

A Hierarchical Routing Approach for Optical Transport Networks

Eva Marín-Tordera, Xavier Masip-Bruin*, Sergio Sánchez-Lopez, Josep Solé-Pareta, Jordi Domingo-Pascual

Advanced Broadband Communications Center

Universitat Politècnica de Catalunya (UPC), Barcelona, Catalunya, Spain

{eva, xmasip, sergio, pareta, joridid}@ac.upc.edu

Contact Author*:

Xavier Masip-Bruin

Universitat Politècnica de Catalunya

Advanced Broadband Communications Center

Av. Victor Balaguer s/n

08800 Vilanova i la Geltrú-Barcelona

Catalunya (Spain)

Tel.: +34 93 896 77 67

Fax: +34 93 896 77 00

E-mail: xmasip@ac.upc.edu

Keywords: Hierarchical networks, Optical Transport Networks, Prediction-Based Routing, aggregation schemes

Abstract

Although the Automatically Switched Optical Networks (ASON) specifications strongly recommend a hierarchical network architecture for these networks, this is still an open issue. The hierarchical network concept involves several mechanisms mainly related with signalling and routing, such as the aggregation scheme, the dissemination process, the updating policy and the routing algorithms. The existing mechanisms for flat networks must be substantially modified to be applied to a hierarchical network architecture. In this paper authors propose a complete hierarchical routing approach mainly focusing on routing concerns, aiming to optimize the global network performance while guaranteeing scalability.

1. Introduction

Optical Transport Networks (OTN) appear as a solution to support the new network requirements owing to broader network expectations produced by both a significant increment of network users and the new emerging Internet applications. When such networks incorporate the automatic switching capability are named ASON, (Automatically Switched Optical Networks). A hierarchical network architecture comes out as one of the hard recommendations stated at the ASON specifications [1] to guarantee network scalability. Therefore, traditional flat network structures must be properly modified to fulfill that ASON recommendation. Main concepts to be modified are those related to signaling and routing, such as the network information aggregation, the network information dissemination, the updating policies and the routing algorithm. All these issues are covered by the ASON Control Plane.

It seems clear that the ASON Control Plane will be based on the Generalized Multi-Protocol Label Switching, GMPLS [2]. Several papers focus on the routing requirements for the GMPLS suite of protocols to support the capabilities and functionality of the ASON control plane [3], [4] and [5]. According to them, the following functionality is expected from GMPLS routing to instantiate the ASON routing realization:

- Support multiple hierarchical levels of Routing Areas (RAs); the number of hierarchical levels to be supported is routing protocol implementation specific.
- Support hierarchical routing information dissemination including summarized routing information.
- Support for multiple links between nodes (and between RAs) and for link and node diversity.
- Support architectural evolution in terms of the number of levels of hierarchies, aggregation and segmentation of RAs.
- Support routing information based on a common set of information elements as defined in [4] and [5], divided between attributes pertaining to links and abstract nodes (each representing either a sub-network or simply a node).

The main advantage of hierarchical routing is to reduce communication overhead while providing efficient routing. The routing algorithm dynamically computes paths supporting the service constraints required by the incoming connection. We adopt the distributed connection management protocol proposed in [6] which implements source routing. While using source routing, routes are computed on the source nodes according to the routing information contained in their network state databases (named Traffic Engineering Databases, TED, when including available resources information).

In this paper authors propose a complete hierarchical network structure mainly focusing on signaling and routing concerns. The remainder of this paper is organized as follows. Section 2 outlines main hierarchical network issues. In Section 3 a clear description of our hierarchical network approach is done. Then, in Section 4 the main contribution on hierarchical routing is proposed. An in-depth evaluation is presented in Section 5. Finally, section 6 concludes the paper.

