Multilayer Traffic Engineering Experimental Demonstrator in the Nobel-II Project

J. Jimenez[1], O. Gonzalez[1], B. Puype[2], T. Cinkler[3], P. Hegyi[3] , R. Muñoz[4], R. Martínez[4], F. Galán[4], R. Morro[5]
[1] Telefónica I+D, [2] IBBT, [3] BME, [4] CTTC, [5] Telecom Italia
{fjjc, ogondio}@tid.es, bart.puype@intec.ugent.be, {cinkler, hegyi}@tmit.bme.hu, {raul.munoz , ricardo.martinez , fermin.galan}@cttc.es, roberto.morro@telecomitalia.it

Abstract

The present paper aims at presenting the specification, implementation and preliminary testing of a basic Multi-Layer Traffic Engineering (MLTE) experiment within the framework of European Project IST NOBEL-II. The main goal of this experimental activity is the validation of a very simple MLTE concept scenario. The main output would be the TE Manager software and additional Interfacing Tools as key components for future MLTE experimentation. The experimentation is proposed by interconnecting multiple research networks of different technologies following an ASON/GMPLS architecture, which allows the dynamic establishment of on-demand optical circuits based on the UNI interface.

1 Introduction

There is a common trend within operators towards a common IP/MPLS network that supports multiple services, which is also transported over an optical infrastructure. The evolution in the optical switching technologies, together with the development of a generalized control plane for transport networks (Generalized Multiprotocol Label Switching (GMPLS) [1]), allows for unprecedented flexibility of the transport stratum. In this context, the IP/MPLS network is enhanced in terms of flexibility and scalability by the GMPLS transport network, allowing IP to use the dynamic optical services.

In order to operate such a network with IP/MPLS and GMPLS, Multilayer traffic engineering utilizes optical layer switching technology as an asset to increase upper layer network performance in terms of throughput, traffic loss, etc. In addition to the routing of traffic flows, which is the traditional goal of traffic engineering, the more complex aspect of MLTE is the design of a logical topology for upper network layers using the underlying optical layer as bandwidth provider. For an IP-over-Optical network for example, IP routers are connected using optical connections (lightpaths), thus forming an IP layer logical topology that can be completely different from the underlying physical topology (optical fibbers and switches) [2].

In this context, a proof-of-concept test scenario is proposed hereafter, which experimentally demonstrates some basic functionality that is required to perform MLTE. Therefore, the goal at this stage is the experimentation and concept validation of a basic TE scenario. In the future, more complex multilayer TE mechanisms could be assessed from the experimentation start point described in this paper.

The present paper is structured as follows: First of all, the technical approach to the demonstration is explained. Next, the basic Multi-Layer Traffic Engineering validation scenarios are described. Then, two network configuration options for the actual implementation are detailed. Finally, the system architecture for the Multi-Layer TE Manager is explained.

2 Technical approach

For the MLTE demonstration, a few technological decisions have been taken. The scenario to perform the experimentation is focused on a packet switched (IP/MPLS) over circuit switched (ASON/GMPLS) multi-layer infrastructure on a single domain network. A simple network of three nodes in each layer is the minimum requirement for a demonstration, and therefore three IP routers and three optical cross connects are needed.

The demonstration uses both data plane and control plane (CP). Control plane interconnection is based on NOBEL2 data communication network (DCN), which is described in section 2.1. It is possible to perform the validation process only with the TE mechanisms provided by the control plane (or even with the basic CP features), where the data plane part of UNI is eventually emulated, but the availability of a Data Plane provides a better framework to demonstrate the multilayer TE concepts. The general scenario proposed to perform the multilayer TE tests is shown in Figure 1.

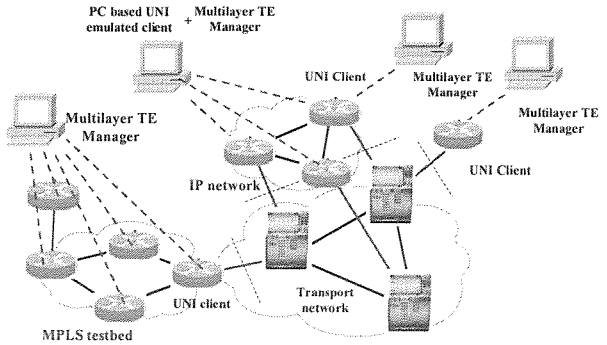

Fig. 1. Basic multilayer transport network configuration for TE mechanism validation.

