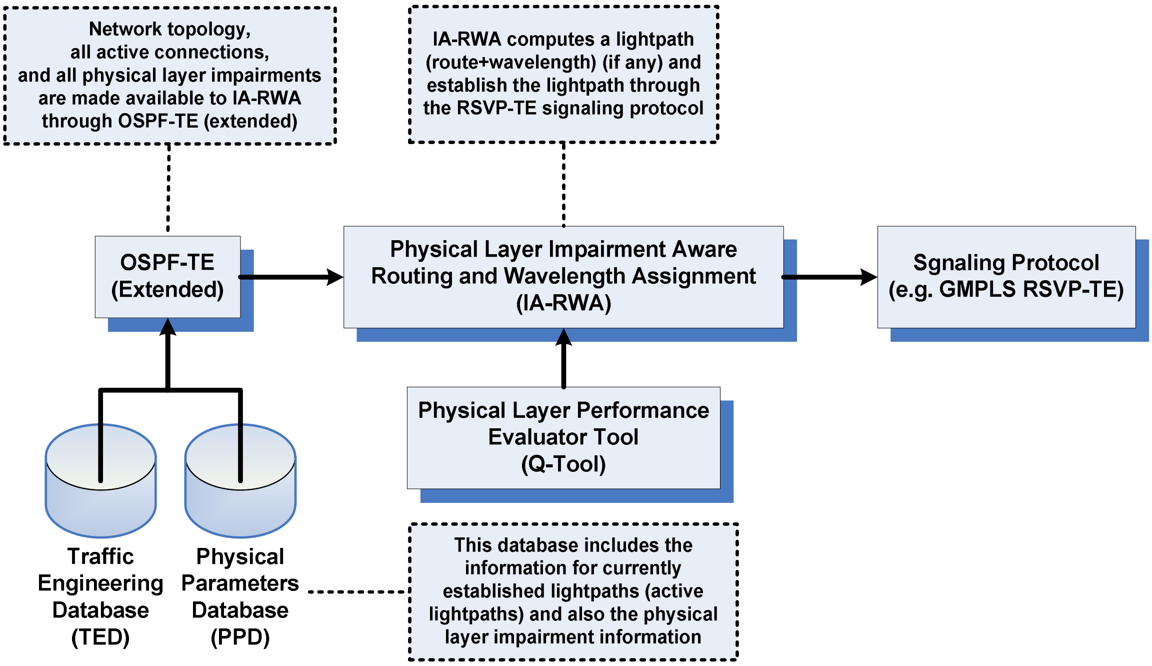

Figure 2-2 Routing (OSPF-TE) based architecture

The routing based approach alone is not considered in DICONET project due to following reasons: 1) DICONET project tries to provide an efficient solution for very dynamic networks, where the inter-arrival time of the connections too small or arrival rate is high, which increases the control overhead in routing approach leading to inconsistent databases. 2) the solution developed in DICONET should also cater to failure scenarios where there are multiple connection requests and need to be rerouted at almost the same time leading to increased control overhead. 3). The important and necessary extensions required in routing based approach is studied in hybrid approach.

2.3. Hybrid architecture

Both stand-alone signaling and routing approaches suffer of some drawbacks which may affect their performance. Therefore, it is a matter of fact that by combining the approaches, in a so called hybrid approach, the aforementioned drawbacks could be overtaken. In hybrid approach both the signaling and routing component of the control plane are extended in order to encompass and disseminate the physical layer impairment information. As explained earlier, several variations of hybrid approach are possible to provide a trade-off between the information carried in OSPF-TE and RSVP-TE protocols, the complexity of protocol extensions, control overhead, and network performance. In general it is sufficient to extend to OSPF-TE to disseminate PLI information that is useful for intelligent RWA computation, while RSVP-TE carries full information to validate the optical feasibility during lightpath setup to accommodate the network status changes and inconsistent PLI information. In this section, we explain one of the possible examples (an extreme case) where both protocols are extended to carry all the required information. The input to the IA-RWA component includes the full network topology, all active lightpaths (currently established lightpaths) and physical layer impairments. This information are kept in PPD database (however we can think about separate databases for each of them, which is a matter of design and implementation). The lightpath establishment process takes place as follows (Figure 2.3):

1. The IA-RWA component computes a route and wavelengths (i.e. lightpath) from source to the destination node, considering the physical layer impairments.
2. The candidate lightpath is forwarded to the extended signaling protocol (i.e. RSVP-TE) to establish the lightpath, while double checking the QoT using proper Q-tool interfaces. Due to non-zero convergence time, the PPD and TED databases used to evaluate the feasibility of lightpath at the source node may not contain up-to-date information PLIs. Also there can be network status changes (either because of a new lightpath establishment or active lightpath torn down between the feasibility computation at source node and the actual reservation in RESV phase of RSVP-TE. Hence, the RSVP-TE protocol again checks the feasibility during RESV phase to accommodate these changes and to make sure that no active lightpath is disrupted due to inconsistent database information.

If the QoT of the candidate path is above a given threshold, then the light path will be established. Otherwise, a PATH_ERROR message will be forwarded back to the source node. Again in this architecture, IA-RWA and Q-tool needs to be implemented on all the nodes in the network as explained earlier.

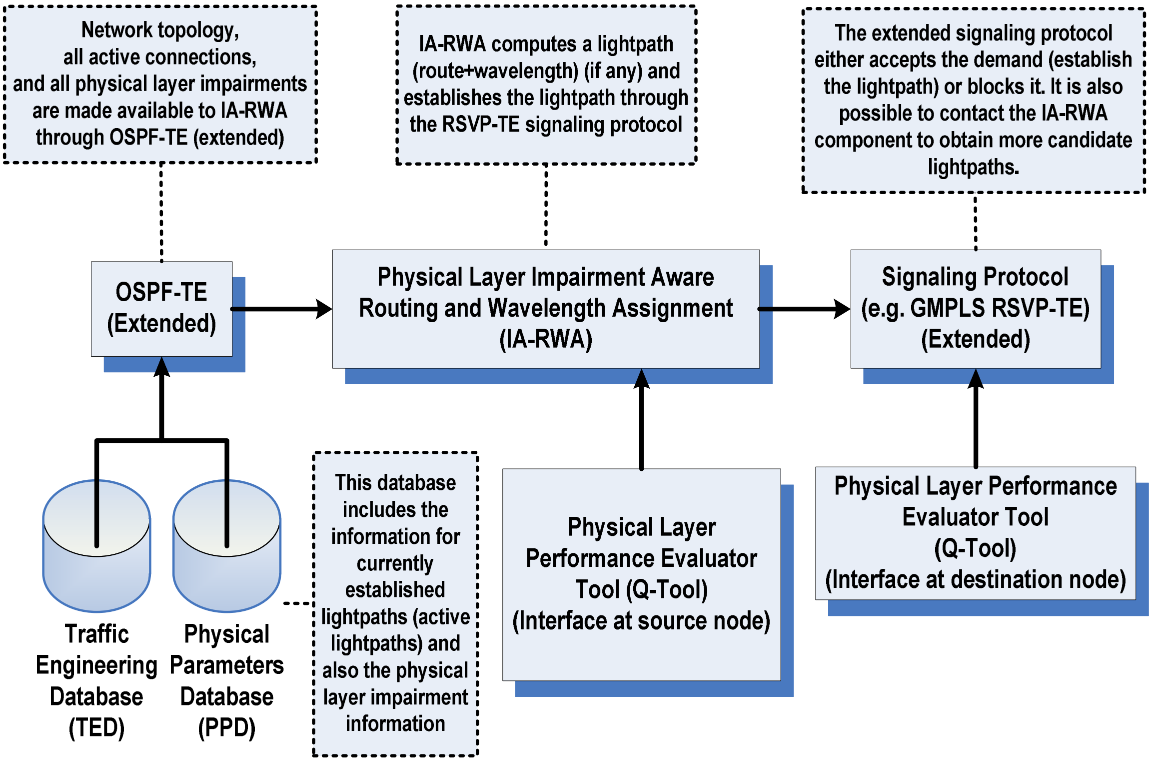

Figure 2-3 Hybrid integration approach

2.4. PCE based architecture

In this integration approach the whole process of physical layer impairment aware routing and wavelength assignment is performed in a single computing element, named as PCE [16]. In the PCE framework, all nodes are running a particular protocol to disseminate their status to the central PCE. We call this software module at each node as PCE Agent. The counterpart software component in the centralized PCE server is named PCE Manager. This protocol can be somehow considered as an SNMP-like management protocol. Therefore it provides the central PCE with the topology information, already established connections and also the physical layer impairments information. Note that monitors can communicate with each node using a local protocol (e.g. SNMP using proper database conventions explained in D5.1) and then the node disseminate these information to the central PCE server. The lightpath establishment process in this approach takes place as follows (Figure 2.4):

1. The source node contacts the PCE server asking for a lightpath establishment to a particular node in the network. We can also think about an operator, which can utilize the PCE interface for establishing a lightpath form a central server (i.e. PCE server user interface).

2. The IA-RWA, Q-Tool and all the required information (i.e. network topology, currently established connections and physical layer impairment information) are already collected in the PCE central database.
3. PCE computes a route and wavelength (i.e. lightpath) for the requested demand using IA-RWA and communicates it with the standard signaling component of the control plane (i.e. GMPLS/RSVP-TE).
4. The signaling component at the source node uses the standard signaling for establishing the lightpath.

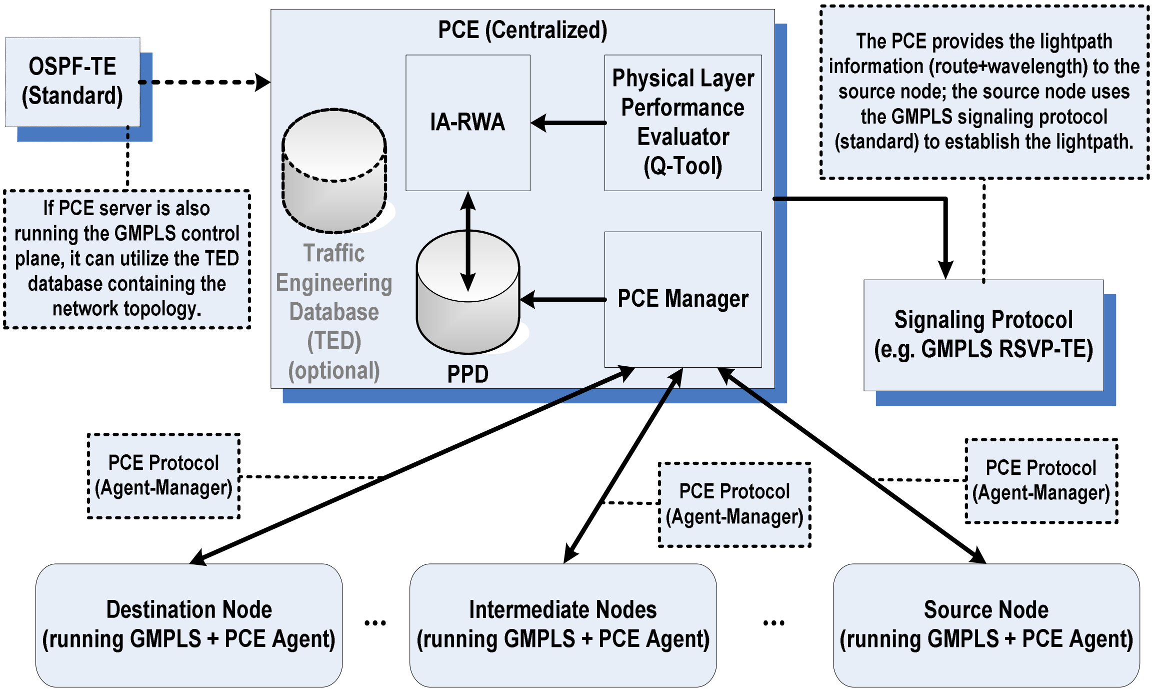

Figure 2-4 PCE based approach

3. Signaling based (RSVP-TE) architecture

In this section, we explain detailed working procedure of standard RSVP-TE protocol, the mechanisms and extensions proposed to introduce PLI-awareness into standard RSVP-TE protocol. In general, before establishing a lightpath for a new connection request, GMPLS protocols should be able to 1) evaluate the feasibility of the new lightpath request w.r.t. linear impairments and w.r.t. to non-linear impairments (i.e., the effect of active connections in the network on the new lightpath request) 2) the effect of new lightpath request on active connections, as the new lightpath may introduce excessive cross on active connections making them unfeasible (i.e., possible disruption of active connections). In this section we briefly discuss the techniques to address these issues and provide brief discussions on the information required and how this information carried in extended RSVP-TE protocol.

3.1. Introduction and basic architecture description

In this architecture when a traffic request arrives at some node, the establishment of the corresponding lightpath inside the network takes place through two main operations:

1. a route computation phase, where a path from source to destination is computed according to some constraints.
2. a signaling phase, which is responsible for the LSP optical feasibility validation and impact on existing channels and setting up the optical channel inside the network itself.

The role of the route computation phase is to provide a set of K paths, from the source to destination, ordered with respect to their length or/and number of hops. In the following we explained the brief working procedure of standard RSVP-TE signalling protocol:

When there is a LSP setup request, a route from source to destination node is computed using the information collected by routing protocol, e.g. OSPF-TE. However, the standard OSPF-TE is not aware of free wavelengths available and the impairment information on each link in the network, but it has the knowledge of availability of network link as long as free capacity is available. Once the route is computed using standard OSPF-TE routing protocol, the signaling session is triggered along the computed route. The source node sends a *PATH* message containing a *generalized label-request* object, and optional objects such as *label-set* and *suggested-label*, etc [16, 17]. The intermediate nodes along the route process *PATH* message, creates a proper state for the session in a local database and updates the *label-set* depending on wavelength availability and forwards it to the next hop. When the *PATH* message arrives at the destination node, it sends a *RESV* message from destination to source containing *generalized label* object with a common free wavelength selected from *label-set*. When the *RESV* message is at some intermediate node, the node configures the local OXC and forwards it to the upstream neighbor. When the *RESV* message reaches the source node, the LSP is established. In addition to the *PATH* and *RESV* messages, other signaling messages such as *PATH_ERROR* and *RESV_ERROR* to indicate the failure conditions to upstream and downstream nodes, respectively, are defined. *PATH_TEAR* and *RESV-TEAR* are also defined to delete the LSP and the related session state along the route. Finally, note that as RSVP-TE we implemented is based on backward reservation method in which resources are reserved during *RESV* message, hence there is no forward blocking (i.e., no reservation of resources in the PATH message hence there is no blocking of resources before the final evaluation and RESV procedure).

3.2. Handling PLIs using extended RSVP-TE

3.2.1. Mechanism description

For a new connection request, the source node computes *K*-explicit routes. The signalling process checks the optical feasibility of first explicit route by sending out a *PATH* message containing several fields describing linear impairments (LIs, e.g.: OSNR, power, CD, PMD, crosstalk, etc.,) and nonlinear impairments (NLIs: e.g., path section characteristics such as signal power, fiber type, and length; signal type such as modulation format and bit-rate; adjacent active channel characteristics such as wavelength, LSP id, crosstalk margin, residual CD, reference FWM crosstalk) and a list of available transponders/wavelengths. Upon reception of *PATH* message, each intermediate node updates these fields considering wavelength continuity constraint and active lightpaths or wavelengths/channels on the path section. If there is no free wavelength on its outgoing link, then it sends *PATH_ERROR* message to source node. If the destination node receives *PATH* message, it evaluates Q-factor for the new lightpath considering both LI and NLIs using the Q-Tool. If the computed Q is above the threshold, the lightpath is feasible and can be established during RESV phase. Then it evaluates the effect of the new lightpath on existing active lightpaths by requesting the re-computation of the Q-factor of existing lightpaths by contacting the destination nodes of the corresponding active lightpaths. If the establishment of new lightpath does not decrease the value of Q-factor below threshold, then the new lightpath can be established. Q-tool requires following inputs and provides the following output at the destination nodes of lightpaths (notations are those of Fig. 3):

- Q-tool input: A snap shot of the topology required to compute the Q-factor for each available wavelength, i.e.,
    - Signal type
    - Selected wavelength
    - Snap shot of the network “linear” topology (comprising only the links of the selected lightpath) and active connections consisting of
        - Path characteristics
        - Active LSPs
- Q-tool output:
    - Q-factor value for a selected lightpath.

3.2.2. Mechanism to avoid potential disruption of existing active lightpaths

The mechanism explained earlier assures that lightpath under consideration is feasible, if computed Q-factor of new lightpath is greater than the threshold Q-value, i.e., Qcomp > Qthr. However, this new lightpath *may introduce* some multi-channel effects such as XPM and FWM, and potentially *reduce Q-factor* of existing active lightpaths, making them *infeasible* in optical domain. Hence we need a mechanism to make sure that the new lightpath is NOT disrupting any active lightpaths before we establish new lightpath. This mechanism can be implemented in several ways. In this document, we present one of the possible mechanisms implemented within DICONET. In this mechanism the destination node of the new lightpath, requests the destination nodes of active lightpaths that are overlapping at least on one segment of the new lightpath, by sending the Q-CHK-REQ with the required information to evaluate the Q-factor of the corresponding active lightpaths. The destination nodes of the active lightpaths evaluate Q and checks if it does not violate the constraints on Q even if the new lightpath is established. If there is no violation in Q requirements, the destination node of active lightpath updates its local database with new information considering the new lightpath setup and responds with Q-CHK-ACK to the destination node new lightpath, otherwise, it responds with Q-CHK-NACK without any local database update. The detailed contents of these messages are discussed later in this section.

The detailed implemented mechanism is shown in an example Fig. 3.1. Let's assume that the total number of wavelengths available are 4, from λ_1 to λ_4 and two lightpaths LSP1 (shown in red color) with route = {7-2-3-4-8} and LSP2 (shown in blue color) with route = {9-3-4-5-10} are established on λ_1 and λ_2 at time T = -2 and -1, respectively and are active. Assume that there is a new lightpath request (LSP3) at time T = 2 with the corresponding route = {1-2-3-4-5-6}. The PATH sent from source node (i.e., node1) carries all the required information to compute the Q of new lightpath and also of the active lightpaths that are overlapping the new lightpath. In this example the PATH message carries information of path and section characteristics of each overlapping segments of lightpaths LSP1 and LSP2. The destination node (i.e., node 6) computes the Q for the new lightpath for each available wavelength. Then node 6 requests for verification of Q of existing active lightpaths LSP1 and LSP2 by sending the Q-CHK-REQ requests to nodes 10 and 8, respectively. If a specific wavelength meets both constraints, i.e., 1) Q of new lightpath is greater than the Qthr and 2) Q of existing active overlapping lightpaths is not below the threshold, then that wavelength is selected and used in the reservation phase. If all available wavelengths are failed either because of its Q or due to the introduction of excessive crosstalk (reducing Q of active lightpaths), destination node sends PATH_ERROR message to source node. Then source node sends PATH message on second candidate path.