2. A Hierarchical Network Structure Overview

A whole hierarchical network structure should be subdivided into routing areas (RAs), (see Fig.1 as an example) containing physical nodes with similar features. The RA nodes should exchange topology and resource information among themselves in order to maintain an identical view of the RA. This information should be

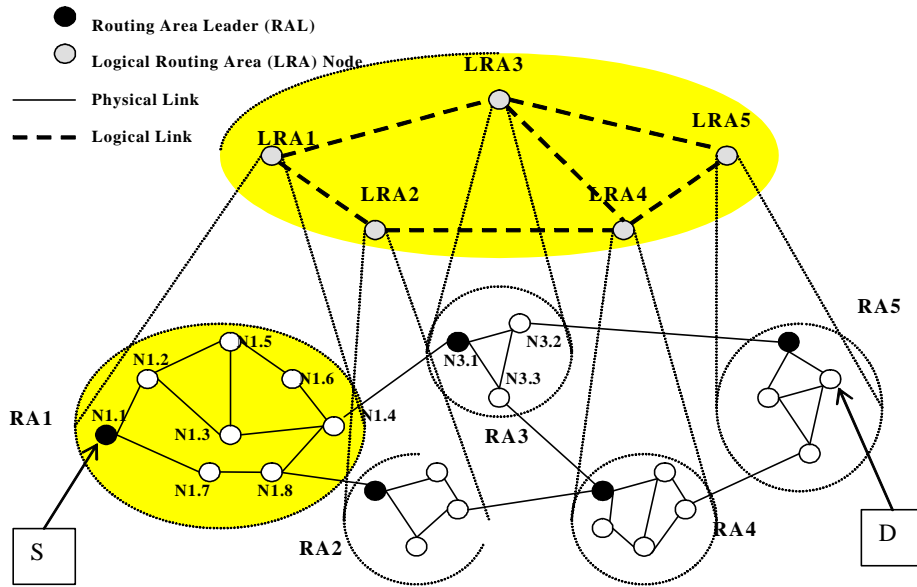


Figure 1. A hierarchical network structure

contained in a Routing Controller (RC) component, which will respond both to requests from connection controllers (CC) for path information needed to set up connections and to requests for topology information from hierarchical mechanisms.

Each RA should be represented by a “Logical Routing Area (LRA) Node” in the next hierarchical level. The required functions to perform this role should be executed by a node called the “Routing Area Leader” (RAL). This node will receive complete topology state information from all RA nodes and will send information up to the LRA node. The propagated information only includes the information needed by the higher level.

There are four basic functional blocks in a hierarchical network structure, the aggregation process, the dissemination process, the update policy and the hierarchical routing algorithm.

A. Network Information Aggregation. As stated above the RAL receives complete topology state information from all the network nodes in its hierarchical level. This information is aggregated before being forwarded to the LRA node. The policy used to define how and which the information is aggregated, is defined by some aggregation scheme.

The main benefit introduced because of using any aggregation scheme is the reduction of the amount of information to be distributed throughout the network. However, a collateral and negative effect of such aggregation scheme is that the information used to compute routes is non-complete, that is, aggregated information does not contain full information about physical links and nodes.

B. Network Information Dissemination. Routing information can be exchanged between adjacent levels of the routing hierarchy i.e. Level L+1 and L, where Level L represents the RAs contained by Level L+1. The links connecting RAs may be viewed as external links, and the links representing connectivity within an RA may be viewed as internal links. Therefore, according to Fig.1 a RAL sends information up to the LRA node. This information summarizes the topology/resource information received by the RAL from all the nodes belonging to the same RA. The communication between adjacent Routing Levels is outside the scope of this paper.

C. Update Policy. Update policies are required to guarantee that the information contained in the network state databases perfectly represents a current picture of the network in order to guarantee an optimal path selection. In general update messages may be triggered by either a periodical refresh (i.e., time-based triggers) or a network change (i.e., threshold-based triggers). While the former does not take into account the network dynamics the latter can drive to a significant signalling overhead in dynamic networks [6], i.e., networks where many new connection setups and releases occur in a short period of time. Thus, new update policies must be developed to reduce this signalling overhead while guaranteeing accurate routing information. However, there is a trade-off between the amount of update messages and the accuracy of the network state information. In fact, the larger the amount of update messages (signalling overhead) the lower the inaccuracy. Since keeping an up-to-date picture of the network is currently not affordable, a certain degree of inaccuracy will always be introduced by any update policy included in the routing protocol.

