Performance evaluation of Threshold-based Multi-layer Traffic Engineering strategies

Víctor López[1], Óscar González de Dios[2], José Alberto Hernández[1], Raúl Duque[2], Carlos García Argos[2], Javier Jiménez Chico[2], Juan Pedro Fernández-Palacios[2], Javier Aracil[1]

[1]Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid
Tel: +34914972268, Fax: +34914972235, E-mail: victor.lopez@uam.es

[2] Telefónica I+D, Emilio Vargas 6, 28043 Madrid

Current backbone networks are based on the IP paradigm, where the IP routers are interconnected through point to point high capacity links. Most of the IP traffic in core routers is pass-through, and may not be processed at the IP layer. Thanks to the advent of reconfigurable optical equipment, the traffic flows can be switched at the optical transport layer instead of consuming IP resources. This work proposes two "Threshold-based" algorithms that aim to reduce the traffic at the IP layer, while efficiently using the optical resources. These algorithms firstly route the traffic using the IP layer, detect the congested links and search for candidate by-passes, which are classified in terms of length and amount of shared traffic. The "Longest" algorithm off-loads the candidate by-passes based on their length. On the other hand, the "Largest" algorithm uses the shared amount of traffic to decide which are off-loaded first. The performance of both strategies is studied in terms of network congestion.

1. Introduction

Internet traffic has been constantly increasing for decades, and, currently, it is almost doubling every year [1]. This traffic is supported in current backbone networks by the IP paradigm, with an architecture where the IP routers are interconnected through point to point high capacity links. With the expected increment of the traffic, the processing capacity required in the IP routers increases linearly and with the number of hops as multiplication factor. Fortunately, most of the IP traffic is transit traffic, that is, when it arrives to a router, its destination is some other remote node. Consequently, such pass-through traffic may potentially not be processed at the IP layer.

The advent of reconfigurable optical equipment has brought an ultrahigh bandwidth dynamic transport layer that can be controlled by the GMPLS protocol stack. The traffic flows can be switched at the optical transport layer instead of consuming IP resources. Switching high capacity flows optically is cheaper and consumes less energy than switching at the IP layer [2]. This way, the optical layer may be used to reduce the CAPEX and OPEX requirements in the network, thus absorbing the traffic growth via the optical layer instead of the IP layer. On the other hand, optical networks have a very coarse per-connection granularity and they lack of the multiplexing benefits of IP networks, so the possibility of resources underutilization increases. In this light, Multilayer Traffic Engineering proposes solutions to deal with the joint utilization of IP and optical resources [3].

This work proposes two Multilayer Traffic Engineering “Threshold-based” algorithms that aim to reduce the congestion at the IP layer, while efficiently using the optical resources. These algorithms firstly route the traffic using the IP layer and detect the congested links. Once these links are detected, the candidate by-passes are searched and classified in terms of length and amount of shared traffic. The “Longest” algorithm off-loads the candidate by-passes based on their length, that is first the longest by-passes. On the other hand, the “Largest” algorithm uses the shared amount of traffic to decide which are off-loaded first. The off-loading process continues until there are no more candidate by-passes, yielding to the final network state. The performance of both strategies is studied in terms of network congestion and cost savings when the core network of an operator runs with the proposed multilayer approaches.

This paper is organized as follows: Section 2 describes the evolution of the transport and IP networks and the motivation for performing the joint operation of IP and transport layers. Next, Section 3 proposes some approaches to solve the MTE problem. Section 4 shows the impact of such mechanisms in the IP congestion and the by-passes established. Finally, Section 5 concludes this article and proposes future work.

2. Network evolution towards IP over GMPLS

Traditionally, different networks have been used for different services. However, current network trends foresee an evolution towards a unique multiservice IP backbone network capable for supporting any kind of service over the same infrastructure. Such evolution together with new traffic demands generated by intensive bandwidth consuming applications, such as P2P video streaming or IPTV, is strongly increasing the traffic volume over the IP backbone. In that respect, current IP backbones, often based on hierarchical architectures and high speed point to point links between routers, are starting to present scalability problems, mainly due to the increase of pass-through traffic in the IP transit routers. In order to solve this problem, operators are starting to deploy ``intelligent'' optical networks, based on ROADMs (Reconfigurable Optical Add Dropp Multiplexers) with GMPLS functionalities, able to dynamically provide high capacity and resilient connections between IP edge nodes.