Due to the update of local databases at destination nodes of affected lightpaths during Q-CHK-REQ and Q-CHK-ACK there exists some more complex issues (such as roll back of updates done at the affected destination nodes) and are handled as follows: When destination node of LSP3 contacts destination node of LSP1 and 2, it updates the information stored in their database if they are not disrupted. If LSP Q1 > Qthr and LSP Q2 < Qthr, and the ACKs are received in that order; then we have to roll back the update done at the destination node of LSP1 (in general if one or more Q-CHK-REQ fails, all successful requests must be rolled back)

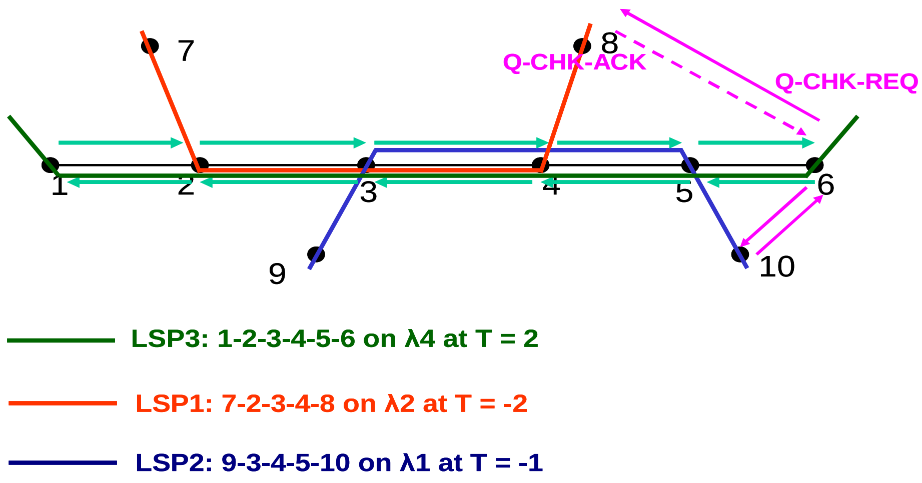

Figure 3-1 An example to handle potential lightpath disruption

3.2.3. Mechanisms to deal with concurrent lightpaths requests

3.2.3.1. Description of the problem

The signaling based architecture also needs to handle the concurrent lightpaths requests and therefore the signaling based approach has to address this issue as well. By looking at the Figure 3.2, it is possible to understand why the concurrent setup of LSPs sharing a few (or all) links may be a problem in the case of the signaling based (RSVP-TE) approach. In fact if the setup requests for LSP-1 and LSP-2 arrive almost concurrently, the corresponding RSVP-PATH messages are sent out one after the other from nodes N1, N2, N3 so that LSP-1 does not see LSP-2 and vice versa. If the node N8 selects a different wavelength for LSP-1 from the one chosen for LSP-2, both LSPs will be established but none of them considers the impact of the other so that LSP-1 and LSP-2 Q-factors will be wrongly evaluated since they will be underestimated. But actually, one or both LSPs might be unfeasible by the optical viewpoint and in such a case the network will not operate properly. Summarizing we conclude that the concurrent setup of LSPs sharing links may lead to the establishment of optically unfeasible LSPs. Therefore the signaling based approach has to provide some mechanisms to manage the concurrent setup of lightpaths. The mechanisms developed to handle concurrent lightpath requests are discussed in the following subsections.

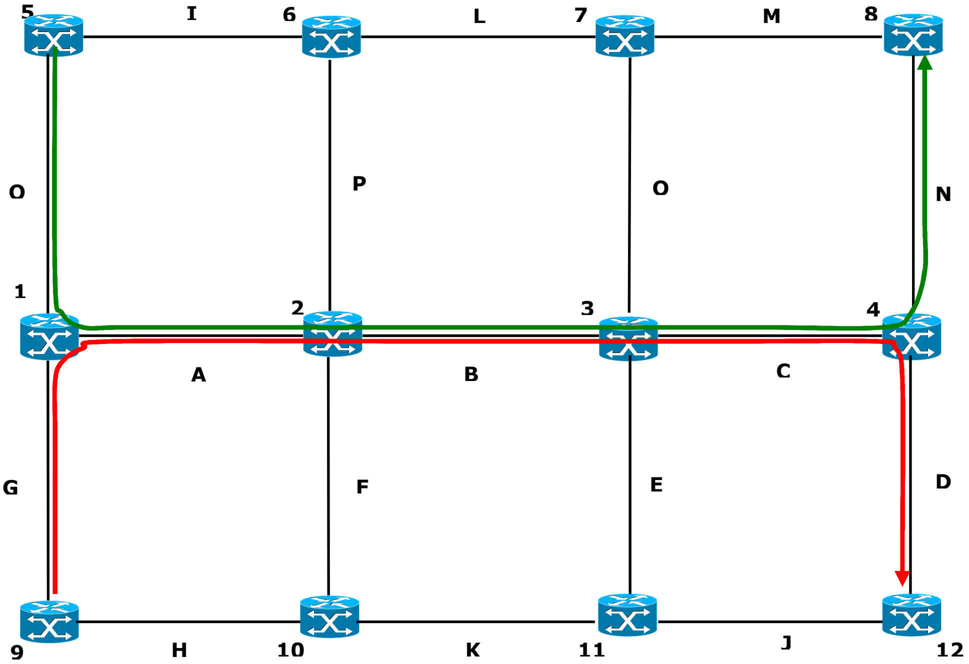

Figure 3-2 Illustration of concurrent lightpath setup in signaling based approach

3.2.3.2. Interface Locking

To assure that over a given path no other LSP is signaled while another lightpath is being set up, links along that path can be locked on the PATH message processing. More precisely during the LSP setup phase at a given node along the path, after processing the PATH message the corresponding output Optical Network Interface Card (ONIC) is locked and it is unlocked when receiving the RESV or PATH_ERROR message. While ONIC is locked, any PATH message, which has same interface as output ONIC, is queued. Figure 3.3 illustrates the flow of signaling messages exchanged for setting up the two lightpaths depicted in the Figure 3.2 when the interface locking mechanism is used.

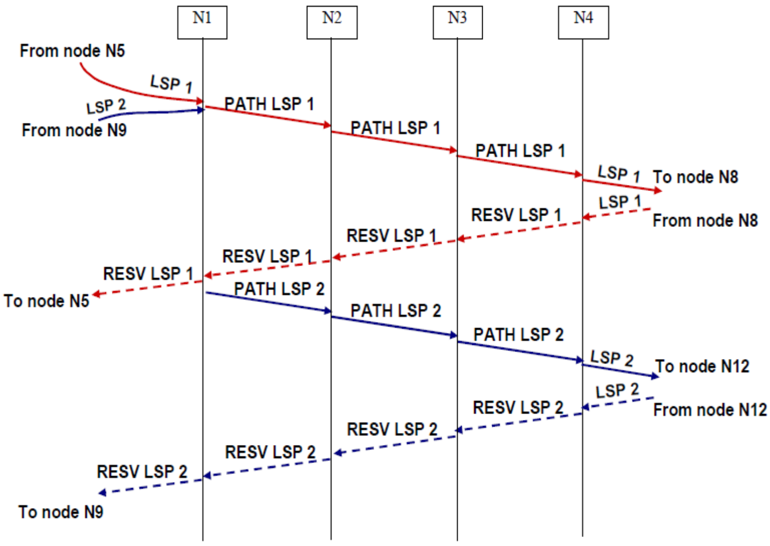

Figure 3-3. Concurrent LSPs setup by using the locking mechanism.

However the processing of the refresh-PATH messages (refer to RFCs on standard RSVP-TE [19, 20]) is different. AS RSVP-TE is soft-state protocol, it requires periodic refresh to update the LSP status and is done using PATH refresh messages. In fact these are processed and sent ahead even though the corresponding output interface is locked since:

1. Blocking a refresh-PATH message may lead to the refresh timeout event in some nodes: this causes in turn the removal of the PSB and RSB associated to that lightpath.
2. The PATH message, which has locked the output interface, saw the already established lightpath to which the refresh-PATH message refers.
3. In the case the message locking the output interface is a setup-PATH, the associated new lightpath (or part of it) may not be seen by the refresh-PATH message, but this is not a problem. In fact during the optical feasibility check, the egress node of the new lightpath calls the destination node of the old LSP (associated to the refresh-PATH message), which stores the parameters of the new LSP as long as it is feasible. Similarly in the case the message locking the output interface is a refresh-PATH, there is no problem at all.

The RESV and PATH_ERROR messages are instead processed at once when received by the RSVP-TE module and as mentioned above they remove the lock added on the processing of the associated PATH message to the output interface of the corresponding lightpath. The locking mechanism instead affects the handling of the RESV_ERROR since:

1. In a node, before sending downstream the RESV_ERROR message (unless it is the destination node), the output interface associated to the corresponding lightpath is locked. This is done in order to avoid concurrent LSPs setup attempts if the egress node on receipt of the RESV_ERROR message tries another wavelength instead of send back a PATH_ERROR message.
2. In a source or intermediate node, a RESV_ERROR message is processed if the output interface associated to that lightpath (to which the RESV_ERROR message refers) is not locked by a setup-PATH or another RESV_ERROR message, otherwise the RESV_ERROR message is queued and will be processed when the interface becomes unlocked. This is also done in order to avoid concurrent LSPs setup attempts. Instead if the interface is locked by a refresh-PATH, the RESV_ERROR message is processed and then sent downstream.

3.2.3.3. The trouble of the locking mechanism: deadlocks

A well-know problem affecting the locking mechanism is the deadlock. A deadlock event (see Figure 3-4) occurs when two LSP requests are generated almost concurrently and the PATH messages of one LSP is blocked by a link locked by the other LSP and vice versa. In such a situations the links remain locked forever because the PATH messages cannot reach their egress nodes and the corresponding RESV or PATH_ERROR messages, which remove the locks from links, will not be sent back.

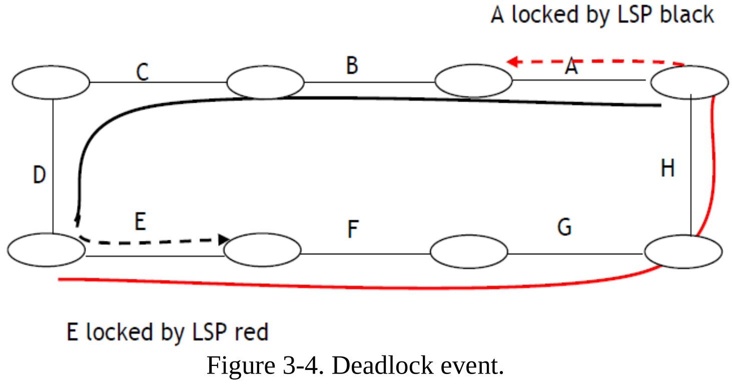

Figure 3-4. Deadlock event.

Of course deadlocks situations must be prevented. There are three possible solutions for them:

1. No PATH message queuing at locked output interfaces: if the PATH message encounters a locked link along a path, such a path will be rejected and the source node will try a new path. This solution is simple but too conservative especially if the inter-arrival time of lightpaths requests is short.
2. Set a timer for each PATH message queued inside a locked interface; if the timeout event occurs for a queued message, the path is rejected and a PATH_ERROR message is sent back. The critical issue of this solution is the determination of the timeout interval.

3. Deadlock explicit detection: if a PATH message encounters a locked interface, it is checked that this message along its way has blocked no interface that is going to be crossed by the PATH message which locked this interface. However we have to highlight that not all deadlocks can be detected but such an event may only occur if the requests of establishing three or more LSP arrive at different network nodes almost concurrently. An example of an undetectable deadlock is illustrated in Figure 3-5 where it is showed the attempt of establishing three lightpaths in a ring network topology. In this example the LSP-1 PATH message finds the output interface locked at node N3 and the deadlock check does not detect anything because LSP-1 PATH message has not blocked the LSP-2 PATH message which is locking the output interface at node N3. The same goes for the LSP-2 PATH message which is blocked by LSP-3 at the node N4 and for the LSP-3 PATH message which is blocked by LSP-1 at the node N2. This particular (and synchronized) sequence of events has led to a undetectable deadlock because:

    i. no PATH message can reach its destination node so that no RESV message is generated and sent back. In this manner locks are not removed from the corresponding locked interfaces.

    ii. no blocked LSP can detect the deadlock event.

However, an undetectable deadlock is a very rare event which may only occurs if the inter-arrival time between LSP requests is very low.

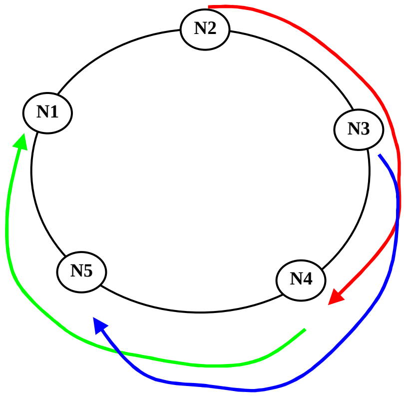

Figure 3-5. Undetectable deadlock.

Since it is possible not to be able to detect all deadlocks, a mixed approach based on the explicit detection together with a timeout for queued messages, sounds to be the best approach to deal with deadlocks.

With regard to bidirectional lightpaths we have to highlight that deadlock events occur frequently during the upstream LSP setup, as illustrated in Figure 3-6 where the setup attempts of LSP-1 (whose setup request is generated by node N1) and LSP-2 (whose setup request is generated by node N4) are showed.

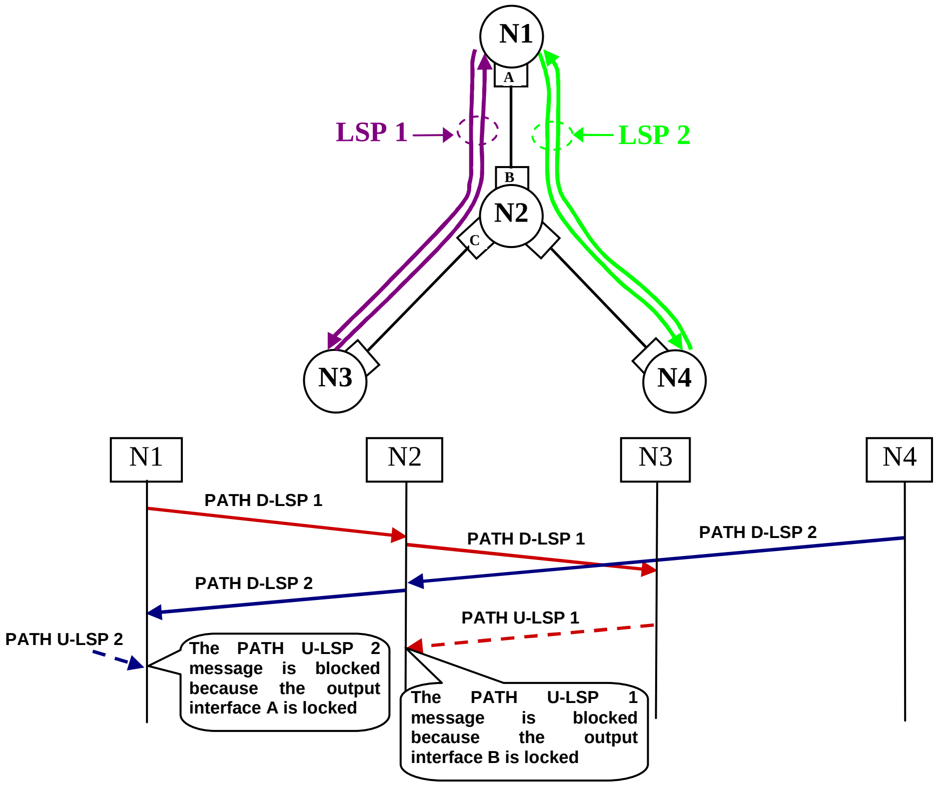

Figure 3-6. Deadlock events happening during the uplink setup of both LSP-1 and LSP-2.

However in order to significantly lower the deadlock events during the uplink LSPs setup, it is enough to lock the output interfaces for both directions at the intermediate nodes after processing the setup-PATH message for the downstream. By referring to the example in Figure 3-6, this means that at node N2 after processing the PATH message for the downstream component of LSP-1, both interface *A* and *C* are locked. The complete flow of signaling messages for this example is shown in Figure 3-7.

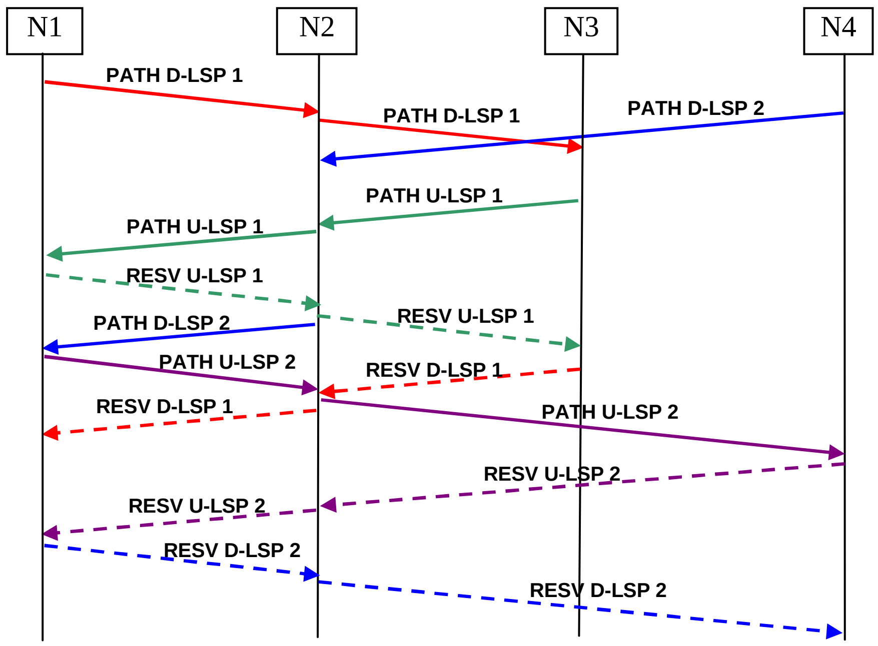

Figure 3-7. Flow of signaling messages for the example considered in Figure 3-6.

As we can deduce, the locking mechanism negatively affects the setup time especially when the inter-arrival time between LSP request is short. If this fact should not be a problem for the setup of a new lightpath, it might be in a restoration situation where the setup time is a very critical issue. Hence, we plan to develop new mechanisms to handle this issue during the implementation in WP5.

3.3. Protocol extensions for carrying PLI-information

The aim of this section is to explain how the GMPLS RSVP messages are extended in order to include information related to transmission impairments [4-7, 9]. First, we start with listing the additional parameters (along their encoding) that the GMPLS RSVP messages have to carry in order to signal a connection inside a transparent/translucent optical network:

Requested service/traffic parameters:

- Service type (encoding type, payload type): this information is encoded by using the
- Standard GENERALIZED_LABEL_REQUEST (RFC3471 [4], RFC3473 [5]) whose format is illustrated in Figure 3-8. Refer to RFC 3471 and 3473 for explanation on each of these fields.

```
0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Length            | Class-Num (19)|  C-Type (4)   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| LSP Enc. Type |Switching Type |             G-PID             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3-8. GENERALIZED_LABEL_REQUEST format.

- Service QoS (channel bit-rate): this information is encoded by using the TSPEC and FLOW_SPEC objects (RFC3471, RFC2210, RFC4328) for the downstream and upstream signaling respectively.

Available wavelength along the path: this information is carried by the PATH message by using the LABEL_SET object (RFC3473 [5]) illustrated in Figure 3-9.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Length            | Class-Num (36)|   C-Type (1)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Action     |       Reserved      |          Label Type     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Wavelength 1                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                               :                               :
:                               :                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Wavelength N                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3-9. LABEL_SET object format.

The information carried by the PATH message in order to check the optical feasibility of wavelengths along the path is:

1. Available set of transponders at the source node, for the available wavelengths along the path.
2. Sequence of optical components along the path along with their parameters.
3. The optical parameters of the affected LSPs (i.e. active LSPs sharing some part of their path with the new lightpath).

This information is encoded through new TLVs (type-length-value) carried by the LSP_REQUIRED_ATTRIBUTES object (RFC4420, DRAFT4420bis). The generic encoding of a TLV is showed in Figure 3-10.

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

```
//          Variable length value                       //
|                                                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3-10. TLV format.