D. Hierarchical Routing Algorithm. The routing problem in optical networks is referred as the Routing and Wavelength Assignment (RWA) problem. The RWA problem splits the lightpath selection into two subproblems, finding a physical route and finding the wavelength in the selected route that may be used to transport the traffic.

Although ASON specifications do not explicitly recommend a routing algorithm, instead a set of features that have to be supported by any routing algorithm running on an OTN are defined. Source routing is one of these recommendations. In such a routing scenario, the routing decisions are taken on the source nodes based on the global network state information contained in their network state databases. As mentioned above, several causes strongly impact on the network state information accuracy. Unlike traditional flat networks where the inaccuracy is basically introduced by the update policy in hierarchical networks such inaccuracy is introduced not only by the update policy but also by the aggregation scheme used to select the information to be disseminated around the network. There are many references for flat networks (some of them also proposing solutions, see for example [7] and [8]) stating that computing paths based on inaccurate (or outdated) network state information strongly impacts on the network performance. Moreover, when addressing hierarchical networks this problem is even more significant since there are more causes driving to have inaccurate network state information.

3. A Hierarchical Network Approach

The main goal of this section is to present an accurate approach describing in detail some of the issues required by a hierarchical network structure, i.e., an aggregation scheme, an update policy and a new routing algorithm, as stated in previous section. Thorough this paper a multifibre hierarchical network is assumed.

A. Aggregation Scheme. There are many possible aggregation schemes. In [9] authors present a review of some of them as well as a new aggregation scheme named Node Aggregation Scheme (NAS) fully devoted to ASON. A brief description of such an aggregation scheme is now introduced.

Consider a network consisting of Q OXCs. According to the proposed hierarchical structure, an optical network is divided into M RAs connected by border OXCs, each one composed of a set of OXCs with similar characteristics. Let $G(Q, V)$ describe the given physical network, where Q is a set of OXCs and V is a set of fibres connecting the nodes. Let $g(q, v)$ describe the given physical RA, where q is the set of nodes (OXCs) in the RA

and v is a set of links connecting the nodes within the RA. Therefore, $g \hat{I} G$, $q \hat{I} Q$ and $v \hat{I} V$. Each fibre supports c different wavelengths, i.e. c different colours from I_1 to I_c . Moreover, it is assumed that wavelength conversion does not exist in any OXC. Thus, an incoming connection is associated to the same wavelength along the end-to-end lightpath. The aggregation scheme works as follows: Firstly, a RA pre-computes all the existing lightpaths between all border nodes along with the network parameters allocated to each lightpath. Secondly, an aggregation scheme in the RAL summarizes this information reducing the amount of data to be flooded throughout the physical network. Finally, the aggregated information from each RA is grouped in a topology database, which will be used by a source node to compute an end-to-end lightpath. The aggregation process will aggregate the information of several network parameters. The following network parameters were proposed for optical networks:

- D : Propagation delay in a link which is proportional to the fibre distance between two nodes.
- AW_p : Number of available wavelengths of each colour in a link

According to the network parameters proposed for optical networks, the aggregation process performed by the NAS defines two aggregated network parameters, the Aggregated Delay and the Aggregated Number of Available Wavelengths. The aggregation process works as follows:

- Aggregated Delay (D_i):
 - Compute all the lightpaths from node i to all border nodes.
 - Add the propagation delay of each link for each lightpath.
 - Select the minimum value among the values computed in the previous step.
- Aggregated Number of Available Wavelength (AW_p^i):
 - Compute all the lightpaths from node i to all border nodes.
 - Select the minimum number of wavelength per colour that is available on each lightpath.
 - Select the maximum value among the values computed in the previous step.

Formally, the D_i and AW_p^i are defined according to (1) and (2) as follows:

$$D_i = \min_{R_{ij}, \forall j \neq i} \left[\sum_{l \in R_{ij}} D(l) \right], \quad (1)$$

$$AW_p^i = \max_{p=1..c} \left\{ \min_{R_{ij}, \forall j \neq i} \left[\min_{l \in R_{ij}} [AW_p(l)] \right] \right\}. \quad (2)$$

where, R_{ij} is a lightpath between border node i and j , l is a link between two nodes belonging to R_{ij} , $AW_p(l)$ is the number of wavelengths of colour p available in a certain link l , and c is the number of colours per fibre.