Due to the current lack of commercial multilayer networks with a unified control plane, the demonstration will be focused on an scenario following an overlay approach. In the overlay model, no routing protocol information (e.g., topology and network resources) is exchanged between the client and the optical network. Thus, only signaling control messages for the connection provisioning are exchanged through the UNI interface. This model is used by the ITU-T ASON architecture [3] as well as the Optical Internetworking Forum (OIF) UNI standard [4] In this context, client IP networks could automatically request optical connections through the UNI interface as a result of Multi-Layer TE processes, and depending on the

actual needs, the appropriate bandwidth and quality of service could be assigned to them.

This scenario, even though more restricted than a true multilayer integrated network, better reflects the current situation in telecom operators in which IP and transport networks are separated and even operated by different departments, and usually with minimal interaction between them. However, the intention is not to perform an integrated network multilayer optimization, but how to exploit dynamic optical network for traffic engineering in the IP/MPLS layer. To achieve such optimization, a peer or augmented model should be used.

The basic scenario comprises one or multiple IP domains, but a single optical/SDH domain, but it could be extended in the future to multi-domain networks, interconnected through E-NNI interfaces.

Finally, an external Multilayer TE manager entity is needed to run the multilayer algorithms, trigger connection establishments and perform TE processes. In a real implementation this software could be included in network nodes, network managers or auxiliary entities. Regarding the MLTE Manager, a flexible and upgradeable centralized system with independent components has been designed, to cope with different network implementations.

2.1 NOBEL2 Data Communication Network

NOBEL2 DCN is composed by 7 domains/test-beds (i.e., Lucent, DT, TID, TI, UPC, CNAF and CTTC), which are interconnected following a star topology throughout a centralized GNU/Linux-based star-hub router provided by CTTC. The interconnection between the star-hub and each one of the 6 other remote peers is achieved using IPSec tunnels on top of the public Internet (CTTC is connected locally). Figure 2 shows the detailed architecture, including test-bed DCN's network addresses (10.X.0.0/16), IPSec tunnel peers and star-hub public addresses (obfuscated), and the IP address used for connectivity monitoring for each DCN.

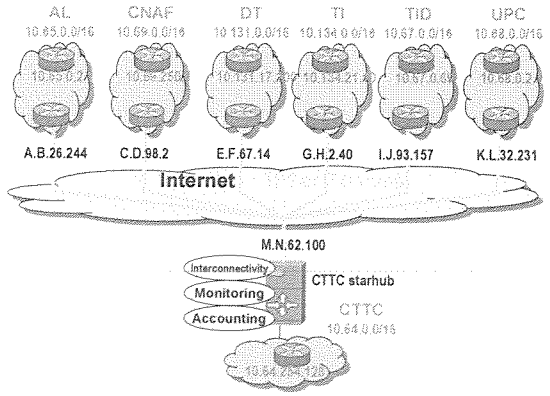

Figure 2. NOBEL2 Physical Control topology

The star-hub is responsible for two main tasks:

- provide transparent connectivity among all DCN domains (configuring the full-mesh virtual topology, as showed in Figure 3)
- to afford an ASON-GMPLS proxy mechanism that allows the "understanding" between domains using different interfaces with the purpose of setting up, maintaining and tearing down connections routed through both domains.

In addition, monitoring (in order to ensure that DCN works properly during experiments) and accounting (to know the utilization level of the star-hub resources) functionalities are also implemented. Both monitoring and accounting systems are accessed through a common intuitive easy-to-use interface based on web (https://nobel-starhub.cttc.es). For more information, please refer to [5].

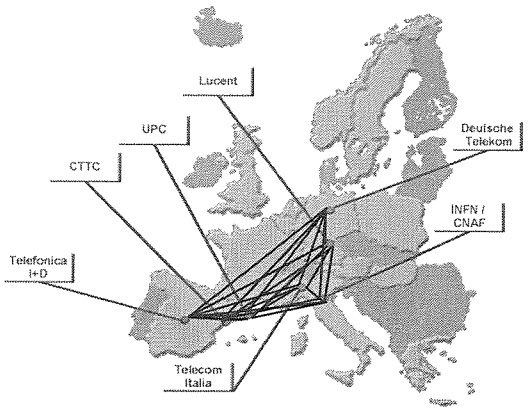

Fig. 3 Full-mesh virtual topology

3 Multi-Layer TE Scenarios

In order to demonstrate the feasibility of the Multilayer Traffic Engineering concept, several scenarios with different complexity, taking advantage of the test infrastructure available in the NOBEL project, are proposed.