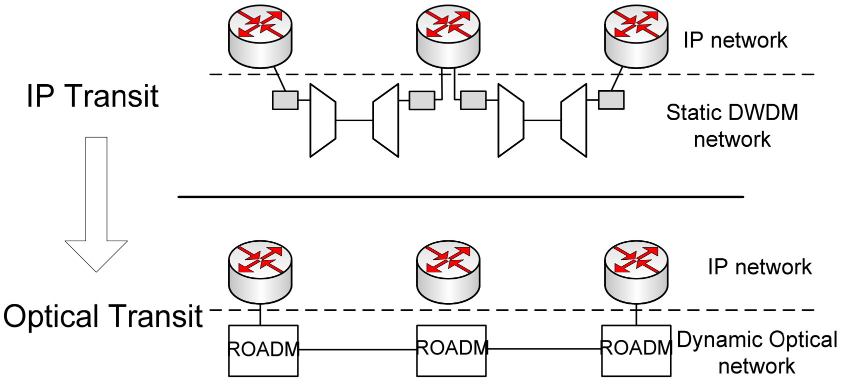


Figure 1: Transit traffic offloading over the transport network

In this new architecture, a wavelength switched network allows the IP routers to cope with the new traffic demands. The next step is to add flexibility by means of a control layer between the IP router and the optical switch, leading to multi-layer capable routers, where, traffic can be dynamically sent through either the IP layer or the optical layer (either into an existing light-path, or a newly created one). Note that this approach is evolutionary, as the IP routers are maintained, and new demands are absorbed by the optical layer. Thus, with the appearance of IP over Intelligent Optical Networks (ION), the number of technical alternatives for resilience and Traffic Engineering is increased. While in pure IP networks all the network engineering mechanisms are exclusively done at the IP layer, in IP over ION both resilience and traffic engineering could be executed in an independent or coordinated way. Next sections describe the main technical alternatives for such a multilayer networking approach.

2.1 Multi-layer Control Plane

The decision over implementing a unified IP/photonic control plane or separated control planes could have a strong impact on the multilayer networking performance:

- A single integrated control plane is the most appropriate solution if the intention is to implement multilayer optimization mechanisms. Nevertheless, this approach will require in most of the cases to a single-vendor deployment
- The use of two separated control planes without an adequate coordination could lead to instabilities and inefficiencies in the network configuration. This is especially problematic in case of network failures that could cause uncoordinated recovery actions.
- Finally, efficient multilayer traffic management is more difficult to achieve in the case of separated layers, but it could be feasible by means of the usage external tools (with algorithms as those presented in this paper) that could be located at the management plane.

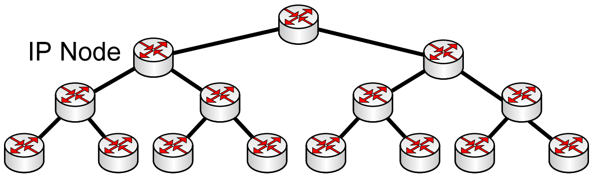
2.2 Multi-layer Data Plane

A further approach for performing multilayer networking is, not only to have an integration of cooperation between control planes of optical and IP layers, but to have an integration of optical and IP equipment. In this light, some IP vendors propose the integration of the DWDM transponders in the IP routers. This way, as well as having an integrated control plane, there is a reduction in network equipment, with a lower investment in optical transponders.

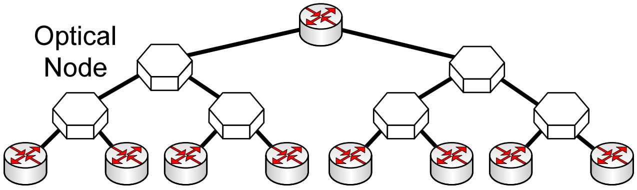
However, there are also potential disadvantages when using this kind of solutions. In WDM networks, the trend is that every provider uses proprietary solutions to optimize the distance of a light-path without regeneration and to the relationships between the transponders and the optical amplifiers. In this context, a multi-vendor scenario would lead to serious interoperability issues both between IP vendors and IP-WDM providers, even with the usage of standards mechanisms such as alien wavelengths. In conclusion, even though in some scenarios it would be feasible to choose for multi-layer data plane integration, there are several interoperability concerns which push for separated but coordinated layers.