Fig.2 illustrates the effects of using the proposed aggregation scheme in terms of entries of the Aggregated TED (ATED). It shows a comparison between the ATED size produced when applying the proposed aggregation scheme and the TED size produced when an aggregation scheme is not used. As expected the ATED size is smaller when applying the NAS scheme.

B. Update policy. The update policy proposed in this paper to be applied to hierarchical optical networks is that defined in [9]. This policy is found by extending those defined for flat networks in [10] (named Absolute change based triggers) and [12] (named threshold based triggers) to be applied to hierarchical networks. In short, a network node triggers an update message when a fixed number N of wavelengths changes their status, i.e. after a

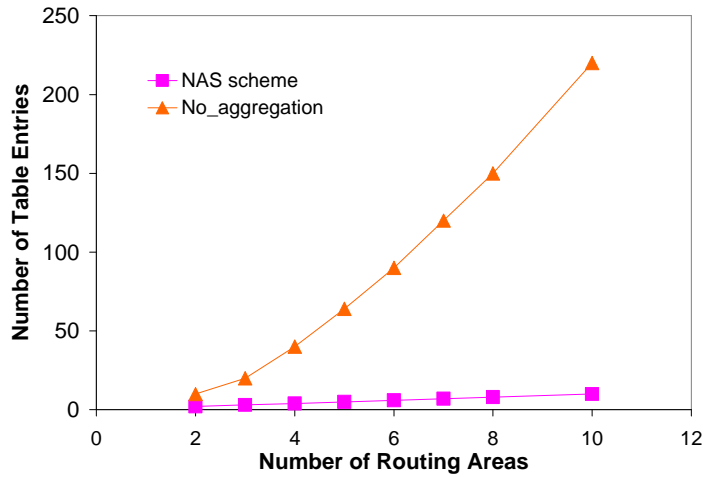


Figure 2. Comparison with/without aggregation

fixed number of N connections are established or released. This update policy is properly modified to be applied to hierarchical routing restricting the update policy within each RA. The update messages sent by the RAL nodes consist of aggregated information and it makes no sense to apply a policy update at this level.

Fig.3 shows the reduction in the amount of update messages flooded throughout the hierarchical network topology of Fig.1, with respect to the N value (number of changes required to trigger an update). As expected, the larger the N the lower the number of update messages. Note that the case of $N = 1$ corresponds to a policy that triggers update messages whenever a change occurs.

C. Routing Algorithm. ASON specifications do not recommend a routing algorithm in order to compute routing paths. However, it defines a set of features that have to be supported by any routing algorithm running over the optical networks. One of them recommends path computation based on source routing. In this paper authors describe and evaluate three different routing algorithms, namely the Balanced Hierarchical Optical Routing (BHOR), the Prediction Hierarchical Optical Routing (PHOR) and the Balanced Prediction Hierarchical Optical Routing (BAPHOR). A preliminary description of all these routing algorithms can be found in [11].