3.1 Logical topology reconfiguration scenario

First, a basic scenario, depicted in Figure 4, demonstrates a logical topology reconfiguration (replacing one IP/MPLS Labelled Switched Path –LSP– with another LSP between different IP/MPLS routers), in order to react to an increase of required bandwidth between certain IP/MPLS routers. With the utilization of multilayer TE managing features, the manager can take the decision to build e.g., by-pass optical connections for those IP traffics that traverse the whole network (thus congesting the in-between routers).

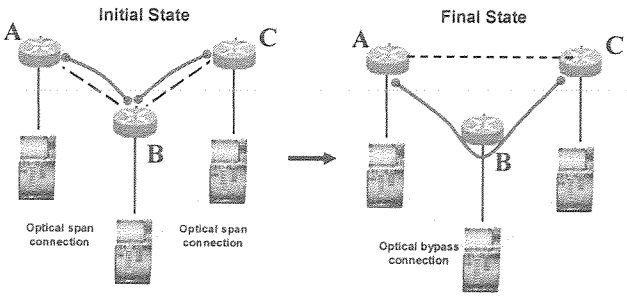

Fig. 4.Optically bypassing node B.

Initially, two optical connections are established, one between A-B and another one between B-C. Traffic from A to C is carried in two hops and it is routed in the IP layer at B. Later, the A-C IP traffic is increased, and when a

threshold is passed, an automatic reconfiguration is triggered, to enforce an optical bypass at B. At this point, a new optical connection between A and C is triggered, and a direct IP link (wavelength-path) from A to C is created.

Two options are possible, one is to tear down A-B and B-C connections before setting up the new one, or a "Make-Before-Break" reconfiguration, that is establishing the new connection before tearing down the initial ones. In the second case, two end to end connections are simultaneously needed for a short period, implying a higher number of resources for the demonstration.

The basic scenario is easily generalized for multiple connections. This way, it would be possible to adjust the topology (e.g. based on A-C traffic volume) on a dynamic basis. For true multilayer routing optimization, information about actual connections at the circuit layer would be needed, but a simplified case can also be demonstrated with IP traffic monitoring, as shown in Figure 5.

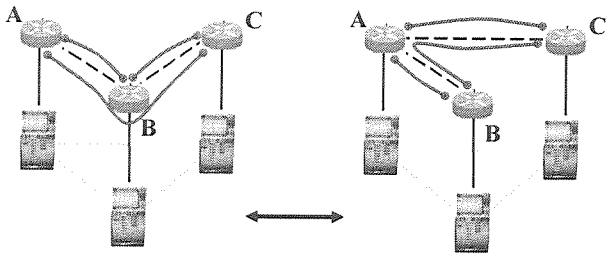

Fig. 5. Automatic reconfiguration when traffic between A and C grows, while the traffic from B to C drops.

3.2 Wavelength Path Fragmentation and De-Fragmentation scenario

The other demonstrated scenario is based on Adaptive Wavelength Path Fragmentation and De-Fragmentation [6], proposed by BME. The idea behind this approach is that if we have no more wavelengths terminated in a node (e.g., node B' in Figure 6) then the new demands from or to that node (B' in Figure 6) will be blocked. To avoid this drawback fragmenting existing wavelength paths is allowed. Fragmentation means, cutting a longer wavelength path into shorter segments. This will use more O/E and E/O ports, however, the virtual topology that the upper IP layer will see will be denser (better connected). This does not only make the topology denser (which results in shorter paths) but it also allows much better flexibility to changing traffic conditions.

De-fragmentation is just the opposite of fragmentation (Figure 6). It assumes that if a wavelength path is terminated in anode where it has only transit traffic using the same outgoing wavelength, then these two wavelength paths can be concatenated into a single one.