In turn, each TLV may contain a sequence of Sub-TLVs, whose generic encoding is illustrated in Figure 3-11.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type        |   Flags     |          Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                        |
//              Value                               //
|                                                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3-11. Sub-TLV format.

For what concerns the description of the path, the PATH message has to carry:

1. The crossed node parameters. Specifically for each crossed node, it is only recorded the type of the optical cross connect (OXC) since every node inside the network stores all the parameters (losses mainly) for each type of OXC (pre-loaded information).
2. For each traversed hop, the parameters of the optical components making up of it. As illustrated in Figure 3-12, a hop between two nodes consists of an output ONIC, an optical link and an input ONIC.

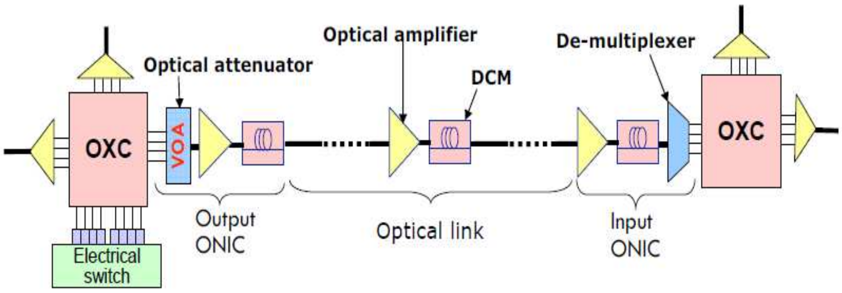

Figure 3-12. General structure of a link connecting two optical nodes.

The parameters required for the affected LSPs and carried in the PATH message are:

- For each hop, the input power at the attenuator
- LSP ID and destination node identification
- Wavelength
- Signal type i.e. the transponder class

For what concerns the RESV messages, they are extended with the optical parameters of the selected wavelength/channel, which are:

- Channel Identifier (wavelength)
- Service Parameters (Rx/Tx Transponder type)

The standard RSVP-TE behavior has also been extended to trigger *Q-CHK-REQ* and *Q-CHK-ACK/NACK* to check the feasibility of affected existing lightpaths when a new lightpath is setup. To avoid potential disruption of existing active lightpaths we have defined a new protocol with the following two messages [8] and is initiated during RESV phase:

- *Q_CHK_REQ* is defined to request the re-computation of Q-factor of existing overlapping lightpaths considering the establishment of new lightpath. It contains both affected and affecting LSP *IDs* and other required information to re-computed the Q-factor considering the new lightpath setup.
- *Q_CHK_RESP* or *Q_CHK_ACK/NACK* is defined to get the response from the affected lightpaths destination node and it only contains whether *Q_CHK_REQ* is successful or not.

3.4. Test cases and simulation objectives

In order to evaluate the performance of the signaling based (RSVP-TE) architecture, we implemented it in a simulation environment developed as an extension of the Java based GLASS (GMPLS Lightwave Agile Switching Simulator) simulator [1] and then a simulation campaign has been conducted. GLASS simulator is developed by NIST, USA using the scalable simulation framework (SSF). It provides network components such as hosts, routers, links, and a number of network protocols. GLASS extends these components with an implementation of MPLS, optical components such as optical cross connects (OXC), edge routers, optical links, fibers, and lambdas. However, the protocols implemented in GLASS do not have support to interact with optical layer. Hence, the protocols included in GLASS simulator suffer from lack of PLI information and corresponding protocol extensions to carry and evaluate the feasibility of lightpaths in optical domain. Here in DICONET project, we have extended GLASS framework to interact with optical layer to get the required information and corresponding protocols to carry this information. In addition several other mechanisms as discussed earlier in this section are implemented. These mechanisms include a) PLI modeling, b) evaluation of Q-factor, c) proper interfaces to Q-tool, d) mechanisms to deal with potential active lightpath disruption, etc., The objective is to check the implementation feasibility of proposed mechanisms and protocol extensions. In addition it is also important to study the performance of different approaches to check the performance for various real-world scenarios to see which protocol extensions works well in those scenarios. In the simulation experiments related signalling based approach and hybrid approach, the following parameters are used:

A statistical processing module ANCLES [2] has been integrated to collect simulation results with target accuracy of *0.05* and confidence level of *95%*. When a node receives a connection request, the *K-CSPF* algorithm computes *K* routes. The Constrained Shortest Path First (CSPF) metric used for route computation is number of hops with $K = 3$. Two kind's of wavelength selection policies are implemented: first-fit and random-fit. In general, each node takes some time to process various RSVP-TE messages and to setup/clear OXC configuration. To simulate real-world scenario, we have considered the following: 1) *PATH* message processing time = *5 msec*; 2) Egress node (where time consuming Q computation carried out) *PATH* message processing time = *80 msec*; 3) *RESV*, *PATH_ERROR*, *RESV_ERROR* message processing time is = 5 *msec*; 4) *Q_CHK_REQ* processing time = *5 msec*; and 5) OXC switch configuration time = *25 msec*. The inter-arrival (I) time is set to 2, 20, and 200 sec. However, the results reported here are only for I = 200 sec. The deadlock mechanism is implemented with timeout value set to 800 msec.

3.4.1. Performance metrics

The parameters used to test the performance of the proposed mechanisms are listed below:

- Blocking probability: The ratio of the number of rejected connection requests to all requested connections.
    - Blocking due to wavelength unavailability called lambda blocking
    - Blocking due to optical unfeasibility (i.e., blocking due to unfeasibility of the lightpath under consideration)
    - Blocking due to affected lightpath disruption (i.e., blocking due to the potential lightpath disruption if the lightpath is accepted into the network)
- LSP set-up time: Elapsed time between the first *PATH* message sent and the *RESV* message received at the source node. This metric reflects how fast a connection request can be established.
- Average number of setup attempts: It is the average number of attempts required for a successfully established lightpath over all connection requests.
- Average number of hops: It is the average number of hops for a successfully established lightpath over all connection requests.

3.4.2. Network topology and characteristics

For what regards the network scenarios, so far we have considered only transparent network where neither wavelength converters nor wavelength regenerators are available in intermediate nodes. Therefore an end to end lightpath must respect the so-called *wavelength continuity constraint.*

The simulation results reported in the Section 3.5 are for the DTAG/T-Systems National Core Network [24] whose topology is depicted in Figure 3-13 while its fundamental characteristics are also summarized in Table 3.1.

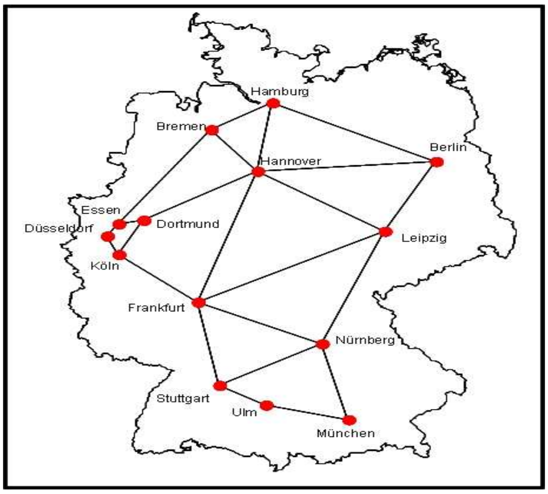

Figure 3-13. DTAG/T-Systems National Core Network.

Table 3.1. Parameters of DTAG/T-Systems National Core Network.

Parameter	Value
Number of Nodes:	14
Number of links:	23
Node degree:	3.29 (min. = 2, Max. = 6)
Link length (Km):	186 Km (min. 37, Max:353 Km)
Span length (Km):	25 Km (Long Aggr.), (min.:0.21, Max:65) 97 Km (Short Aggr.), (min: 37, Max.:149)
Number of Spans:	7.5 (Long Aggr.), (min: 1, Max:17) 1.87 (Short Aggr.), (min: 1, Max: 3)
Path length (Km):	410 Km (min.:37, Max.:874)
Hop count:	2.35 (min:1, Max:5)

3.4.3. Traffic models

Simulations have conducted by employing a dynamic traffic model where lightpath set-up requests are generated according to a Poisson process. Following this model, lightpath requests arrive at an average rate of $1/\alpha$, where α is the average request inter-arrival time expressed in seconds. The traffic requests considered in simulations are uniformly distributed among all the nodes.

An installed lightpath has an average exponential duration of μ seconds.

The traffic load ρ is defined as the average network resource (link wavelength) usage computed in percentage as:

$$\rho = \frac{\overline{N}_c \times \overline{L}_c}{M \times W} \times 100\%$$

where $\overline{N}_c$ is the average number of active connections and equals μ/α, $\overline{L}_c$ is the average number of hops in the network (considering only shortest paths between nodes and a uniformly distributed traffic matrix), M is the number of network links and W is the number of wavelengths inside a fiber.

3.4.4. Q-tool and Q-factor computation

In this deliverable, we use a temporary implementation of Q-tool while we wait for the complete implementation and integration of Q-tool with the signaling based architecture. The results and performance study of this deliverable are updated later with the final implementation/integration of Q-tool in Jul/Aug 2009. In the temporary Q-tool, the Q-factor is computed as follows: $Q = 10 \log_{10}[1/N_{tot}]$, where N_{tot} = single channel effects + multi channel effects. The single channel effects = a * l, where as multi channel effects

$$= b * \sum_{i=1}^{N=Numhops} l_i * N_i \big| \lambda = \pm k$$

Where a and b are computed using reverse engineering, i.e., take for example a value of Q_{thr} = 18 and L_{max} = 1000; then 18 = 10 $\log_{10}[1000/N_{tot}]$, which gives N_{tot} value. Now assume that b = 0 and then we get a = N_{tot}/l. Note that b can be adjusted to tune the multichannel effects in this fake implementation. In the simulation we set *k = 4*.

3.5. Performance study

The results from the simulation experiments and a performance comparison of both signalling based and hybrid approaches are discussed in Section 4.3.

4. Hybrid Architecture

The hybrid approach is based on both extensions of the routing and the signaling protocols. The rational behind the hybrid architecture is to improve the routing computation phase in order to minimize the likelihood of performing several attempts during the signaling phase before setting up the optical channel. The PLI information carried in the PATH messages by RSVP-TE messages may be accurate if there are very frequent changes in the network or due to concurrent lightpath requests. However, in present optical networks the frequency of lightpath arrivals and departures to the network is not too high. Hence, it is important to consider these factors when designing and developing the hybrid architecture.

4.1. Protocol extensions for carrying PLI-information

In hybrid architecture both RSVP-TE and OSPF-TE protocols are extended. As discussed earlier these protocols need to be extended considering various factors such as, the overall network performance, overall control overhead, number of protocols and number of changes needed in each protocol, the deployment and standardization considerations, stability and robustness of various databases. As most of the impairment information that is required to compute Q is carried using RSVP-TE, the main information that can be useful to make intelligent decisions is wavelength availability information. OSPF-TE is extended correspondingly to carry this information. OSPF-TE can also be extended to carry other PLI information, however, given the dynamicity of connections (connections arrive and depart from the network at large time scales) in the current optical networks that may not be needed. Hence, for this situation it may not be required to disseminate PLI information using OSPF-TE. As the dynamicity is less, the frequency of changes in PLI and logical topology information is also less. However, if wavelength availability information is available at the time of route computation, we can implement intelligent RWA algorithms and hence can reduce number of attempts leading less setup time. Hence, the DICONET consortium decided to make minimal changes to OSPF-TE considering the above mentioned factors.

The detailed extensions RSVP-TE protocol is discussed in Section.3. In the following we discuss, the changes made to OSPF-TE to realize hybrid approach. In the hybrid-approach an extended Constrained Shortest Path First (CSPF) algorithm is implemented considering the wavelength availability constraint in the route selection. The information related to links wavelength availability is distributed to all network nodes by mean of the OSPF-TE protocol [RFC 3630 [26]] introducing a new TE-Link Sub-TLV [RFC 4203[25]] that lists the available wavelengths in the link. A proposed encoding of the information is defined in section 3.4 of draft-ietf-ccamp-rwa-wson-encode-01 [27]. However, the implementation in GLASS simulator is as follows. The available wavelengths list is simply a 'list' without real encoding in GLASS. The LSA is updated and flooded to all nodes at each link status change (wavelength allocation/deallocation). The update frequency of the information could be an issue leading to some discussions. It can be trimmed (e.g. no new update in X ms) as it is currently done for the 'available bandwidth' info, but the routing constraint is much more sensitive to errors w.r.t. the wavelength availability information than in the bandwidth availability information. The wavelength availability is used by the routing to validate a route and to provide the initial set of usable wavelengths to the signaling (the actual wavelength selection however is made anyway in the signaling phase). It is important to note that providing the initial set of wavelength is just a way to reduce the amount of info carried by signaling messages (as PATH message need to carry PLI information regarding the available wavelengths). It works the same even with no initial information provided and the 'standard' dynamic wavelength selection at the destination node using wavelengths collected in PATH message.

4.2. Test cases and simulation objectives

The test case and simulation objectives are same as discussed in Section 3.3.

4.3. Performance study of signalling & hybrid approaches with simple Q-tool

Figure 4.1 shows blocking probability (BP) vs. load for first-fit and random-fit for both signalling based approach and hybrid approach. Note that BP has the following main contributions: 1) blocking due to wavelength continuity constraint called wavelength blocking, i.e., no free common wavelength along the route (as shown in Figure 4.2); 2) optical failures due to the unfeasibility of connection request under consideration (as shown in Figure 4.3); 3) optical failures due to the active connections, i.e., setting up of new lightpath may introduce excessive crosstalk on active connections potentially disrupting them, hence the new connection need to blocked (as shown in Figure 4.4); and 4) due to deadlocks as explained earlier. BP is lower in case of hybrid approach due to exact information regarding wavelength availability which reduces number of attempts below K = 3 before a successful feasible path is found as shown in Figure 4.2. However, in case of signalling based approach, as there is no wavelength availability information during route computation time, the route is computed only based on shortest path. When the path message reaches the destination nodes, it realizes that there is no common free wavelength along the path leading to more number of attempts ultimately reaching the maximum number of allowed attempts (which in this case K = 3) and hence more blocking due to wavelength blocking as shown in Figure 4.2 leading to higher blocking probability. As the load increases the BP for both architectures increases because of the following reasons: 1) the number of active connections in the network increases and hence the chances of finding a free common wavelength along the route decreases leading to higher wavelength blocking as shown in Figure 4.2 (which is also observed in results presented in Fig 4.9 and 4.10); 2) as the number of active connection increases in the network the effect of these active lightpaths on the new lightpath requests increases (due to multi-channel effects) leading increase in blocking due to the optical failures as shown in Figure 4.3; 3) as the active connections in the network increases the number of lightpaths that may be potentially disrupted due to the setup of new connection increases leading to increase blocking due to affected lightpaths as shown in Figure 4.4. Note that the blocking due to optical failures is higher in case of hybrid approach as the average path length/number of hops in this case is higher compared to signalling based approach due the computation of route using CSPF with wavelength availability information. However, the wavelength blocking in case of signalling based approach is higher compared to hybrid approach as signalling based approach computes shortest path without wavelength availability information and when the PATH message reaches the destination it realizes that there is no common free wavelength leading to more attempts and higher wavelength blocking as shown in Figure 4.2. The blocking due to deadlocks in all architectural scenarios is zero for inter-arrival time, I = 200. However, it has a very small, but, nonzero value (not shown in this report) having very little impact as the inter-arrival time (arrival rate) decreases (increases), i.e., for I = 2 and 20. The percentage of blocking due to deadlocks is more in case of signalling based approach compared to hybrid approach. The reason is: in signalling based approach there is high wavelength blocking leading to more setup attempts, which increases the chances of deadlocks due to presence of more PATH messages in the network.

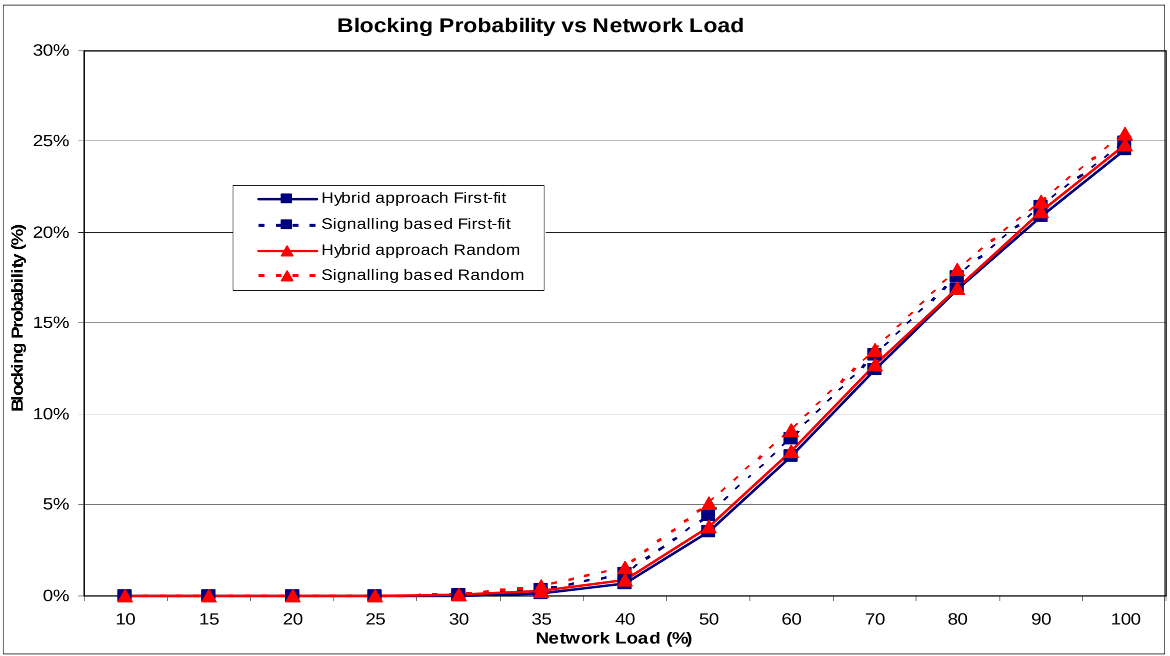

Figure 4-1 Blocking probability vs. network load for RSVP-TE and hybrid approach

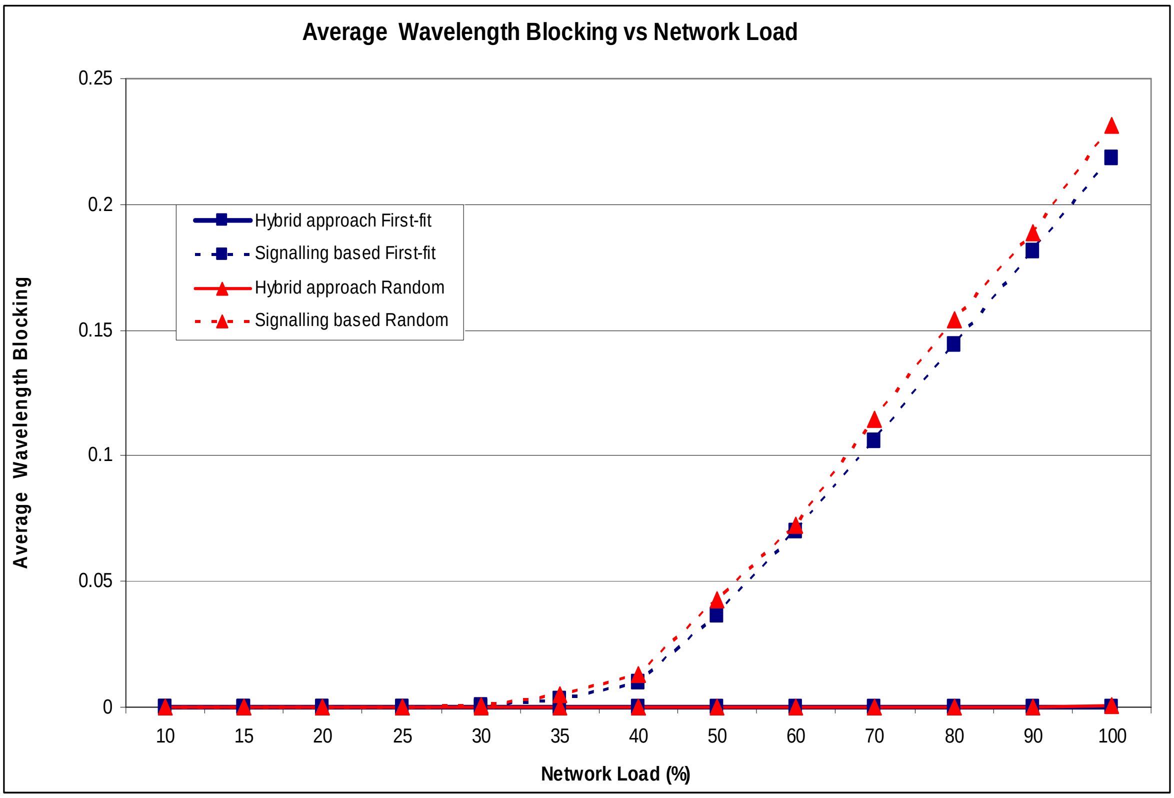

Figure 4-2 Percentage blocking due to wavelength unavailability vs. network load for RSVP-TE and hybrid approach