3. Multi-layer Traffic Engineering Mechanisms

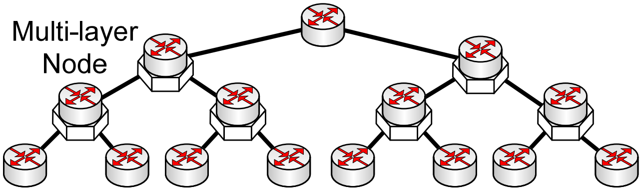
There are at least three different alternatives when considering an IP over WDM network. The current "All IP" architecture routes all traffic using the logical IP topology; the full offloading solution based on all optical transit nodes assumes that all IP edge nodes are connected with a full-mesh topology, that is, a light-path is provisioned between every two IP nodes. The third option comprises the Multi-layer Traffic Engineering (MTE) solution, which is a hybrid solution of the previous two solutions.



(a) Current All IP Nodes



(b) Optical Transit Nodes



(c) Multi-layer Nodes

Figure 2: Architectural IP offloading alternatives

The MTE solution proposed in this article transmits IP traffic using the available IP resources (just like IP over WDM does), but when the end-to-end traffic reaches a given traffic threshold an end-to-end light-path is established for such demand. Note that the full-mesh topology solution reserves a light-path for every aggregation pair, whereas the MTE solution only reserves one of such direct connections if the threshold capacity is reached. Therefore MTE solution comprises a more efficient use of the network resources.

3.1 Definition of MTE engineering mechanisms

Let us consider a network topology *T = [V; L]* consisting of a vertex set *V* and an arc set *L* [4]. It is assumed that each vertex represents a multi-layer capable router. Each arc *(x; y) € L* has associated a non-negative real number *l(x; y)*, which refers to its load. Let us define *D* as the demand matrix, where each value associated to the position *(i; j)* denotes the traffic demand from node *i* to node *j*, *d(i; j)*, *i; j € V*. The objective of this MTE problem is to by-pass all connections in the optical domain which exceed a given capacity threshold (C_{min}). This threshold can be chosen based on QoS parameters or economic cost, such that if exceeded, the cost of the optical offloading is lower than the cost of the equivalent IP resources.

Our MTE algorithm takes as a starting point a given logical topology at the IP layer, which may correspond to the current network situation. First, the algorithm maps the IP traffic over this logical topology provided by the optical layer using a given routing algorithm. Once the load in each link is known (*L*), the algorithm runs through the links of each node, finding *n* consecutive links where $l(i; j) > C_{min}$. Let us call this set of consecutive links as candidate paths set *CP* for optical by-passing. Note that the shared bandwidth among the links $l_1,\ldots, l_n$ is $BCP = min(l_1 - C_{min},\ldots, l_n - C_{min})$. Once the *CPs* are found, the algorithm has to decide the order to extract them. We propose two cases: “Single Threshold Largest By-Pass” and “Single Threshold Longest By-Pass”.

The “Single Threshold Largest By-Pass” algorithm sorts the candidate paths by its shared bandwidth (*BCP*) and, if there are paths with the same shared bandwidth, by length. On the other hand, “Single Threshold Longest By-Pass” algorithm firstly shorts them by length and, secondly, by shared bandwidth. Once the list is sorted, path candidates are extracted.

4. Impact of MTE on the IP over WDM architecture

This case study compares these MTE algorithms with the normal network operation, where IP resources are map over the topology given by the optical layer. Let us call this procedure: “All IP” solution. When the IP over WDM technology is used, no information exchange is done between the layers, so the IP layer just chooses a routing algorithm and sends its traffic based on this protocol. This case study routes the traffic using Shortest Path (SP) and Equal Cost Multi-Path (ECMP) algorithms. Essentially, ECMP is the same algorithm as SP, but instead of providing a single shortest path, ECMP uses the set of paths with minimum distance, thus enabling load balancing at the IP level.

The other extreme solution is to create a full mesh topology between all nodes in the network. This means that for each end-to-end demand, a lightpath is established between the source and the destination. Consequently, a network with N nodes would require $N(N - 1)/2$ ligthpaths. Obviously, this full mesh solution is feasible only when the end-to-end network load is too high, but it is out of the scope of this study.

This case study is done using the well-known backbone network topology: the NOBEL reference network (Figure 3) [5]. However, similar results are found in NSFNET backbone network [6]. The NOBEL reference network has 28 nodes, which are connected by 42 bidirectional links.