The best is to change the virtual topology when the considered wavelength paths do not carry any traffic. However, if the demands have much smaller bandwidths than the channel, many of them can fit into a single channel and therefore there will be hardly any moment when the reconfiguration can be performed. Therefore, we assume operations on "live" wavelength channels, i.e., traffic streams will be interrupted, which is tolerated by some services that allow resending the lost data units (packets) upon the interrupts. For some critical services this is not allowed, these services are to be carried over uninterruptible connections. Some traffic flows allow these interrupts and they resend the lost packets (e.g., TCP).

In case of fragmentation, the interrupts last while the ends of the cut wavelength path find the data unit delineation (frame/packet headers) again, typically within a time from μs to ms. However, if the fragmentation was neither a local one, nor a make-before-break one, then the wavelength path has to be torn down, and two new shorter wavelength paths are to be set up that lasts for seconds.

In case of de-fragmentation all the buffered packets will be lost, the end node has to determine loss of delineation and find new data unit delineation, while the two wavelength paths to be concatenated are to be torn down and a new wavelength path set up instead.

For both cases we investigate the hardware and control plane support as well as the local and make-before-break approaches.

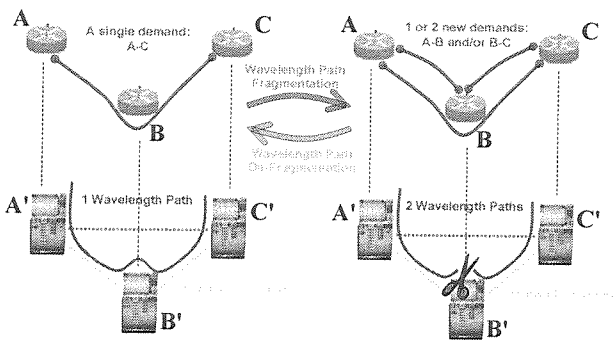

Fig. 6. Wavelength Path Fragmentation and De-Fragmentation..

4 Network Configuration

Two parallel approaches are being implemented in the context of IST NOBEL-II project for MLTE demonstration.

4.1 Integrated Multilayer Scenario

The first network implementation is focused on a IP/MPLS over an ASON/GMPLS network. For simplicity reasons, the MPLS capabilities of the upper layer will be disabled, and thereafter considered as IP, as it has been agreed that the presence of MPLS adds complexity to the basic demonstration. However, MPLS features are still considered as an important issue for further steps. On the other hand, the ASON/GMPLS network can be either TDM or lambda capable. In addition, the first demonstration targets a single domain scenario, but the intention is to make multi-domain scenarios also feasible in the future.

In principle, Telecom Italia optical/SDH test-bed is used as transport domain providing on-demand optical connectivity to the upper layer. The IP layer is composed by three different domains: the Telecom Italia domain, the Telefónica I+D domain and the IBBT domain. The IP domains are interconnected to the optical network through UNI interfaces.

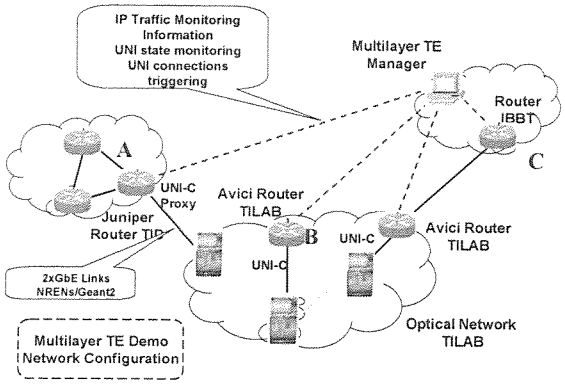

Fig. 7. The MLTE Demonstration Network.

As previously described, it has been agreed to follow an ASON approach, without an integrated control plane, for multilayer testing. This requirement is due to the fact that most currently deployed test-beds and production networks are composed by commercial nodes, which are quite complex to integrate with external testing or managing tools. Thus, any interaction from optical/SDH network with client networks will only be possible through the UNI, mainly the OIF public UNI. The main conclusion is that the optical/SDH network is opaque by definition, and very limited information is available for TE purposes.

It should be further investigated whether to extend this multilayer TE approach for other open interfaces, namely GMPLS private UNI, or alternative mechanisms to retrieve some information from the optical network topology and connection status. These mechanisms could include interfacing the network managers (commercial or developed ones), interfacing the MIBs in the nodes, or in case of using emulators, collecting information from the database. On the other hand, the IP network is expected to be accessible by multiple monitoring tools.