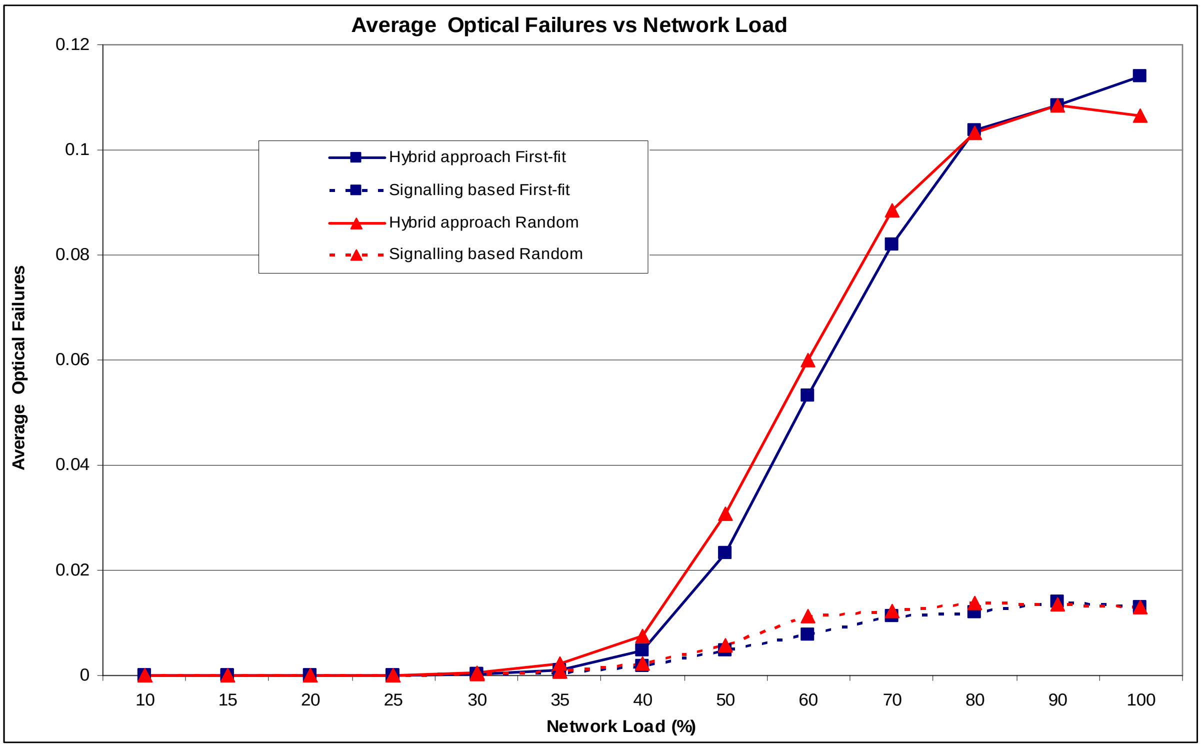

Figure 4-3 Percentage of optical failures (due to unfeasibility of lightpath itself) vs. network load for RSVP-TE and hybrid approach

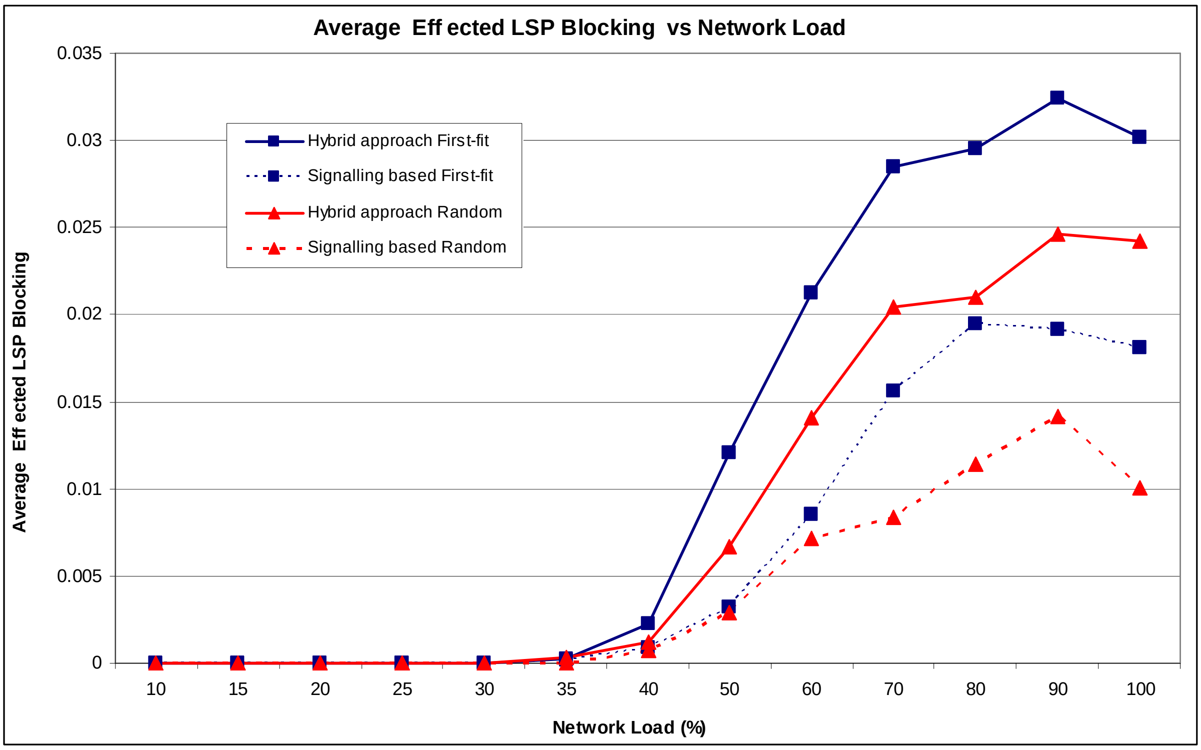

Figure 4-4 Percentage of blocking due to active (potentially affected) LSPs vs. network load for RSVP-TE and hybrid approach

The average setup time depends on the average number of attempts and average number of hops in the route. As the average number of hops in the route is higher in case of hybrid approach (as shown in Figure 4.7) the average setup time is high, as shown in Figure 4.5. It also depends on the network and load, which needs to be examined with final Q-tool implementation. The wavelength availability information in hybrid approach reduces the wavelength continuity constraint failures at the intermediate and destination nodes due to wavelength unavailability leading to less average number of attempts, as shown in Figure 4.6. As the load increases the average number of hops decreases, as the chances of finding a free common wavelength on longer hop paths decreases leading to decrease in lower setup time. The average number of attempts in case of signalling based approach is higher due to lack of wavelength availability information and it increases with the load due to increase in both lambda failures and both kinds of optical failures.

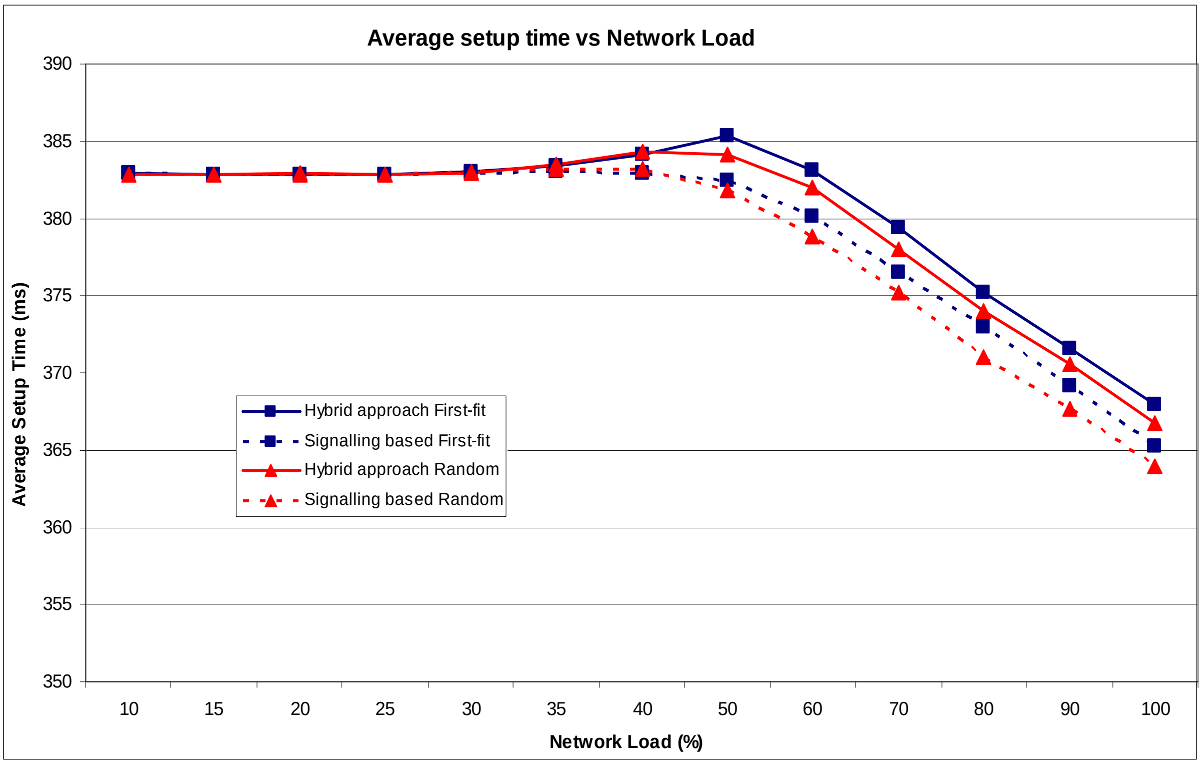

Figure 4-5 Average setup time vs. network load for RSVP-TE and hybrid approach

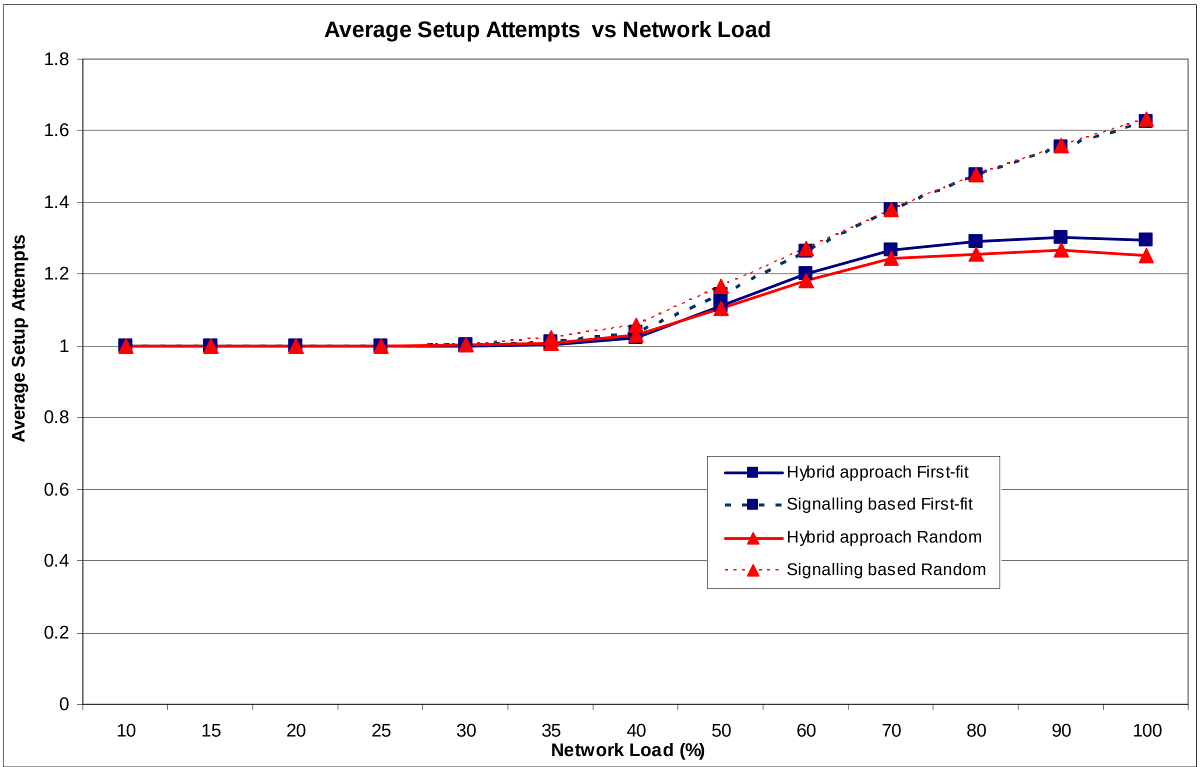

Figure 4-6 Average number of setup attempts vs. network load for RSVP-TE and hybrid approach

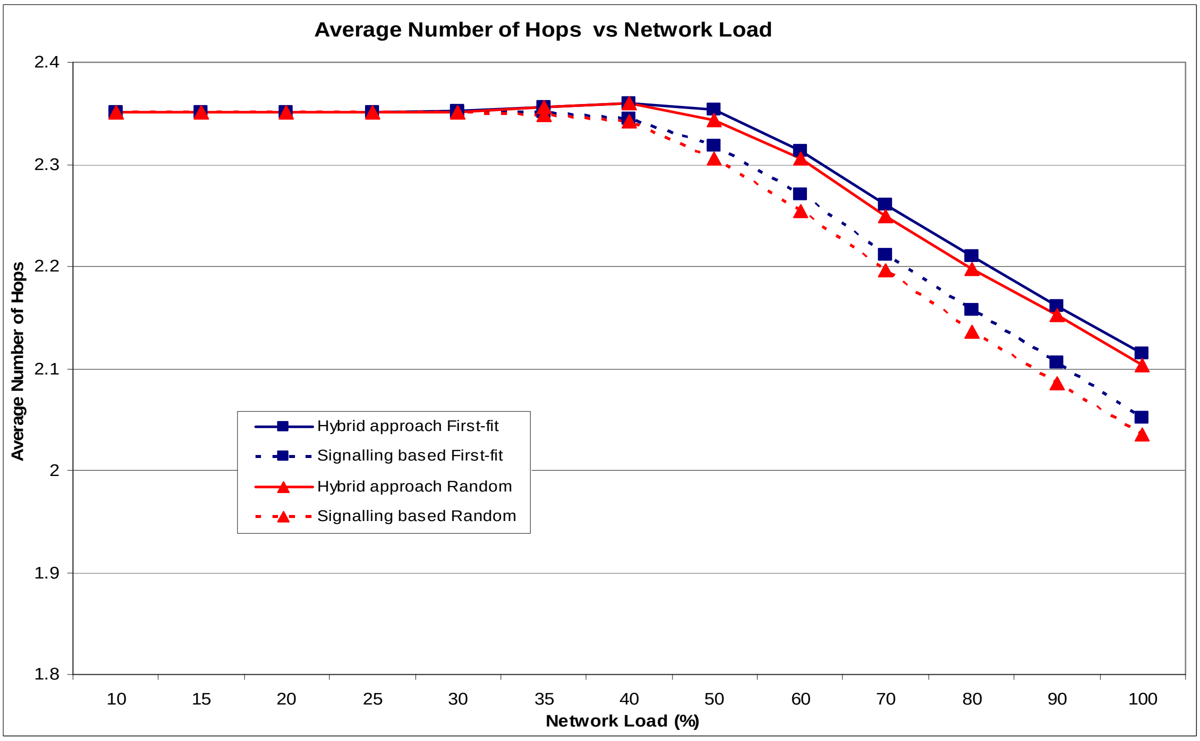

Figure 4-7 Average number of hops vs. network load for RSVP-TE and hybrid approach

The average number of PATH messages sent by each node in the network again depends on the average number of hops and average number of attempts. Note that PATH messages counted are not per LSP, but, the PATH messages sent by each node. As the average number of hops in hybrid approach is higher (due to the reason that the longer paths are more affected by wavelength unavailability failures in the signalling based approach), the average number of PATH messages is high as shown in Figure 4.8.

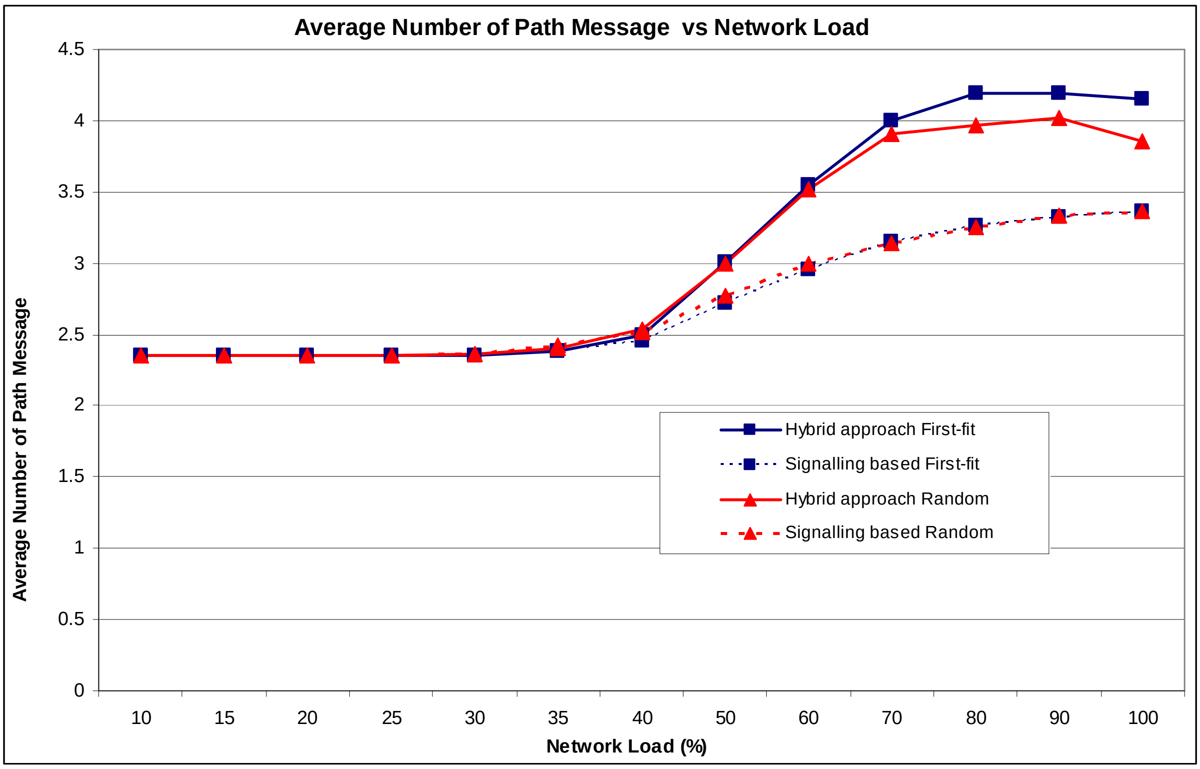

Figure 4-8 Average number of PATH messages vs. network load for RSVP-TE and hybrid approach

The blocking probability vs. hop count for various approaches (for load = 60) is shown in Fig. 4.9. The corresponding percentage of different hop count requests are shown in Fig. 4.10. As it is expected that there very less number of connection requests generated in simulation studies as the hop count increases. The blocking probability for large hop count connections is higher than smaller hop count connections as expected for all approaches. However, the BP for signalling approach is higher compared to hybrid approach due to lack of wavelength availability information during the route computation phase.