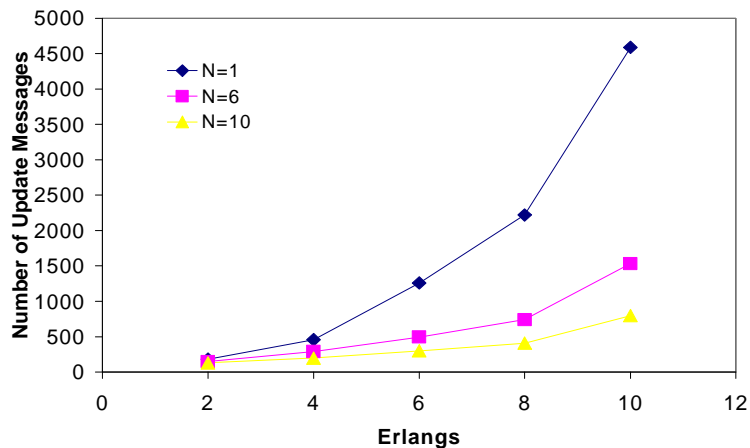


Figure 3. Number of update messages

4. Proposed Hierarchical Routing Algorithms

There are three routing algorithms evaluated in this paper. The BHOR algorithm was defined from the algorithms already proposed in [12] and [9]. In fact, the algorithm proposed in [12], named ALG3, was modified to be applied to hierarchical networks and the algorithm proposed in [9] was also modified to remove routing inaccuracy concerns. The PHOR algorithm is inferred from the Prediction Based Routing mechanism (PBR) presented in [13]. Many algorithms are inferred from the PBR mechanism, all of them based on predicting the resources availability. The main advantages of the PBR are that neither a monitoring process is needed nor any updating messages must be flooded. Results in [13] show the impact on the blocking probability when using an algorithm inferred from the PBR mechanism. The last algorithm evaluated in this paper, BAPHOR, is generated by combining the two above proposed algorithms so including their main benefits while avoiding their potential negative effects.

4.1 BHOR description

According to the ASON specifications routes are dynamically computed on the source nodes reacting to any incoming request. Therefore, when a source node is required to establish a new connection it selects both the k -shortest paths to the destination and an available wavelength. Then it sends a setup message piggybacking the explicit route along the selected lighthpath.

There are three basic components impacting on how optimal the lighthpath selected is:

- the length of the selected lighthpath (Hn)
- the degree of congestion (Cd)
- the degree of obstruction (Od)

The length of the selected path is simply represented by the number of hops. As known, the larger the path is the larger the potential congestion. The degree of congestion can be easily measured by checking the wavelength availability on each node. In fact Cd stands for the minimum number of available wavelengths per colour for every path between the source-destination node pair. Generally speaking high values of Cd for a particular wavelength (colour) mean that this wavelength is not congested. However, this behaviour means that the source nodes must know the network resources availability when computing the paths. Unfortunately, because of the update policy this information may not be accurate enough to guarantee succesful lighthpath decisions. The degree of obstruction tries to minimize the impact of such inaccuracy on the lighthpath selection process. The degree of obstruction Od is measured per wavelength on an end-to-end path and represents the number of links on the path where such a wavelength is defined as potentially obstructed wavelength (POS). This parameter aims to define those wavelengths which are worst updated. For example, if the updating process was time-based those wavelengths which are not updated for long time would be defined as POS. However, being aware that our hierarchical network proposal includes a threshold-based updating, the POS definition must take into account the threshold value, more specifically, a value lower than the threshold value which authors represent by a percentage p_r of that. Being B (any link is a bundle of B fibres) the total number of a certain I_i on a link, R the current number of available (not assigned to an already established lighthpath) I_i on this link we can say that according to the threshold-based update policy, a wavelength I_i is defined as POS, namely I_i^{pos} on a certain link,

when R is lower or equal than p_r . Therefore, the Od value represents the number of links on the end-to-end path where such a wavelength has been defined as POS.

The three parameters defined above should be included in the expression used to compute the lightpaths. So, we propose to represent the weight associated to each link by the factor Od/Cd . This factor stands for a balance between the number of potentially obstructed wavelengths and the real congestion. Since longer paths than the shortest ones can be selected, the length of the path is also included in the path decision. Hence, in order to avoid those paths that are either widest (in terms of wavelength availability) but too long or shortest but too narrow, the weight factor of each path is modeled by $W(I_i)$ according to

$$W(I_i) = Hn \left(\frac{Od}{Cd} \right) \quad (3)$$

By computing $W(I_i)$ we will know the weight per wavelength of an end-to-end path in a RA. However in a hierarchical network a wavelength could be defined as POS on different hierarchical levels, so Eq.(3) must be modified to be applied to different hierarchical levels as well. Let $W_j(I_i)$ be the parameter defined in Eq.(3) per each hierarchical level, let n be the number of hierarchical levels and let $W_h(I_i)$ be the new weigh factor for the end-to-end lightpath. Then, $W_h(I_i)$ can be computed as follows:

$$W_h(I_i) = \sum_{j=1}^n W_j(I_i) \quad (4)$$

Finally, the BHOR algorithm will select the wavelength that minimizes the $W_h(I_i)$ factor.