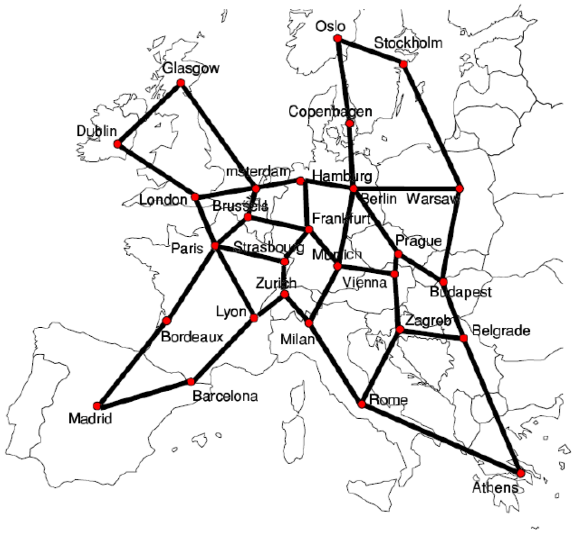


Figure 3: NOBEL reference network

In order to study network performance, a random traffic matrix is computed with a uniform distribution [0; 1]. Once the matrix distribution is computed, its values are scaled such that the amount of traffic between every end-to-end nodes is multiplied by an increasing factor in order to simulate different traffic loads.

4.1 IP layer congestion reduction

The objective of this algorithm is to detect those links that exceed a bandwidth threshold (C_{min}) in order to by-pass them using the optical layer. Figure 4 illustrates the amount of overloaded links (BW > C_{min}) in the “All IP” architecture and when the MTE algorithms are applied. Let us remark, that not only SP routing, but also ECMP is depicted. The X axis shows an increasing traffic rate until all links of the IP layer are congested. The NOBEL's network owns 42 bidirectional links and when there is a maximum rate between the end-to-end nodes of 1 Gbps, the IP layer is completely congested.

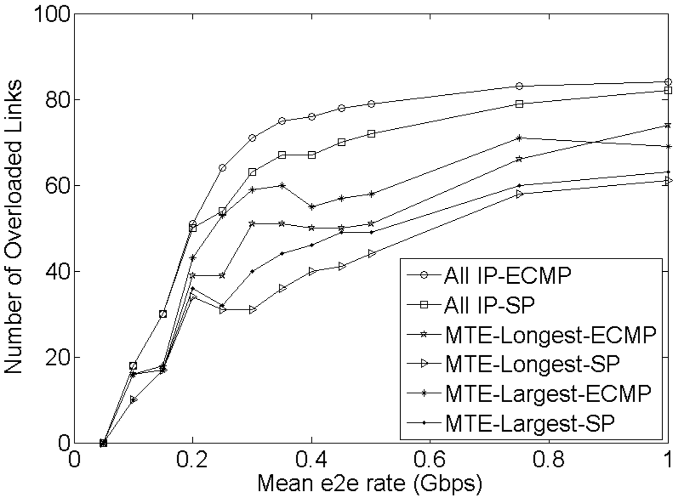


Figure 4: Number of Overloaded links (*BW* > C_{min}) in each topology

It is worth noting that the overload in the “All IP” topology differs when the routing algorithm changes. ECMP routing should split the traffic among the links and reduce the IP congestion, but this has the opposite effect for the MTE mechanisms. As ECMP balances the amount of traffic at the IP layer, it is more difficult to find a shared bandwidth greater than C_{min}. However, both routing mechanisms are affected similarly by the load increment. Once the performance of the “All IP” solution is outlined, the performance of MTE algorithms needs to be analyzed. The Largest algorithm reduces the number of overloaded links, but a less number than the Longest approach. The average reduction of both algorithms is around 30% for all simulations.

However, not only the congestion reduction, but also the impact at the optical layer is important. Figure 5 depicts the number of by-passes established by the MTE algorithms. The number of by-passes established with the Longest algorithm is much larger than the number that the Largest algorithm creates. This behavior is related with the algorithm itself. The Longest algorithm establishes the longest candidate bypasses at the beginning and then it continues with the shorter ones. On the other hand, the Largest algorithm creates the by-passes with the greatest amount of traffic at the beginning, thus, reducing the amount of candidate paths that exceed the minimum bandwidth (C_{min}). This reaches a more fragmented network configuration, yielding to a more overloaded IP topology.