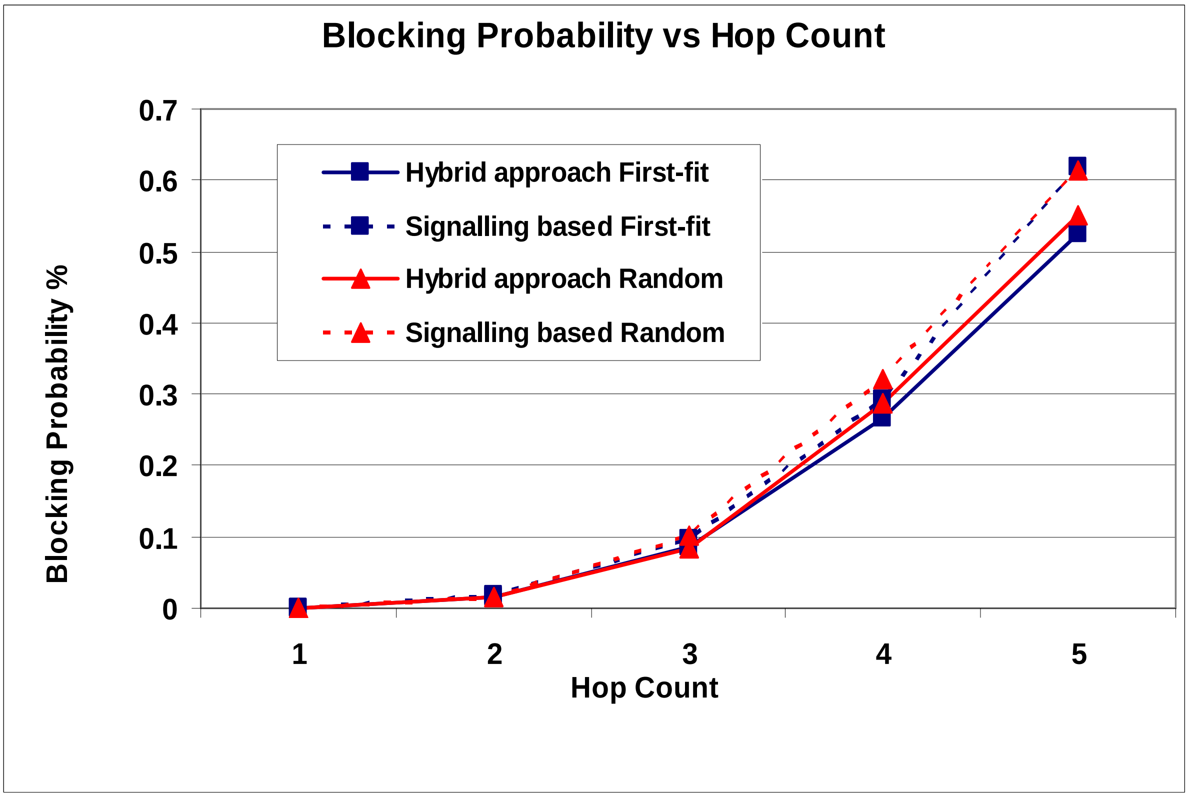

Figure 4-9 Blocking probability vs. hop count for RSVP-TE and hybrid approach (load = 60)

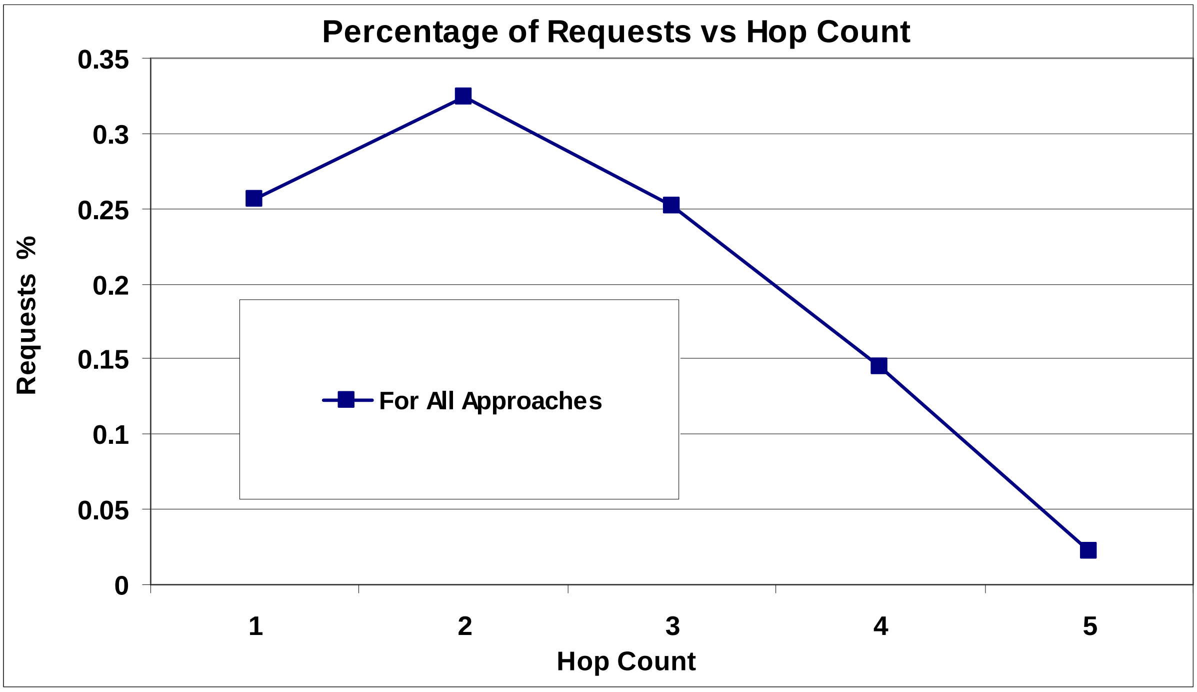

Figure 4-10 Percentage of connections vs. hop count for all approaches

4.4. Performance study of signalling & hybrid approaches with final Q-tool

The final Q-tool from WP3 is integrated with signalling and hybrid control plane architectures. We have conducted extensive simulation experiments after integration to study the performance of signalling and hybrid architecture. As the computation of Q-factor using Q-tool takes longer time, we have run the simulation experiments for the DTAG network described in Sections. 3 and 4 using 8 number of wavelengths, first-fit wavelength selection policy, and inter arrival time = 2.0 msec. All other parameters are the same as described in Section 3 and 4. In the following we compare the simulation results for signalling and hybrid approaches using simple Q-tool (Section. 3.4) and final Q-tool from WP3 for comparison purposes though simple Q-tool does not take into account all real PLI models.

Blocking probability vs network load for both approaches using simple and real Q-tool is shown in Figure 4-11. The blocking probability for hybrid approach is lower than the signalling based approach as expected. As the load increase the blocking probability of both approaches increases.

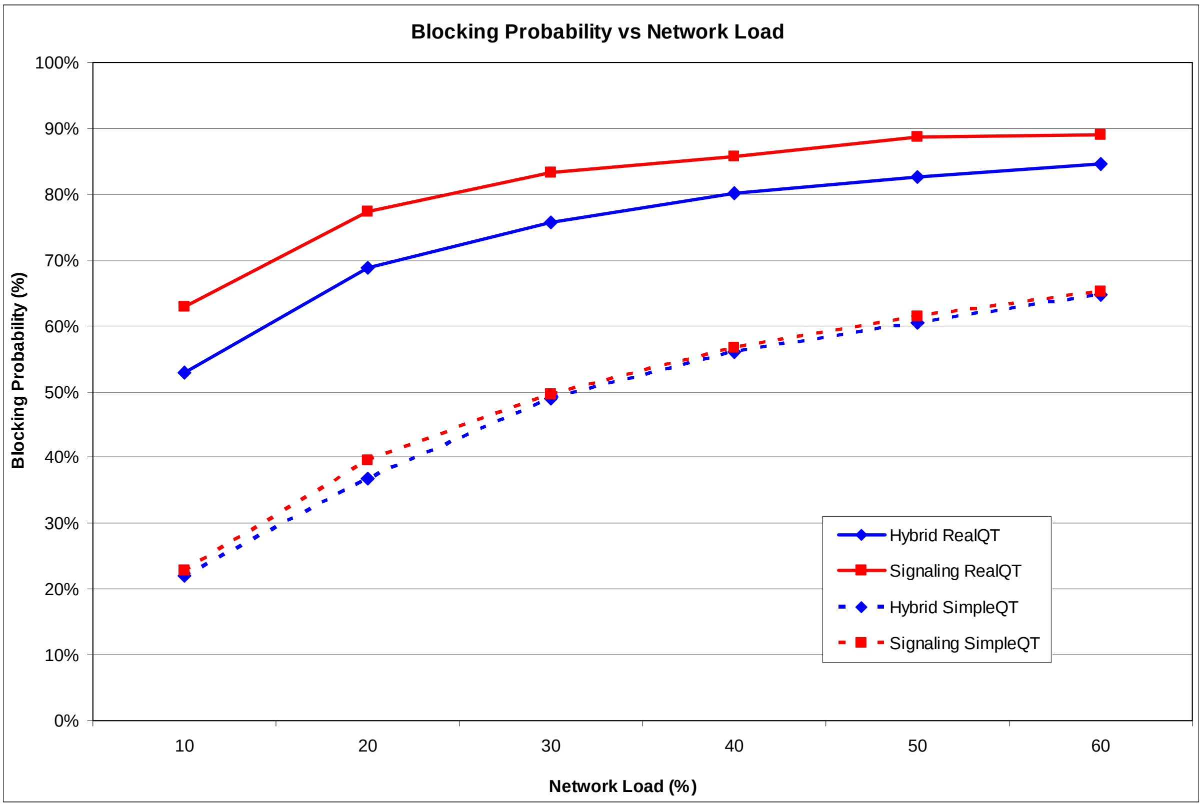

Figure 4-11 Blocking probability vs. network load for RSVP-TE and hybrid approach

Wavelength blocking vs network load for different approaches is shown in Figure 4-12. As wavelength availability information is disseminated using extended OSPF-TE in hybrid approach, the wavelength blocking is zero. However, as the load increases the wavelength blocking of signalling approach increases slightly due to unavailability of common free wavelength along the route.

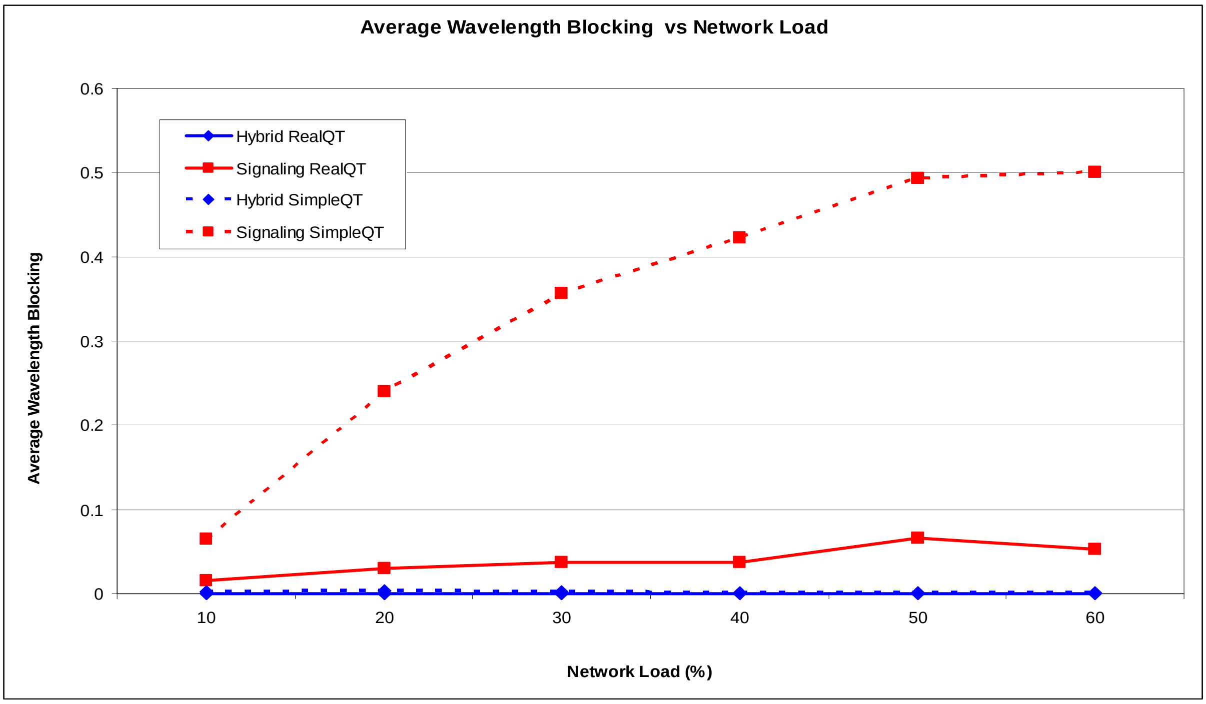

Figure 4-12 Percentage blocking due to wavelength unavailability vs. network load for RSVP-TE and hybrid approach

Average blocking due to optical impairments vs. network load is shown in Figure 4-13. As PLIs are evaluated at the source node in case of hybrid approach, the blocking due to optical reasons is negligible. Whereas in signalling approach at source node PLIs information is not available. Hence it computes the path without the knowledge of PLIs. When the PATH message reaches the destination node there are some failures due to optical reasons.

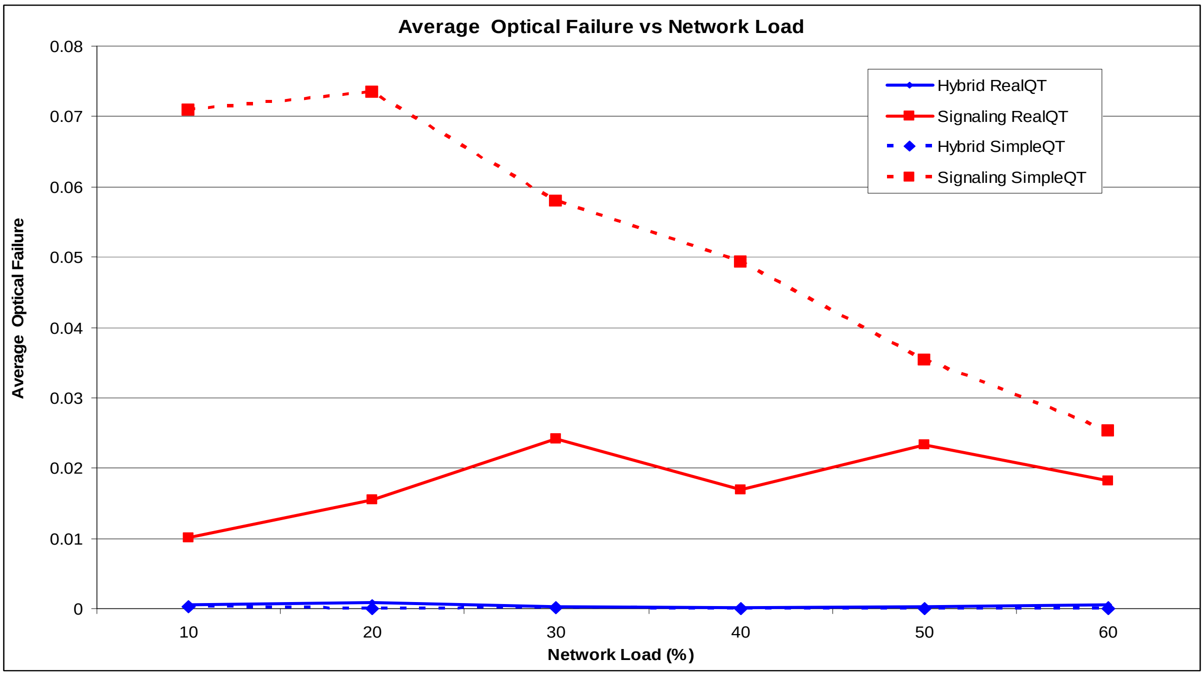

Figure 4-13 Percentage of optical failures (due to unfeasibility of lightpath itself) vs. network load for RSVP-TE and hybrid approach

Blocking due to affected LSP vs. network load is shown in Figure 4-14. In signalling approach, this kind of blocking is too high as establishment of new lightpath potentially disrupts active LSPs. Whereas the affected LSP blocking in case of hybrid approach is almost negligible due the feasibility evaluation at the source node itself.

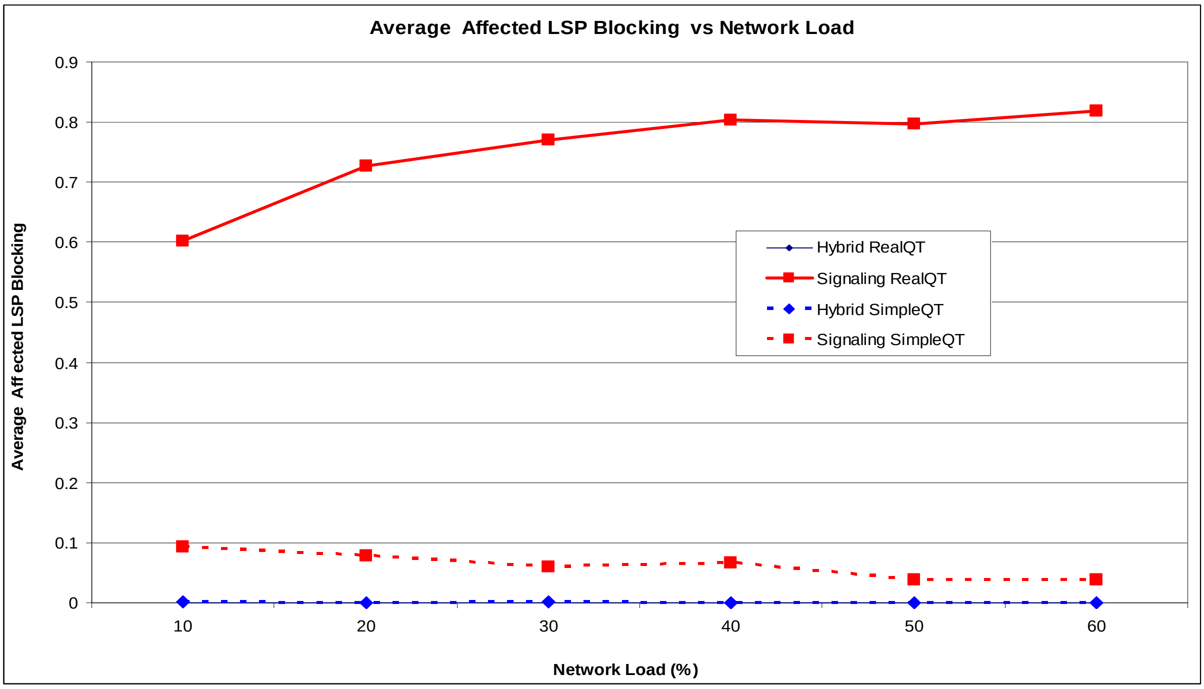

Figure 4-14 Percentage of blocking due to active (potentially affected) LSPs vs. network load for RSVP-TE and hybrid approach

The blocking due to route failures due to unavailability of route is shown in Figure 4-15. In hybrid approach this kind of blocking is too high, as many requests are failed at the source node. As the wavelength availability information and PLI information is available at the source node, it computes the path on which free common wavelength is available. If there is no free continuous available wavelength from source node to destination nodes the requests are blocked. Hence this kind of blocking is high. Whereas in signaling based approach, k-shortest paths are computed at the source node. However, wavelength continuity constraint is checked during the signaling phase and feasibility evaluation is done at the destination node. Hence, blocking due to route failures is less in case of signaling approach. Note that overall blocking is mainly due to the above four contributions. Apart from the there could also request failures due to deadlock mechanism implemented. However, the failures due to deadlocks are almost negligible in all cases.