4.2 PHOR description

As stated above the PHOR is based on extending the initial work presented in [13] where the Prediction Based Routing is proposed for optical transport networks. Before deeply describing the PHOR authors briefly present a short description of the PBR to make the PHOR understanding easier.

A. Prediction Based Routing Overview

The main idea of the PBR mechanism is based on extending the concepts of branch prediction used in computer architecture [14]. The prediction of branch instructions is carried out based on the previous branch instructions behaviour. Considering this idea, the PBR mechanism is based on predicting the route and wavelength assignment between two nodes according to the routing information obtained in previous connections setup. Thus, an important effect of the PBR is that network state information is not used to take routing decisions, so the update process is completely removed. It is worth mentioning that topology update messages are not removed. However these messages are not a problem since typically they are based on refresh time about 30 minutes.

The routing decision is taken from the 'history' of each path. Every source node, for its possible destination node must keep previous information about both wavelength and route allocated to them. This history is repeated all through the time and is stored in a history register, which will be used as a pattern of behaviour. This history register is used to "train" a new table, named Prediction Table (PT). There is one of such registers for every wavelength on every path to every destination node. We name these registers as wavelength registers (WR).

We define a cycle as the basic unit of time where the history state is susceptible to be modified. We propose a method to register the history of the network state in every source node based on assuming that for each cycle, each WR is updated with a 0 value when this wavelength is used on this path on this cycle. Otherwise, the

register of an unused wavelength on a path is updated with a 1. It must be noticed that the expression “a path is used” means that being this path selected by the prediction algorithm it is really available, i.e. correct prediction decision. On the other hand, “a path is unused” when no incoming connection is assigned to this path. The WR length depends on the old wavelength information (per cycle) we want to keep stored.

The PTs are the base to be able to predict a wavelength and a path. In the source nodes one PT is needed for every feasible route between any source-destination node pair. The PT for a wavelength on a path is accessed by using the WR content as an index. In every source node there is the same number of wavelength registers than of PTs. For example, a source node sends traffic towards two different destination nodes and every source-destination node pair has two different paths (two shortest-paths). Moreover, if we assume the existence of 6 wavelengths then 24 PTs are needed on the source node, one for every path and wavelength.

Every entry in the PTs has a counter, which is read when accessing the table. This value is compared to a threshold value. If the value from the table is lower, the prediction concludes that the connection is accepted and the wavelength is used on this path. Otherwise, the path is predicted to be not available. The counters are two-bit saturating counters, where 0 and 1 account for the availability and 2 and 3 account for path unavailability. Saturating counter means that when counter has a value of 0 and it is decreased its new value is also 0, and when its value is 3 and it is increased its value remains at 3. The use of two values to account for the availability or the unavailability has been well studied in the area of branch prediction and is out of the scope of this paper. As it is presented in [14] a two bit counter gives better accuracy than a one bit counter.

As explained above, there is one PT in the source nodes for every wavelength on every path and for every destination. The PTs have to be updated with the same index used on the prediction. When a new connection request is set up the PT of the selected wavelength and path is updated, decreasing the counter. On the other hand, when the connection request has been blocked the counter is increased. The rest of the PTs of the unused paths are not updated. Note that when a connection request is set up only the PT of the wavelength and path used is updated, but all the wavelength registers corresponding to that destination are updated (0 for the used wavelengths and 1 fro the unused wavelengths).

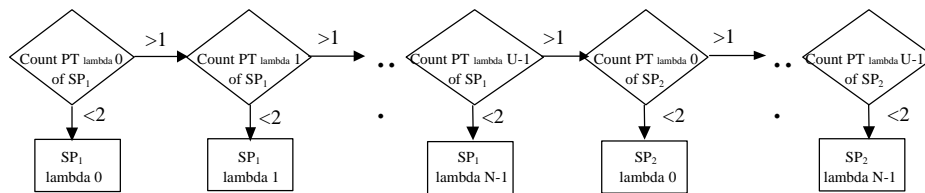


Figure 4. RWP Flowchart

It is worth noting that the PTs updating process in the source nodes is made immediately with the prediction. For this reason it is not necessary to flood any update message through the network to keep network state databases perfectly updated.