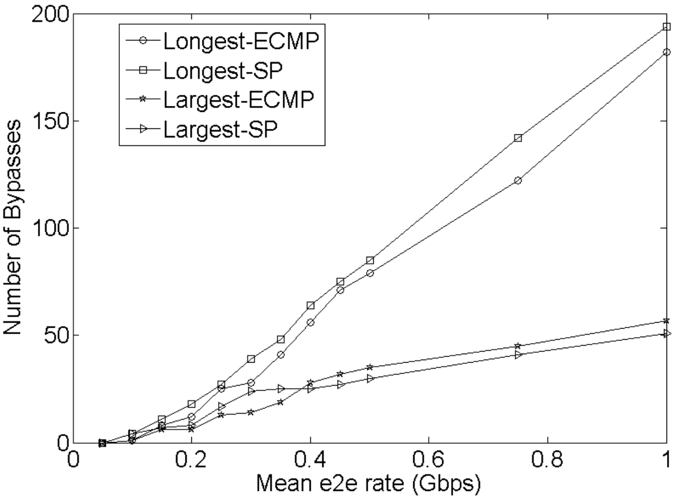


Figure 5: Number of established by-passes

Finally, let us focus on the amount of off-loaded traffic. Figure 6 shows the percentage of the traffic that is off-loaded to the optical layer. At low loads, there is no off-loading at all, since the amount of traffic is high enough to be groomed through the end-to-end ligthpaths. As the system becomes overloaded, this percentage of traffic is higher, consequently reducing the amount of traffic at the IP layer and absorbing the traffic increment just by the WDM layer. The better performance of the “Longest” algorithm is because it is able to send more congested paths to the optical network, yielding to a higher percentage of extracted traffic in comparison with the Largest algorithm.

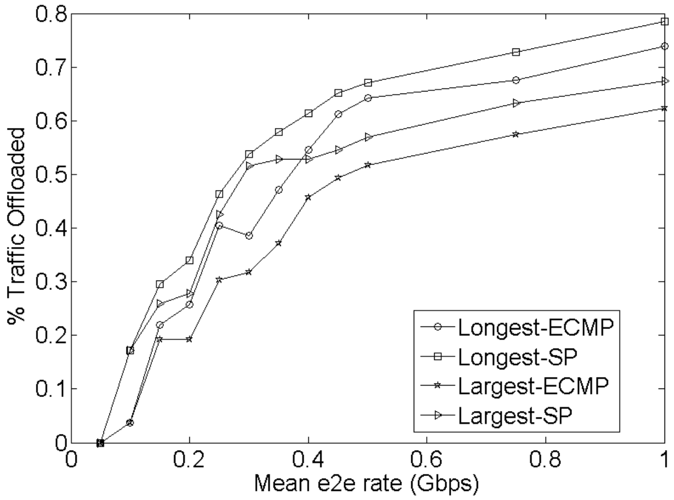


Figure 6: Percentage of off-loaded traffic over the total traffic

5. Conclusion

The main contribution of this article is the definition of two MTE algorithms to deal with the congestion of the IP layer in network operator. The novelty of such algorithm is that it, firstly, uses the IP resources to provide service to the incoming traffic, as current IP layer does. If there is congestion at the IP layer it searches the possible paths to be off-loaded using the optical layer.

In light of the results, we can see that the Longest algorithm achieves a better performance in terms of congestion reduction, but it uses more extensively the optical resources than Largest algorithm.

As future work, we will study the impact in the amount of resources at the IP and optical layer, as well as the definition of new mechanisms based on optimization.

Acknowledgements

This work was carried out with the support of the BONE project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme. The authors would also like to acknowledge the support of the Spanish project: Multilayer networks (TEC2008-02552-E).

References

[1] Cisco, white paper "Global IP Traffic Forecast and Methodology, 2006–2011"

[2] J. Baliga, R. Ayre, K. Hinton and R. S. Tucker, "Photonic Switching and the Energy Bottleneck", Photonics in Switching, August 2007.

[3] J. Enriquez Gabeiras et al. Is Multilayer Networking Feasible? Optical Switching and Networking 2009.

[4] L. R. Ford and D. R. Fulkerson, *Flows in Networks (Rand Corporation Research Studies Series)*. Princeton University, June 1962.

[5] "Next generation Optical network for Broadband European Leadership" (NOBEL), FP6-506760, http://www.ist-nobel.org/

[6] V. López, "End-to-end Quality of Service provisioning in multilayer and multidomain environments", *Ph.D. Thesis*, 2009.