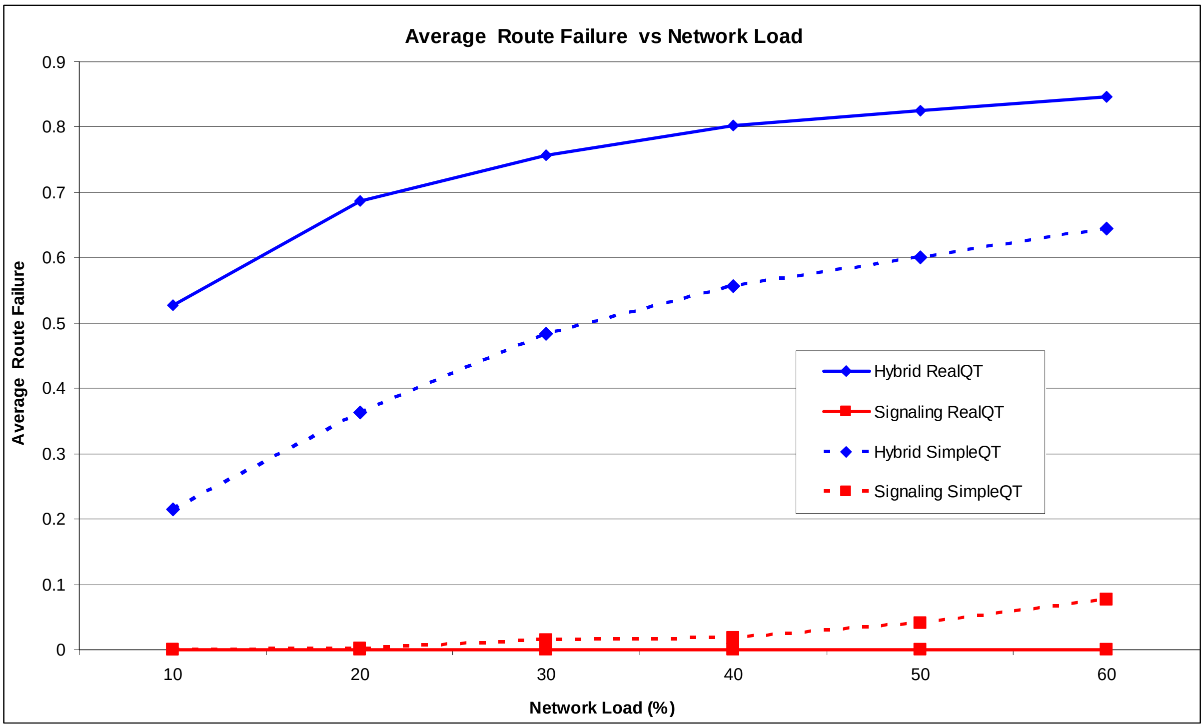

Figure 4-15 Percentage of blocking due to route failures vs. network load for RSVP-TE and hybrid approach

Average setup time vs network load is shown in Figure 4-16. As explained earlier average setup time mainly depends on average number of hops and average number of setup attempts. Average setup time for hybrid case is slightly lower than signalling case. In hybrid case the average number of attempts is 1 and average number of hops is slightly higher than that of signalling approach due to longer routes because of using wavelength availability information in route computation. Whereas in signalling approach average number of attempts is higher and average number of hops is lower than that of hybrid approach. See Figures 4-17 and 4-18 for average number of setup attempts and average number of hops. Note that average setup time is computed only for accepted connections. Though we did not show here, the total number of connections accepted and the total setup time in case of hybrid case higher than that of signalling case.

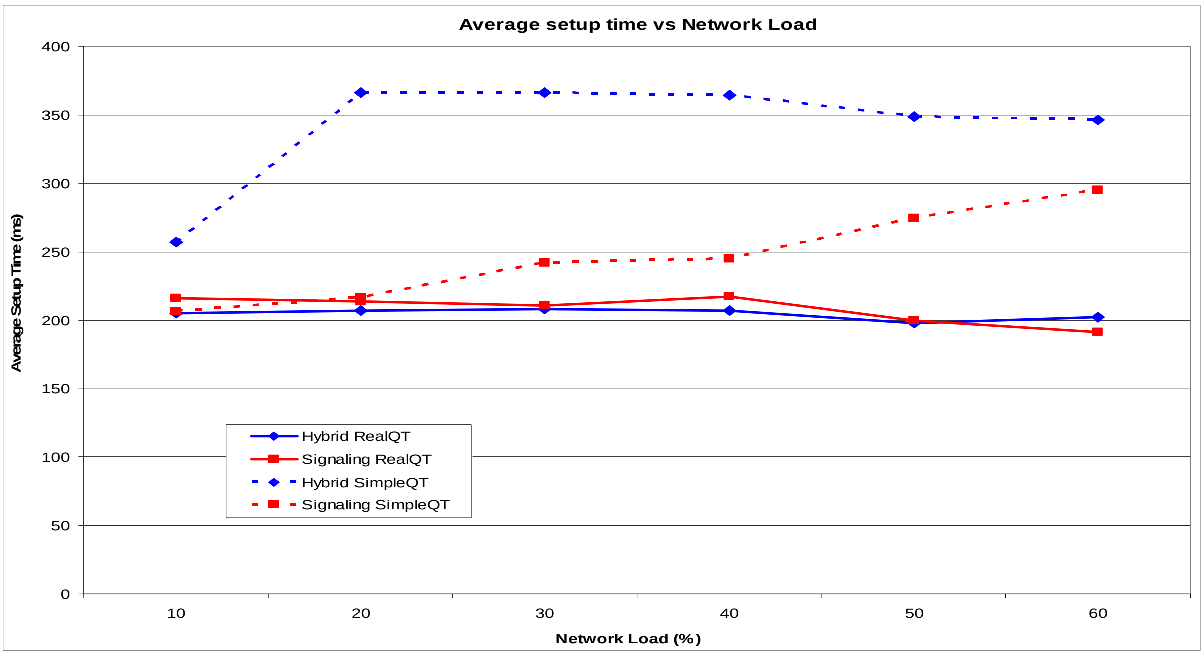

Figure 4-16 Average setup time vs. network load for RSVP-TE and hybrid approach

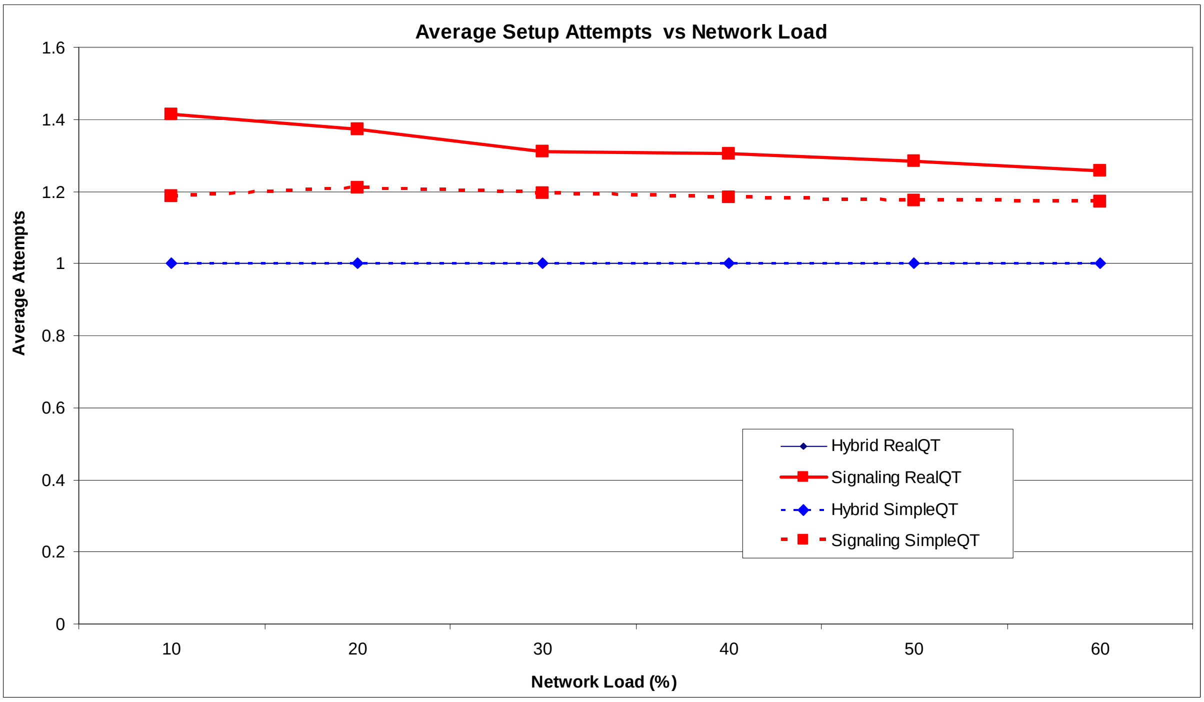

Figure 4-17 Average number of setup attempts vs. network load for RSVP-TE and hybrid approach

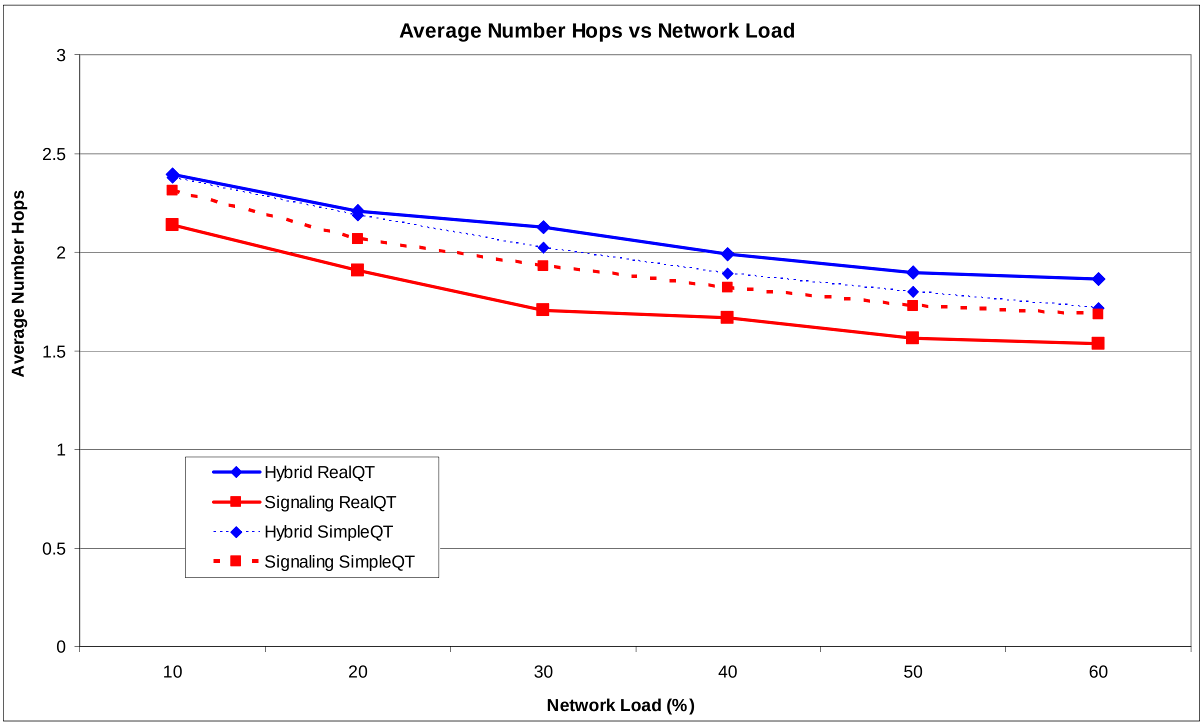

Figure 4-18 Average number of hops vs. network load for RSVP-TE and hybrid approach

The average number of PATH messages vs. network load is shown in Figure 4-19. It depends on the average number of hops and average number of attempts. PATH messages are counted at each hop. Hence it is basically multiplication of both average number of hops and average number of setup attempts. See Figures 4-17 and 4-18 for average number of attempts and average number of hops.

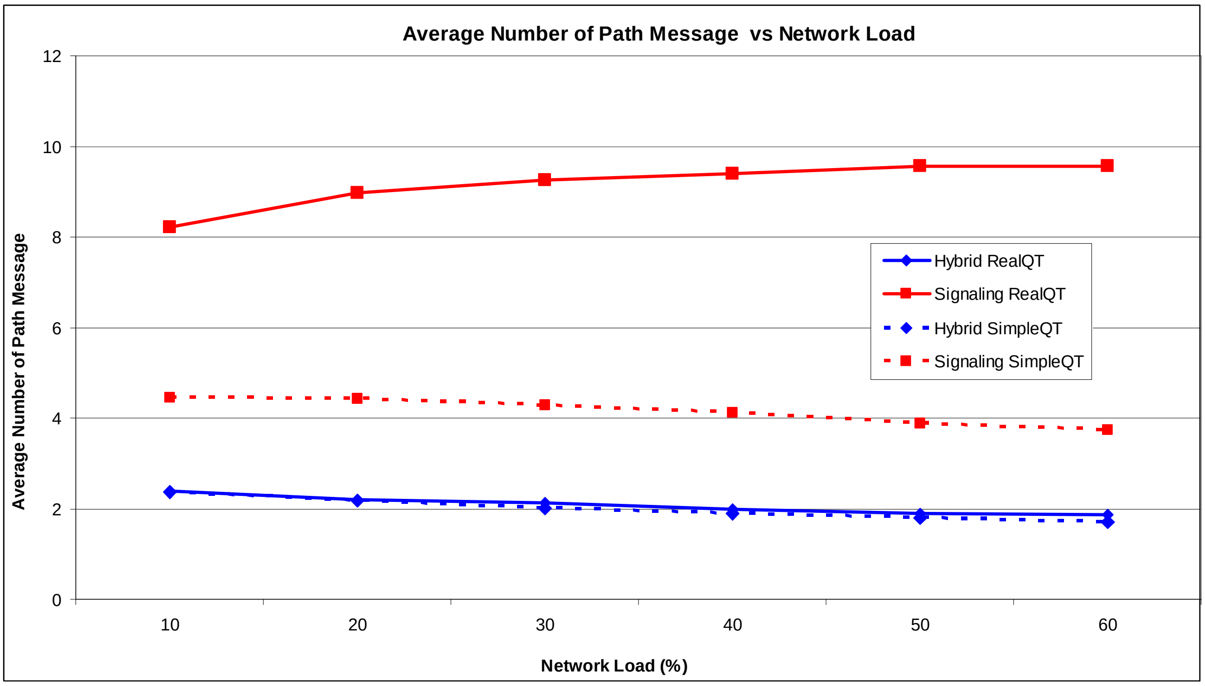

Figure 4-19 Average number of PATH messages vs. network load for RSVP-TE and hybrid approach

Blocking probability vs. hop count for load=60 is shown in Figure 4-20. It shows that blocking for longer hops LSPs is higher compared lower hop LSPs. This mainly due to unavailability of wavelengths or affected LSP blocking. Blocking probability in case of hybrid in slightly lower than that of signalling approach for a given LSP length. This is mainly due to the limit on number of shortest paths (K=3). In hybrid case either it finds a path with continuous wavelength or rejects the connection. However, in signalling case the paths checked for wavelength continuity constraint during signalling phase and optical feasibility is evaluated at the destination nodes. So, the number of attempts required may be higher than k=3 resulting in more blocking.

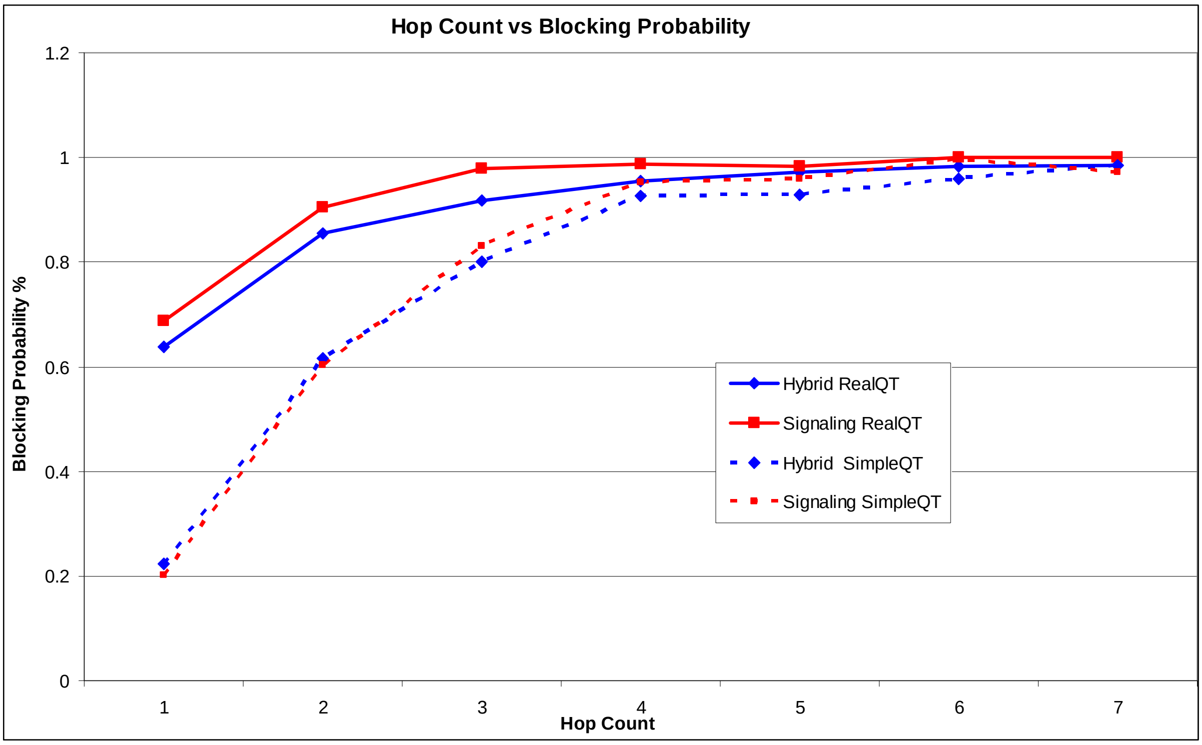

Figure 4-20 Blocking probability vs. hop count for RSVP-TE and hybrid approach (load = 60)

5. PCE based architecture

The goal of this section is to study the PCE architectures for considering physical layer impairments. In Section 5.1, we first analyze the state of the art of PCE architecture, and identify two possible approaches to include physical layer impairments by extending either OSPF-TE or PCEP protocols. In Section 5.2 the protocol extensions for these two approaches are given accordingly. Section 5.3 defines the test cases and simulation objectives. Section 5.4 provides the simulation results.