B. RWP: The Prediction-based routing algorithm

We define in [13] the Route and Wavelength Prediction algorithm, RWP, inferred from the PBR mechanism. When a new request arrives at the source node demanding a connection to a destination node, all the PTs of the corresponding destination are accessed. It must be noticed that one prediction table, PT, and one wavelength

register, WR, exist for every wavelength on every path to every destination node. We assume that two shortest paths ($k=2$) are computed for every source-destination node pair, SP_1 and SP_2 . The PTs are accessed by one index per table which is built with the wavelength histories contained in the WR. As a consequence of reading the PTs, the 2-bit counters are obtained. In Fig.4 we present a flow chart depicting the RWP performance supposing U wavelengths in every link. The RWP algorithm starts always considering the value of the counter of the PT of the first wavelength on the shortest path, for instance SP_1 . If the counter is lower than 2, and this wavelength is free in the node's outgoing link towards SP_1 , the prediction algorithm decides to use this wavelength on this path. Otherwise this wavelength is not used. In this last case, the value of the counter of the next PT is examined. Notice that next PT corresponds to the second wavelength on SP_1 . When the counters of the PTs of all the wavelengths of SP_1 have been examined, that is, either the counters always are greater than 2 or all wavelengths on the outgoing link towards SP_1 are not available, the prediction algorithm checks the PTs of the next path, SP_2 , and so on. When the prediction algorithm, after checking all PTs, decides that all the feasible wavelengths on the two paths are blocked, then it tries to forward the connection request through the first available wavelength on the outgoing link towards one of the two shortest paths either SP_1 or SP_2 . The information about the outgoing links of the source node is always known by the source node.

Wavelength registers (WR) are updated depending on which wavelength is used and whether the request is blocked or not. Also the PT of the used wavelength and path is updated by either increasing (means connection blocked) or decreasing (means connection not blocked) the counter of the corresponding entry in the PT.

It is worth mentioning that counters of every wavelength on all the feasible paths between a source-destination node pair can be read, so allowing the prediction to be made, before a new connection request reaches the source node. It is a very significant factor which substantially reduces the cost involved with the PBR mechanism. In fact, even though several tables must be accessed to make the prediction, these accesses can be done offline. For every possible new request, the decision of which path to use is already done

C. The PHOR: Extending the RWP to hierarchical optical networks

In this subsection we shortly describe the PHOR algorithm [11] (Prediction Hierarchical Optical Routing), which is based on modifying the RWP algorithm to be applied to hierarchical networks. The main advantages of introducing the PBR concept in hierarchical networks is that neither update messages are required nor any aggregation process.

The algorithm works as follows. The k -shortest routes, A and B (assuming $k = 2$) are precomputed in the source nodes for every destination node. If such a destination node belongs to the same RA the path is completely defined. Otherwise, if the destination node belongs to a different RA, the route is specified by both the route from the source node to the last node on this RA and the different RAs for the rest of the route. For example in Fig.1 assuming $k = 2$, if the source node is the N1.1 and the destination node is the N5.4, the two shortest routes are N1.1-N1.2-N1.5-N1.6-RA3-RA5 (A) and N1.1-N1.7-N1.8-RA2-RA4-RA5 (B). There is one WR and one PT for every wavelength for these two routes. We assume that being A and B link disjoint routes, A stands for the shortest route and B is equal or larger than A.

Wavelengths on each route are weighted according to the minimum number of available wavelengths of every colour per link along the lightpath. This weight is used to order all different possibilities to set up the lightpath. It is important to note that this information is from the point of view of the source node N1.1. This source node only knows how many wavelengths has assigned in every link but it does not know the real

availability of the links because there are not update messages. The PTs are checked in this order. The decision of which wavelength and route are selected is done depending on the value of the counters of the PTs and the availability of the node's output links.

4.3 BAPHOR description

The algorithms presented in the previous subsection have some advantages and weaknesses. In order to mix the benefits while reducing the weaknesses of such algorithms we propose a hybrid routing algorithm named BAPHOR (Balanced Prediction Hierarchical Optical Routing) [11]. The main idea underlying such algorithm is that the aggregated network state information of the external RAs can be replaced by a prediction about the availability through the external RAs. On the other hand, the network state information within the RA is flooded by an update policy and utilized by a balanced routing algorithm. Summarizing, the aggregation schemes are not necessary because the network state information is not flooded between different RAs. Nevertheless, updating is needed into every RA. Such scheme makes the dissemination process easier since dissemination is only limited to RAs scenarios.