Telecom Italia owns two Avici routers with UNI-C capabilities developed for OIF tests in the last years. In addition, they are able to perform on-demand provisioning in the optical layer. IBBT domain will be directly connected to one of these routers through an IP link.

On the other side, Telefónica I+D will interconnect a Juniper router to the optical domain, and will make use of a UNI-Proxy solution, developed by PSNC (Polish Supercomputing National Center) within the context of IST MUPBED project and in an inter-project collaboration, for dynamically requesting for optical circuits (see Figure 8).

The interoperability between the Proxy and the Avici routers within the Telecom Italia test-bed has been already demonstrated [7].

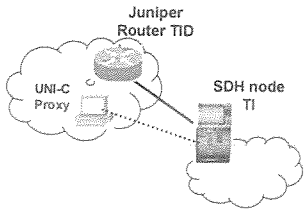

Fig. 8. The UNI-Proxy solution for dynamic SDH/optical circuits.

4.2 Local emulated Scenario

The second scenario has been proposed and implemented by IBBT. The proposed scenario is composed by three Alcatel service routers (Alcatel 7450/7750) connected by a 1 Gbit/s fibber ring. There are no optical cross-connects in this scenario, but the optical layer is emulated by enabling/disabling some of the optical ports of the service routers. The reason behind this alternative network scenario is to prove that the proposed approach is useful in different networks, and to provide an alternative test-bed for development of the manager software and evaluation of further TE techniques.

The service routers run OSPF natively to support TE driven rerouting of traffic; MPLS is also supported. Optical port configuration is done through the CLI interface of the routers. Linux PCs attached to each service router are used to interface the MLTE manager system to the service routers (e.g. implement a UNI-C proxy).

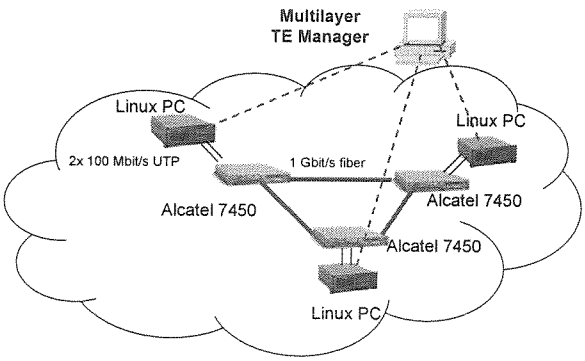

Fig. 9. IBBT Local Demo Scenario

4.3 Multilayer TE Manager

In both cases, a multilayer TE managing system is installed. The multilayer TE manager is conceived as software implemented and running in an IP router or a separate application running remotely on an external PC.

In the second case, a centralized model is a feasible option, so that the centralized multilayer TE manager is connected to each of the IP routers in the test-bed (or at least those with UNI features to be triggered and traffic to be monitored), through direct physical interfaces, through IP tunnels, through an IP network or even through the Internet.

This way, the manager has information available at the IP layer on all the significant points of the IP network, and would be able to interact and monitor the optical network through the UNI interfaces.

On the other hand, given the technical limitations of the centralized TE manager or due to implementation decisions, it is possible to implement multiple multilayer TE managers (e.g., one per test-bed) so that an information interchange infrastructure should be built to allow them to share information. For the first implementation step, a centralized Multilayer TE manager is implemented, although the distributed scenario is recommended in case a multi-domain framework, which is considered in future extensions.

4.4 Network and software configuration

The optical/SDH network configuration comprises setting up both data plane for SDH/Optical equipment, configuring topology and routing information relative to the control plane, checking availability of UNI-N capable interfaces availability, and installing additional control plane monitoring tools such as Clearpond [8].

The configuration of IP networks requires information about the equipment as well as the control plane interconnection (IP tunnels and configuring the addressing schema), as well as the availability of UNI-C interfaces. In addition, TE information comprising topology data (Virtual Topology) and needs to be configured, as well as IP forwarding tables and routing protocols such as OSPF.

The configuration of the TE manager includes the hardware needs for the TE manager (a PC), as well as interfaces for interconnection to routers and UNI triggering tools, monitoring tools, etc.

For all the interconnection purposes, the NOBEL 2 DCN, explained in section 2.1 has been configured used.