5.1. Introduction and basic architecture

The PCE is an entity which takes care of all routing computations inside an Autonomous System (AS). The PCE is described in RFC 4655. It can be located in a single node, or distributed among multiple nodes for scalability and reliability. Typically, the PCE has a Traffic Engineering Database (TED) which it uses to store network information. All other elements, such as NMS, normal nodes in the network, etc. can request PCE to compute a route. In such case, those elements can be called PCC (Path Computation Client). Computation is done upon request from the Path Computation Client (PCC) using the PCE communication protocol (PCEP). PCEP is used for both PCE-PCE and PCE-PCC Communications. The specification of PCEP is detailed in RFC 5440[9]. The progress of setting up a connection can be as following (see Figure 5.1):

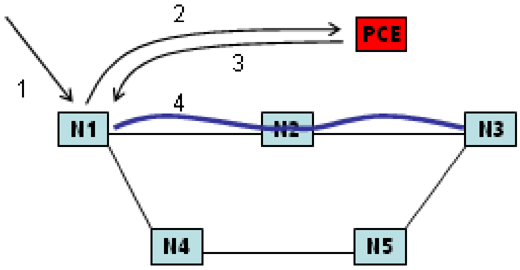

Figure 5-1The progress of setting up a connection

1) Head node (N1) receives a request to set up a connection to N3;
2) N1 plays PCC role, builds a PCReq (Path Computation Request) message, and sends it to PCE to request an available route from N1 to N3;
3) PCE receives the PCReq message, tries to compute an available route, and returns the computation result to PCC with a PCRep (Path Computation Reply) message;
4) N1 receives the PCRep message, sets up a LSP following the route carried in this message.

PCEP is based on TCP. TCP port 4189 has been assigned to this protocol by IANA. PCE uses this port to listen to TCP connection request from PCCs. If the TCP connection has been accepted and set up successfully, a PCEP session can be set up based on this TCP connection. After the PCEP session has been setup successfully, PCC can request PCE to compute a route.

PCE functions can be added to a normal node, or realized in a separate PC server. The basic PCE architecture is shown in Figure 5.2.

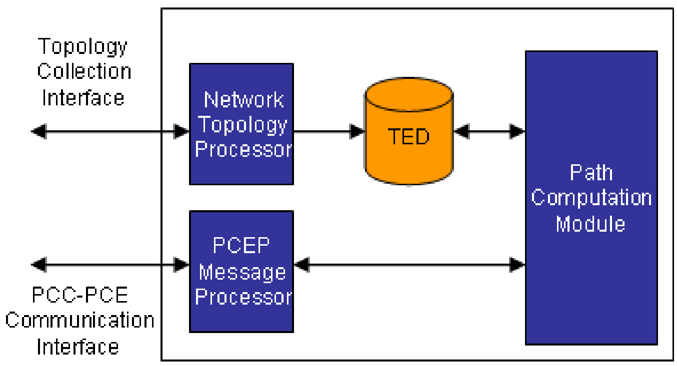

Figure 5-2 PCE basic architecture

There are two main interfaces in PCE. One is Topology Collection Interface and the other one is PCC-PCE Communication Interface. PCE gets network topology information from Topology Collection Interface and stores it to Traffic Engineering Database (TED). This interface can be realized through GMPLS OSPF-TE. The TED, and the path attributes received from PCC-PCE Communication Interface, can be provided to Path Computation Module to compute an available route. The functions of PCEP Message Processor are:

- Listening to TCP connection requests from PCCs;
- Processing PCEP session messages;
- Receiving and parsing PCReq messages coming from PCCs, building and sending PCRep messages to PCCs based on computation results.

A PCEP message processor module should be added to the nodes that play PCC role. The functions of this module are:

- Starting the progress of setting up a TCP connection with PCE;
- Processing PCEP session messages;
- Building and sending PCReq message to PCE based on LSP attributes
- Receiving and parsing PCRep message coming from PCE.

5.1.1. Approaches for considering PLI information

According to the basic PCE architecture in Section 5.1, two protocols can be extended to include physical layer impairments. One is the OSPF-TE routing protocol and the other one is the PCEP protocol. In order to store impairment information there is a need for a new database. PPD will be introduced in the nodes and all the impairments-related information will be stored in it. Monitors will also report the detected impairments parameters to PPD through the monitor-control plane interface.

The PCE has to incorporate two engines. One for routing LSPs in the virtual topology using some constraint based routing algorithms and using the OSPF-TE TED are input for computations, and another for computing lightpath setup by means of Impairment Aware RWA in the optical layer, using the PPD as input. Both engines may have some policy implemented. In the IP/MPLS layer, the TED needs to be kept synchronized between all routers in the network and the PCE. The tacitly assumed dissemination method is using OSPF-TE.

In order to enable IA-RWA, the PCE should be stateful with regard to its optical layer component, meaning that all details on all lightpaths in use in the network must be maintained in this PPD database. This is in order to be able to incorporate complex inter-channel effects such as four wave mixing (FWM) and Cross-Phase Modulation (XPM) into the IA-RWA computations. The IP/MPLS layer information can be stateless, keeping only track of the total resource usage and other aggregated metrics in order to improve scalability. The number of Label Switched Paths in the IP/MPLS layer is typically much larger than the number of lightpaths in the optical layer.

5.1.2. OSPF-TE extension approach

Figure 5.3 shows the PCE architecture by extending OSPF-TE protocol to disseminate the physical layer impairments information. Each node floods the information from TED and PPD by using the extended OSPF-TE protocol. The TED and PPD in PCE are updated by the extended OSPF-TE protocol. Standard PCEP is used for communications between PCC and PCE. PCE communicates to the network planning & operation tool (NPOT) for routing and wavelength selection as well as lightpath quality checking. If the lightpath can be established, standard RSVP-TE signaling protocol is used for lightpath setting up.

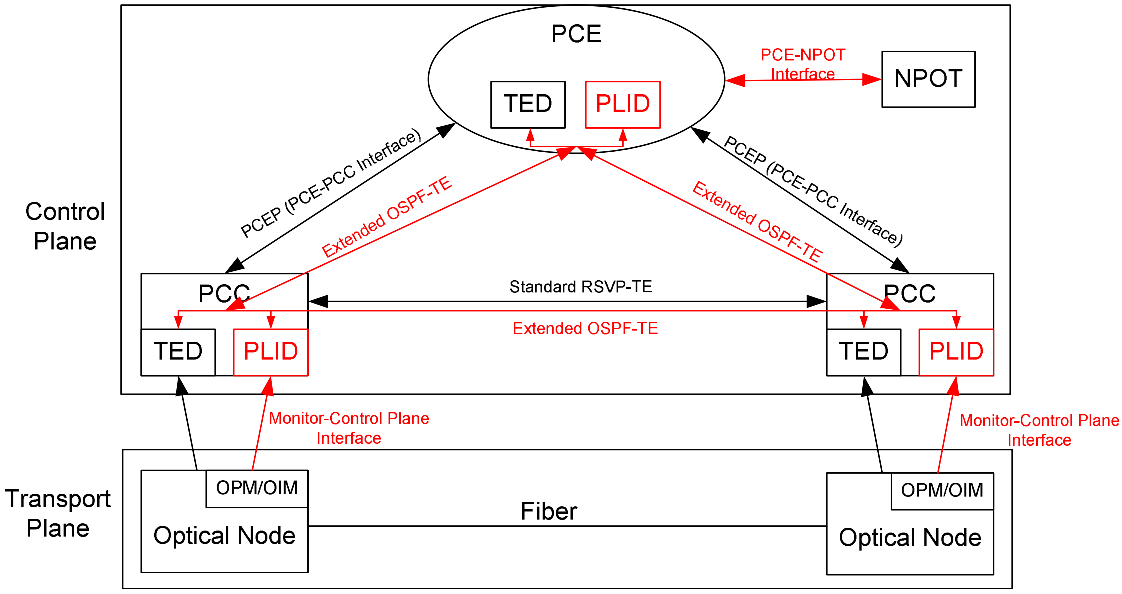

Figure 5-3 OSPF-TE Extension for Including Physical Impairments in PCE Architecture

5.1.3. PCEP extension approach

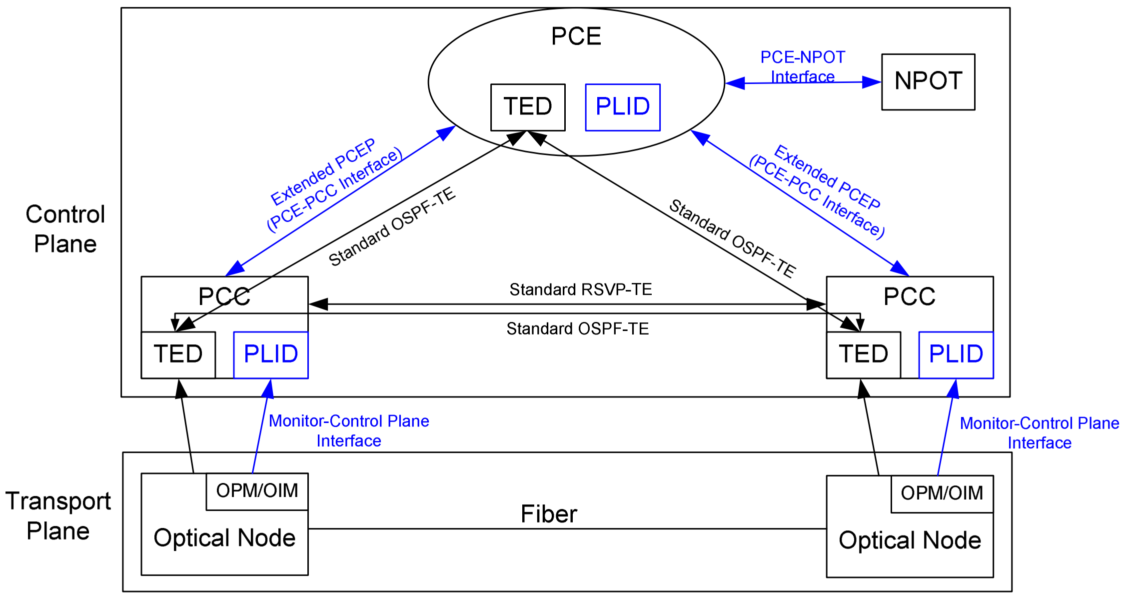

Figure 5-4 PCEP Extension for Including Physical Impairments in PCE Architecture

Figure 5.4 shows the PCE architecture by extending PCEP protocol to disseminate the physical layer impairments information. PCC updates the PPD information to PCE by using the extended PCEP protocol, while standard OSPF-TE protocol is used to flood the TED information. Standard RSVP-TE signaling protocol is used for lightpath setting up.

5.2. Protocol extensions for carrying PLI-information

In this sub-section, we define the protocol extensions for carrying PLI-information for both OSPF-TE extension and PCE extension approach.

5.2.1. OSPF-TE extension

5.2.1.1. 10# LSA introduction

IETF RFC2370 defines enhancements to the OSPF protocol to support a new class of link-state advertisements (LSA) called Opaque LSAs. Opaque LSAs provide a generalized mechanism to allow for the future extensibility of OSPF. Opaque LSAs consist of a standard LSA header followed by application-specific information. The information field may be used directly by OSPF or by other applications. Standard OSPF link-state database flooding mechanisms are used to distribute Opaque LSAs to all or some limited portion of the OSPF topology. Opaque LSAs are types 9, 10 and 11 link-state advertisements. IETF RFC3630 defines the Traffic Engineering LSA using type 10 LSA. In our OSPF-TE extension proposal, we use this type 10 LSA to carry and disseminate the impairment information. The header of Traffic Engineering LSA (a kind of 10# LSA, defined in RFC3630) is as follows (Figure 5.5):

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             LS age            |    Options    |      10       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       1       |                 Instance                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-5 Traffic Engineering LSA header (The red portions identify an LSA uniquely)

A Traffic Engineering LSA contains one top-level TLV. There are two kinds of top TLVs defined in RFC3630, i.e. router address top TLV and link top TLV. A top TLV contains a set of sub-TLVs. There are many kinds of sub-TLVs defined in RFC3630/ RFC4203. The Length field of a TLV defines the length of the value portion in octets (thus a TLV with no value portion would have a length of zero). The TLV is padded to four-octet alignment; padding is not included in the length field (so a three octet value would have a length of three, but the total size of the TLV would be eight octets). Nested TLVs are also 32-bit aligned. Unrecognized types are ignored.

The above rules also apply to the following extensions. It is allowed that using multiple LSAs to describe one TE link's different attributes. For example, we can use a LSA to describe a TE link's local index, remote node id, remote index, bandwidth .etc, and use another LSA to describe the TE link's impairment information.

5.2.1.2. Top TLV

We can use the current defined top-level TLV and add some new sub-TLV to carry impairment information, and new top TLV can also be defined to carry impairment information. In this document we adopt the latter since the former may cause non-impairment information flooding when the impairment information updated. A new top TLV is defined as follows (Figure 5.6):

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type = 3            |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Sub-TLVs                           |
.                                                               .
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-6 Impairment TLV

5.2.1.3. Sub-TLVs

Impairment parameters may depend on a link, a wavelength or a waveband, therefore we defines 3 Sub-TLVs for Link ID Sub-TLV, Wavelength Sub-TLV, and Waveband Sub-TLV, and 2 Impairment Parameter Sub-TLVs. These sub-TLVs are defined as follows: (as shown in Figures 5.7 to Figure 5.11).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = 30            |          Length = 4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Link ID                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-7 Link ID Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = 31            |          Length = 4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Wavelength                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-8 Wavelength Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = 32            |          Length = 4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Reserved    |   Waveband    |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-9 Waveband Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = 33            |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Impairment Parameter Lists...                |
.                                                               .
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-10 Impairment Parameter Sub-TL

Each element of Impairment Parameter Lists is:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|I| Application |F|     IT       |          Upper Limit          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Upper Limit (Cont.)          |          Lower Limit          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Lower Limit (Cont.)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5-11 Impairment Parameter Element

Therein F = 0 denotes that the impairment's value is a unique number, i.e. there is no Lower Limit field in this element. F = 1 denotes that the impairment's value is a range, i.e. there is Lower Limit field in this element.

5.2.1.4. Flooding mechanism

Each node creates the impairment LSA based on the local physical impairments information (e.g. from monitors) and flood the Impairment LSA to the whole network. If the impairment parameters change, update the related LSA.

5.2.1.5. Usage of new sub-TLVs

We can use the combination of the above defined TLV to create the Impairment LSA. There are ordering requirements for these sub-TLVs:

- If there is no Link ID sub-TLV in an Impairment LSA, it means this LSA describes the node (node id is the advertising router indicated in the LSA header) related impairment information.
- If there is a Link ID sub-TLV in an Impairment LSA, the sub-TLVs following it are related to the link.
- If there is a Wavelength sub-TLV in an Impairment LSA, the Impairment Parameter Sub-TLV following it describes the wavelength related impairment information.
- If there is a Waveband sub-TLV in an Impairment LSA, the Impairment Parameter Sub-TLV following it describes the waveband related impairment information.

5.2.1.6. Examples

- If a node wants to flood the impairments related to the node, the Impairment LSA can be constructed as follows:
  - LSA Header | Impairment TLV header | Impairment Parameter Sub-TLV
- If a node wants to flood the impairments related to a link, the Impairment LSA can be constructed as follows:
  - LSA Header | Impairment TLV header | Link ID sub-TLV | Impairment Parameter Sub-TLV
- If a node wants to flood the impairments related to a link and a wavelength belonging to the link, the Impairment LSA can be constructed as follows:
  - LSA Header | Impairment TLV header | Link ID sub-TLV | Impairment Parameter Sub-TLV | Wavelength sub-TLV | Impairment Parameter Sub-TLV
- If there are impairments relatd to multiple wavelengths, repeat the combination of (Wavelength sub-TLV | Impairment Parameter Sub-TLV).
- If a node wants to flood the impairments related to a link and a waveband belonging to the link, the Impairment LSA can be constructed as follows:
  - LSA Header | Impairment TLV header | Link ID sub-TLV | Impairment Parameter Sub-TLV | Waveband sub-TLV | Impairment Parameter Sub-TLV
- If there are impairment parameters related to multiple wavebands, repeat the combination of (Waveband sub-TLV | Impairment Parameter Sub-TLV).

5.2.2. PCEP extension

In the Path Computation Request (PCReq) message, the PCC must be able to request a lightpath with some specific minimum signal parameters (e.g. OSNR, ...) depending on its QoS requirements. In order to accomplish this, two additions can be made to the PCReq format. First, the Request Parameters (RP) Object, which specifies the characteristics of the computation request, can contain a flag indicating a physical layer (PL) computation request. If the PCE is not capable of PL computations, a PCEP Error (PCErr) Message should be returned. Second, the physical layer parameter bounds for the path can be included in METRIC Objects (of a new type, or consider the them as TE metrics) containing the constraint values (e.g. minimum OSNR,. . .) for the requested path. Similar TLV formats as suggested for OSPF-TE can be used. The same format can be used in the PC Reply (PCRep) message to specify the computed parameters for the connection.

<table>
<tr><td colspan="4">OBJECT HEADER</td></tr>
<tr><td>Reserved</td><td>Flags</td><td>Type</td><td>Value</td></tr>
<tr><td colspan="4">optional TLV</td></tr>
</table>

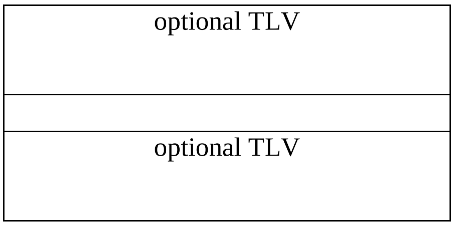

Figure 5-12 PCEP Notification Object

The PCEP provides Notification (PCNtf) messages, which can be PCE or PCC initiated, in order to notify the receiver of certain events in the network, such as PCE overload conditions or cancelled PCReqs. PPD changes could be sent using a new type of PCNtf message, being a PPD database update notification (PPDNtf). In order to send the information from the local database to the PCE, we added two new objecttypes (PCC IMPAIRMENT SEND and PCE IMPAIRMENT REQUEST) in the pce notification object for the PCE Notification message. The impairment parameters are added as optional TLVs to this object. Together with a path request, it would be advisable to send the latest monitor information from the local PPD. In order to avoid overloading the PCE with PPD update events during multiple consecutive PCReqs, a policy using a delay timer or some minimum required PPD change before sending notifications can be used. When the PCE receives a PCReq, it consults its (global) PPD for further impairment information needed to compute the requested path. If the connection has a high requirement for accurate information, entries in the global database may be too old. In this case, the PCE must query these nodes for their latest database information using a PCE IMPAIRMENT REQUEST. Furthermore, a policy to notify the PCE in case of monitoring alarms can be implemented using PCC-initiated notifications.