4.5 Scenario Procedure

At the beginning of the test, an initial virtual configuration of the IP layer is built up. Traffic sent on the IP layer is continuously monitored by means of CLI commands or SNMP requests to the routers MIB.

Based on the monitoring information, and upon a certain threshold, the MLTE Manager decides to change the virtual topology, and therefore sends a connection request (and possibly connection teardowns) to the optical network. The triggering interface and information retrieval are implemented by command line connections to the UNI clients.

After a new optical LSP is established, a new adjacency is created between the client routers, so that routing tables need to be updated. For that purpose, OSPF is continuously running at the IP layer, also taking into account that a special care must be taken with IP addressing. Preliminary tests show convergence times of 4 seconds.

After a new optical LSP is released, routing tables should be updated again, as routes should be dynamically removed (with a make-before-break approach).

Finally, basic optical network monitoring will be achieved by monitoring the UNI-triggering modules feedback (establishment/teardown), as well as the information provided by the Clearpond display software [8], which traces the OIF UNI messages and creates a drawing of the network.

5 Multilayer TE Manager System Architecture

The multilayer TE manager is conceived as the hardware and software infrastructure able to perform traffic engineering processes and to take traffic engineering decisions on a multilayer network. It is designed to demonstrate the feasibility of utilizing multilayer TE tools on a real network environment. This section will deep on the requirements and basic architecture of the MLTE manager.

5.1 Requirements for the Multilayer TE manager

Basing on the test scenarios and from the technical limitations imposed by the network infrastructure, the MLTE manager needs:

- Topology (IP and optical/SDH). Information about the physical topology of the network (both intra and interlayer links) is needed. We assume the topology does not unexpectedly change over time, so it is from a static configuration file. Nevertheless, the topology must be updated when simulating physical links failures.
- Virtual topology. Due to the dynamic nature of the tests, it is necessary that the TE manager keeps updated information about the virtual topology constructed over the circuit layer. For this reason, there is a mechanism to monitor the establishment of circuit switched end-to-end connections between IP routers.
- Traffic performance at the IP layer. For scenarios based on the IP layer performance and resources usage, traffic monitoring tools needs to be used to check the current operation of the IP layer. The TE manager needs to have this information available for decision triggering.
- Circuit bandwidth/resources information updates. In case multilayer routing scenarios are to be implemented, the TE manager needs to be informed of the occupation of the circuit layer resources (or at least knowledge of the physical routes associated to every single switched connection).

To accomplish these requirements, the TE manager needs to be interconnected to external agents, concretely:

- Monitoring tools are needed to monitor both the IP layer (e.g. Ethereal, Iperf and other specific tools), and the circuit layer connectivity and, if needed, connection routes.
- Triggering systems are mandatory to allow on demand connection establishments at the circuit layer. In an overlay model, these triggering systems will usually make use of the UNI interface.
- In addition, the TE manager has access to a database where all the required information (both static and dynamic) is kept up to date. This information comprises both topology and connections.

Finally, hardware/software requirements to install and implement the TE manager are not be very stringent in terms of processing power or platform needs. The TE manager and possibly the database) should be feasible to be running on a dedicated PC. This PC should need to be connected to external entities, such as monitoring tools or triggering mechanisms. For this reason, separate DCN-wise network connections could be needed for TE management. Additional workstations might be needed to install monitoring tools, triggering mechanisms or databases, but some of these mechanisms could also be located in the main TE manager server or integrated in network equipment (routers or switches).

5.2 Multilayer TE manager design architecture

Figure 10 shows a proposal for the multilayer TE manager software architecture:

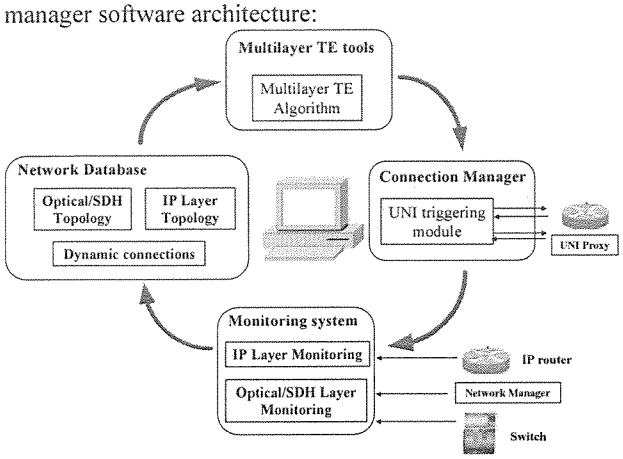

Fig. 10. Software architecture for the multilayer TE manager infrastructure.