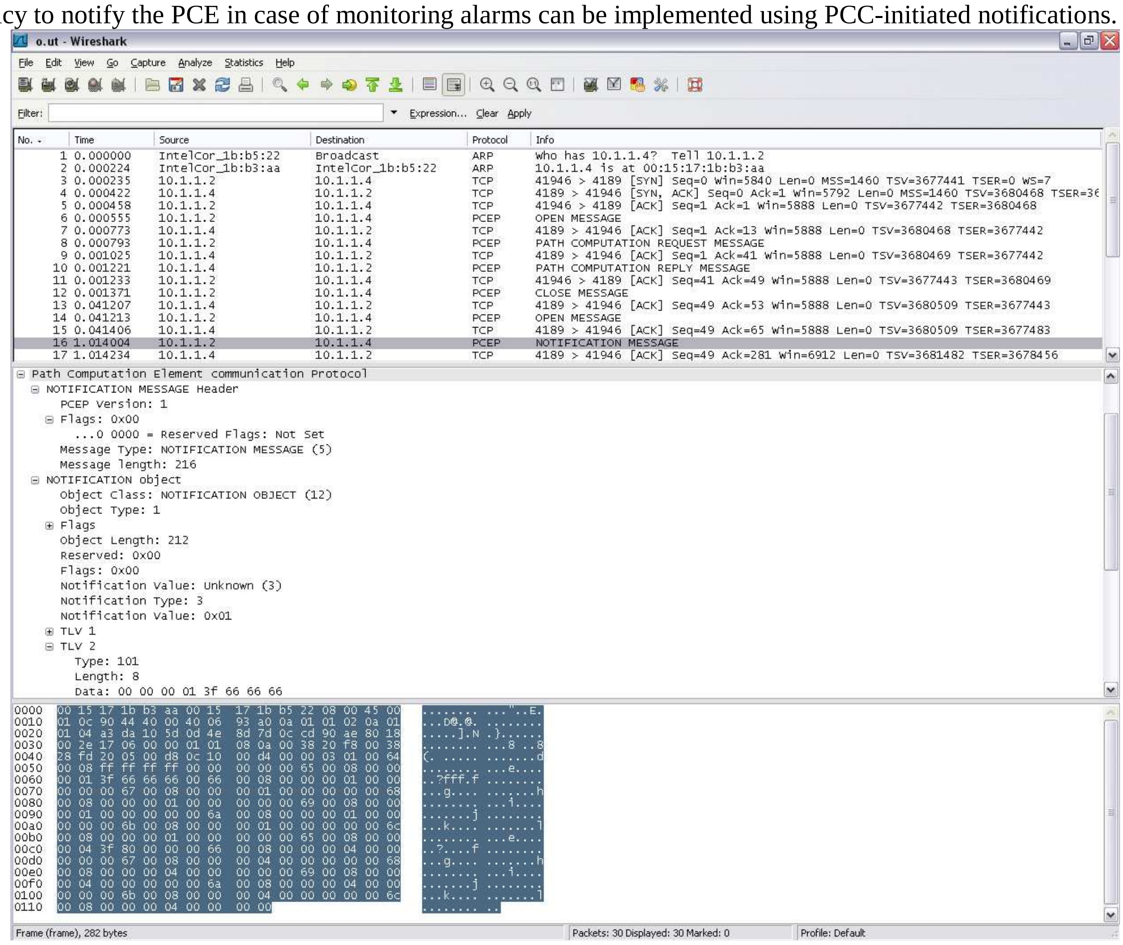

Figure 5-13 Wireshark trace of the PCC-PCE communication

In Figure 5.13 we show a wireshark trace of a test run of our PCE implementation, showing communication between the PCC (10.1.1.2) and the PCE server (10.1.1.4). Wireshark contains filters for PCEP as of version 1.0.3, but the IANA specified TCP port for PCEP (4189) is not yet recognized. There are 3 subwindows in the wireshark display. The top window shows the recorded packets, the center window shows the analysis of the selected packet and the bottom window shows the raw packet data. There are 2 sessions displayed. The first 2 packets are the Ethernet ARP request and reply, packets 3-5 are the establishment of the TCP connection. PCEP specifies that the connection between PCE and PCC remains established even when no PCEP sessions are active. Each PCEP session starts with an OPEN message (packets 6 and 14). This OPEN message specifies the session ID (incremented each session) and parameters such as the keepalive (defaults to 30 seconds) and deadtime (defaults to 4 times the keepalive) timing. In the First PCEP session, the client requests a path computation through a PATH COMPUTATION REQUEST message (packet 8), and the PATH COMPUTATION REPLY is sent (packet 10). The client then closes the connection using a CLOSE message (packet 12). The second session sends a NOTIFICATION message (Message Type 5) to the server with a PPD update. This message is shown in detail in the center subwindow. The message contains a NOTIFICATION object (ObjectClass 12) with Notification Type 3. This is our PCC IMPAIRMENT SEND object. This object carries the impairment TLVs, where TLV with type 101 (specifying the Q-factor) is shown. The wavelength ID is 1, the value is 0.9 in IEEE-754 floating point notation

5.3. Test cases and simulation objectives

In this deliverable, we will only evaluate the performance of the PCEP extension approach, while the OSPF-TE extension approach will be studied in T5.2.

In order to evaluate the proposed dissemination scheme based on the PCEP extension, we implemented the PCE server and PCC client in C++. Most of the functionality contained in RFC 5440 was implemented, although some messages and objects were simplified. Our test setup consists of 3 linux (Debian Lenny) PCs emulating one PCE server (10.1.1.4) and 2 clients (10.1.1.2 and 10.1.1.3) connected through an ethernet switch in star topology. The database system used is MySQL 5.0.
We ran an experiment to measure the database update process. The PCC at 10.1.1.2 requests new paths from the PCE at 10.1.1.4 with a poisson distributed arrival time normalized to 1 (In the experiment the arrival time is 1 second). The monitor update arrivals are also poisson distributed from 1 to 0.01 (1 second to 10 milliseconds). The client at 10.1.1.3 compares the client PPD database and the PCE PPD database at regular intervals of 1 second. Note that network delay barely influences the database polling, since it was well below 1 millisecond.

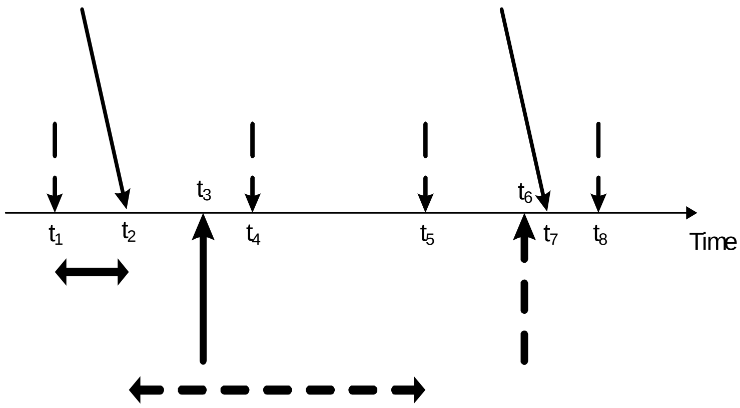

Figure 5-14: Timing Diagram

Figure5-14 shows the timing diagram of our study. The arrows above the horizontal time axis represent the monitor updates/registration in the local database (acquisition time, short dashed line) at times t_1,t_4,t_5 and t_8 and the update messages/registration in the global PPD (long continuous line) at times t_2 and t_7. The lines for the update messages are slanted to indicate the propagation delay of these messages in the network. Note that the global PPD stores its registration time of the value (t_2 or t_7), and not the monitor acquisition time of the value (which would be t_1 or t_5 respectively). In operation it would of course make more sense to store the acquisition times of the monitored values. If we compare both the local and global database at time t_3, the global database will be up-to-date and the value registration time – acquisition time = $t_2 - t_1$ will be positive (continuous horizontal line). If we compare both databases at time t_6 the database will be outdated, and registration time – acquisition time = $t_2 - t_5$ will be negative (dashed horizontal line).

5.4. Performance study

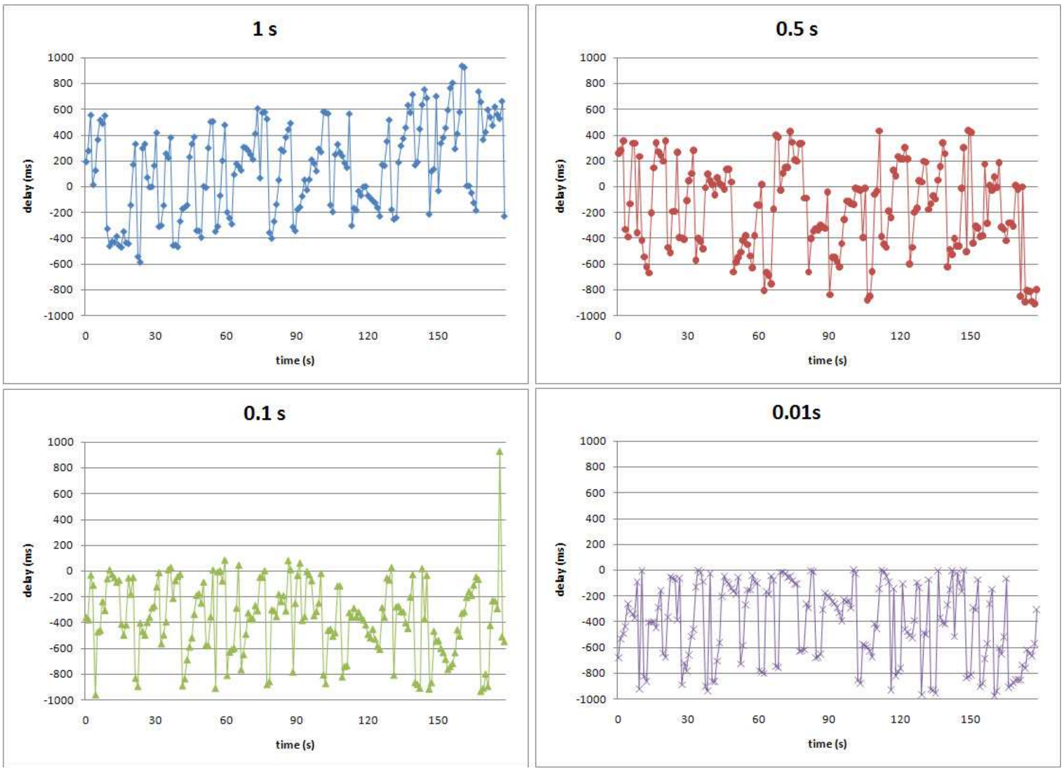

Figure 5-15 Database delay between server and client

Figure 5.15 shows the measured database delays between the PCC and the PCE. The times displayed in the labels are the monitor interval times. The vertical axis is the delay in ms, the horizontal axis is the time in seconds, where roughly 3 minutes of monitored activity is shown. The delay in the vertical axis is the difference between the timestamp in the server database and the timestamp in the client database. If this value is positive, the server database is up-to-date. Note that the request timing puts a limit on the maximum time discrepancy between both databases.

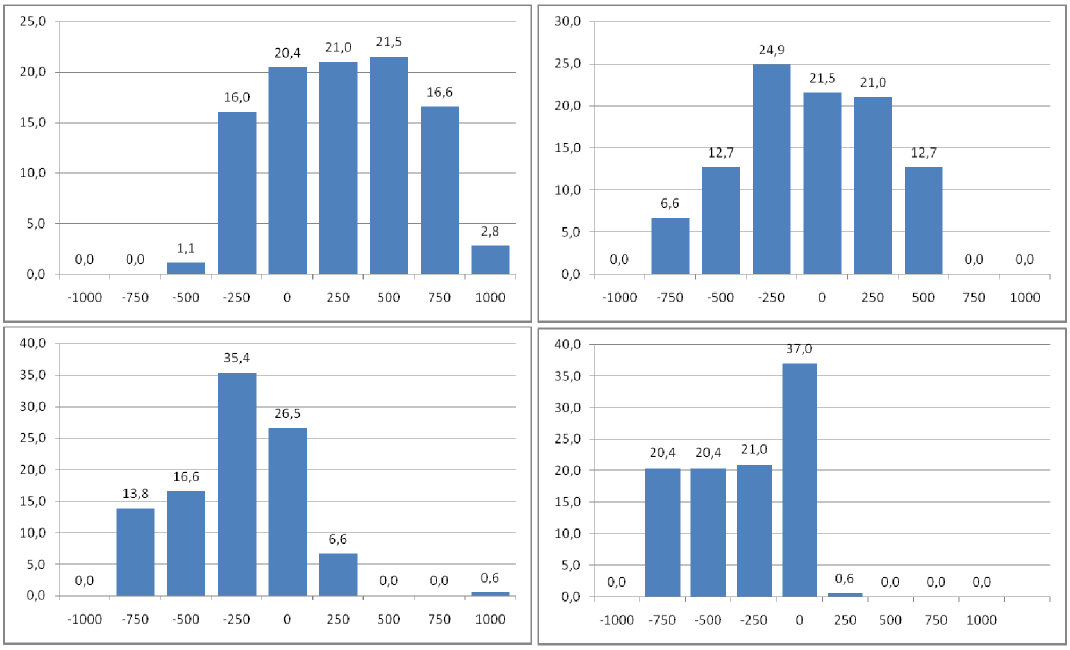

Figure 5-16: Statistical analysis of timing data

Figure 5-16 shows the probabilities (in percent) of the database delay being in specific intervals. When both the monitor updates and the path requests are distributed with 1 second intervals, the average delay was 137 ms, with 62% of monitored values positive. When the monitor update speed is increased to 0.5 seconds, the average delay drops to -180 ms and only 34% of values are positive. When the monitor delay drops to 0.01, the average delay equals -469 ms, with 99.4% of values negative (meaning the PCE database is practically always outdated). These figures all agree with what would be expected from analytical statistical analysis of two Poisson distributed processes. The faster the arrival of update messages with regard to the monitor updates, the more up-to-date the PCE database will be. In DICONET, highly agile networks are considered with very frequent lightpath requests and teardowns. We have now only considered the age of the database entries. Whether the scheme of updating the PCE database solely when a path requests is sent will be sufficient will depend on the monitor update speed, the dynamicity of the path requests and the allowed age of the database entries in the PCE. The more dynamic the network, the more up-to-date the PCE database will need to be, since the parameters heavily depend on the lightpaths set up in the network. The presented solution scales with this dynamicity of the traffic, and will update the PCE PPD faster if the network is more dynamic.

6. Conclusions

Table 6.1 Comparison of various control plane architectures

Metrics		**Signaling based approach**	**Routing based approach**	**Hybrid approach**	**PCE based approach**
Performance	*Blocking probability*	High	Medium	Low	Low
	Average number of attempts	High	Medium	Low	Low
	LSP set-up time	High	Medium	Low	Low
	Path computational complexity	Low	High	Very high	Very high
Engineering	*Control Plane Load*	Low	Medium	High	Medium
	Robustness to TED inconsistency	High	Low	Medium	Low
	Protocols to modify	RSVP-TE and implementation of new protocol for handling possible active lightpath disruption	OSPF-TE	Both RSVP-TE & OSPF-TE	PCEP
	Distribution of PLI information	Local	Globally flooded	Globally flooded	Communicating PLIs information to PCE server
	Impact on control overhead	Low	High	Very high	Medium
	Impact on existing protocol modules	High	Low	Very high	Medium
	Standardization Efforts	Initial efforts started in IETF in recent years	Some existing efforts inside IETF	No	Initial stages

Table 6.1 provides a qualitative comparison of various approaches w.r.t. to both engineering and performance metrics. The *performance metrics* refer to the metrics useful to verify how good a certain solution performs compared to others, while the *engineering metrics* refers to the metrics that gives a measure of the architectural impact of the solution on the overall network.

Based on the qualitative comparison presented in Table 1 and considering the resources allocated to exploring all possible control plane architectures with the resources, the DICONET consortium decided (during Trento meeting in February 2009) to explore three possible control plane architectures further: 1) signaling based, 2) hybrid, and 3) PCE based. The stand alone routing based approach is not studied in detail in DICONET due to various reasons pointed out in Section. 2. The signaling based architecture will be implemented and evaluated separately with modification to RSVP-TE protocol to carry both linear and non-linear impairments required by Q-tool to evaluate Q-factor of a) the new lightpath request and b) the existing active lightpaths that could be affected by the establishment of new lightpath. New mechanism (as explained in Section. 3) to detect the possible disruption of existing active lightpaths are also be implemented. The hybrid architecture will be implemented by considering the changes made to RSVP-TE protocol in signaling based architecture together with the introduction of few minimal changes to OSPF-TE (e.g. wavelength availability information, etc.) that will help to improve the overall performance of the network without affecting too much the complexity of the overall system. The PCE based architecture is also considered as the emerging architecture which requires changes to OSPF-TE and PCEP to compare and contrast the advantages of centralized control plane architecture with the other two distributed architectures. Finally, note that this report (D2.3) will be updated in July/Aug based on extensive studies and feedback from other work packages to provide a quantitative conclusions too.

7. References

[1] GLASS simulator, http://www-x.antd.nist.gov/glass/Main.htm

[2] ANCLES simulator: http://www.telematica.polito.it/ancles

[3] DICONET Milestone Report, M2.2, December 2008.

[4] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

[5] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource reSerVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[6] A. Farrel, D. Papadimitriou, J.-P. Vasseur, A. Ayyangar, "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using Resource reSerVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4420, February 2006.

[7] A. Farrel, D. Papadimitriou, J.-P. Vasseur, A. Ayyangar, "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using Resource reSerVation Protocol-Traffic Engineering (RSVP-TE)", draft-ietf-ccamp-rfc4420bis-03.txt, May 2008.

[8] Chava Vijaya Saradhi et al., "Control Plane Issues in Cross-layer Optimized Dynamic Optical Networks" ICTON 2009, invited paper.

[9] G. Martinelli, *et al.*, "GMPLS Signaling Extensions for Optical Impairment Aware Lightpath Setup," *IETF draft*, *draft-martinelli-ccamp-optical-imp-signaling-01.txt*, February 2008.

[10] E. Salvadori *et al.*, "Towards Deployment of Signalling based Approaches for Impairment Aware Lightpath Setup in Transparent WDM Optical Networks," Proc. of ECOC 2008.

[11] R. Martínez *et al.*, "Challenges and Requirements for Introducing Impairment-Awareness into the Management and Control Planes of ASON/GMPLS WDM Networks," *IEEE Communications Magazine*, vol. 44, no. 12, pp.76-85, December 2006.

[12] E. Salvadori *et al.*, "Distributed Optical Control Plane Architectures for Handling Transmission Impairments in Transparent Optical Networks," *to appear in IEEE/OSA Journal of Lightwave Technology*.

[13] Ch. V. Saradhi and S. Subramaniam, "Physical Layer Impairment Aware Routing (PLIAR) in WDM Optcial Networks: Issues and Challenges," to appear in *IEEE Surveys and Tutorials*.

[14] A. Farrel and I. Bryskin, *GMPLS: Architecture and Applications*, Morgan Kauffman, 2005.

[15] P. Castoldi *et al.*, "Centralized vs Distributed Approaches for Encompassing Physical Impairments in Transparent Optical Networks," *Proc. of ONDM*, pp. 68-77, May 2007.

[16] F. Cugini *et al.*, "Implementing a Path Computation Element (PCE) to Encompass Physical Impairments in Transparent Networks," *Proc. of OFC* 2007.

[17] F. Cugini *et al.*, "A Novel Signaling Approach to Encompass Physical Impairments in GMPLS Networks," *Proc. of IEEE Globecom*, pp369-373, 2004.

[18] D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC4328, January 2006.

[19] R. Braden *et al.*, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification," RFC 2205, September 1997.

[20] R. Braden and L. Zhang, "Resource ReSerVation Protocol (RSVP) -- Version 1 Message Processing Rules," RFC 2209, September 1997.

[21] J. Strand and A. Chiu, "Impairments and Other Constraints on Optical Layer Routing," *IETF RFC 4054*, May 2005.

[22] DICONET Milestone Report, M2.3.

[23] DICONET Deliverable Report D3.3.

[24] DICONET Deliverable Report D2.1.

[25] K. Kompella and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", *RFC 4203*, Oct 2005.

[26] D. Karz, K. Kompella, D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2," *RFC 3630*, Sep 2003.

[27] G. . Bernstein, Y. Lee, D. Li, and W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", Mar 2009.

[28] DICONET Deliverable Report D5.1.