The reference architecture has four principal modules:

- **Multilayer TE tool**: the main TE software applications, implementing the multilayer TE algorithms as well as any other TE specific software tool. This system will have the "intelligence" of the TE manager.
- **Connection Manager**. In charge of UNI triggering to the ASON/GMPLS transport network. With UNI triggering does not necessarily mean UNI-C implementation, but capability to trigger an external UNI-C process request (either to a router, a UNI Proxy or an emulator based implementation). The UNI triggering module must be able to collect the retrieved information from the network after a UNI request (including establishment failures or teardowns).
- **Monitoring system**. In charge of three different functions: IP traffic and connectivity monitoring, UNI connection establishment monitoring and, if possible, circuit layer monitoring. IP monitoring should be achieved by interfacing routers information tables, by signalling sniffing or any other monitoring mechanism. UNI establishment/teardown monitoring is achieved by interfacing the UNI triggering module. If the manager is the only agent making requests to the network, it would be easy to monitor the current status of the IP virtual topology just by noting down the information related to the connections it has already established.
- **Topology database**. Maintains both a physical (transport) and physical/logical (IP) topology and updates it dynamically by collecting feedback from the monitoring system and giving input to the multilayer TE tools.

The Multilayer TE manager is an open platform and the interfaces of the modules are defined to allow for future re-use. The idea is that at a first step, the modules have a basic functionality, enough to demonstrate the TE concept, and more complex functionality can be added later on. Therefore, it should be understood as the base for a platform to test MLTE algorithms and ML Recovery strategies developed in NOBEL2 [9].

6 Conclusions

The present article shows the design and implementation details of a Multi-Layer Traffic Engineering demonstrator developed in the IST NOBEL-II Project. The lack of an integrated control plane has proven to impose stringent requirements in the deployment of MLTE mechanisms. Anyhow the overlay approach has been demonstrated to be feasible and is adequate for operator's networks with separated IP and optical layers.

The experimental tests of the demonstration have shown that when MLTE reconfigurations are triggered, one of the key issues for a good performance is the convergence time of the OSPF routing in the IP layer.

The main achievement has been to provide a base platform to test MLTE algorithms and ML Recovery strategies. This platform can be enhanced in the future and used by upcoming European projects.

Acknowledgments

This work has been co-funded by the European Commission through the European Project IST NOBEL-II.

References

1. RFC 3945 "Generalized Multi-Protocol Label Switching GMPLS Architecture".
2. B. Puype et al. "Multi-layer Traffic Engineering in Data-centric Optical Networks, Illustration of concepts and benefits," in Proc. ONDM 2003, Budapest (2003), pp. 221-22
3. ITU G.8080 "Architecture for the automatically switched optical network ASON", G.8080/Y.1304.
4. OIF UNI2.0 "UNI2.0 Signalling Specification", http://www.oiforum.com/public/UNI_2.0_ja.html
5. R. Muñoz et al. "Experimental interconnection and interworking of the multi-domain (ASON-GMPLS) and multi-layer (TDM-LSC) NOBEL2 Test-beds", ECOC 2007
6. T. Cinkler et al. "Multi-Layer Traffic Engineering through Adaptive Lambda-Path Fragmentation and De-Fragmentation: The "Grooming-Graph" and the "Shadow-Capacities"", in Proc of Networking 2006
7. MUPBED Deliverable D4.4 "Demonstrations preformed in the MUPBED test bed year 2", August 2006.
8. Clearpond Monitoring Software, available at : http://www.clearpondtech.com/ipccmon.htm
9. IST NOBEL-II Project: www.ist-nobel.org

BroadBand Europe

3 - 6 December 2007
Antwerp, Belgium

Organized by

IBBT

The Interdisciplinary Institute for BroadBand Technology

Editors:
Peter Van Daele
IBBT-Ghent University (B)
Peter Vetter
Alcatel-Lucent (B)

ISBN 978 9076546094
www.bbeurope.org

Hosted and sponsored by

Alcatel·Lucent