A. Wavelengths Registers, Prediction Tables and Database Tables

In every source node there is a Database Network State Table containing the information about availability of all the internal links of its RA. This database is updated depending on the frequency of updating, that is, the parameter N. Moreover, although there is one WR and one PT for every route and wavelength for every possible destination in every source node, now the source and the destination nodes are nodes in belonging to different hierarchical levels. For example, from the Fig.1, a possible route between RA1 and RA5 is RA1-RA3-RA5, and in the node N1.1 of the RA1 there would be one PT and one WR for every wavelength for the route RA1-RA3-RA5.

B. The Hybrid Algorithm BAPHOR

The BAPHOR algorithm performs the lightpath decision targeting to minimize the $W_h(I_i)$ value, as it is done in the BHOR algorithm. However, the $W_h(I_i)$, that is the hierarchical $W(I_i)$ value, is computed in a different manner. The BHOR calculates the $W_h(I_i)$ value by adding the $W(I_i)$ value of each hierarchical level which is $H_n\left(\frac{Od}{Cd}\right)$. In the BAPHOR algorithm, in the first hierarchical level (into the RA) the $W(I_i)$ value is also calculated as $H_n\left(\frac{Od}{Cd}\right)$, but for the others hierarchical levels the value to be added is the value of the corresponding PT counter. This is expressed in Eq. (5) when the number of hierarchical levels is n.

$$W_h(I_i) = H_n\left(\frac{Od}{Cd}\right)_{j=1} + \sum_{j=2}^n PTcounter(j) , \quad (5)$$

The $W(I_i)$ values and PT counter values can be mixed in this manner because both stand for high availability when they are low, and stand for light availability when they are high. Assuming k-shortest paths with $k = 2$, for every possible source-destination node pair the two shortest routes, A and B are precomputed. The algorithm selects the route, A or B, and the wavelength that minimizes the $W_h(I_i)$ value. When two different $W_h(I_i)$ have the same value, the algorithm selects the route and wavelength with higher $Cd(I_i)$ value, that is the route and wavelength with more resources availability.

5. Performance Evaluation

Before evaluating our proposal, we present two different examples to illustrate its performance both based on the topology shown in Fig1. The former depicts the BOHR behavior when the NAS is used to aggregate the network state information. The latter illustrates the BAPHOR behavior.

5.1 Illustrative examples

Considering that every RA includes control functions with signalling capabilities, we assume $B = 10$ fibres per link and 4 wavelengths per fibre. Update messages are sent according to $N = 6$ and a wavelength is defined as POS according to a percentage $p_r = 50\%$ (i.e., when the minimum number of available wavelengths on this link is lower than or equal to 3). Suppose that incoming call requests arrive between nodes S and D in Fig.1.

A. Example 1: Illustrating the BHOR performance.

Shaded areas in Fig.1 show the whole network perspective seen by nodes belonging to Routing Area 1 (RA1). Each node has a topology database containing both complete topology information of the RA1 and information summary of the rest of the network. This is represented in Fig.1 by three different types of links, which are grouped on either physical or logical links. For example, physical link N1.1-N1.2 stands for available resources in the link connecting nodes N1.1 and N1.2. The physical link N1.4-RA3 stands for the available resources in the link connecting RA1 to RA3 through node N1.4. Finally, logical link RA1-RA3 stands for the aggregated information of RA3 from the node directly connected to RA1, i.e., N3.1. This aggregated information of the logical link RA1-RA3 has been obtained according to the information of the topology and available resources database shown in Table 1 and an aggregation process applied to that database, which is described as follows.

Table 1. Topology/Available Resource Database in RA3

Link	λ_1	λ_2	λ_3	λ_4	D
N3.1-N3.2	2	3	4	7	1
N3.1-N3.3	1	5	3	2	1
N3.3-N3.2	4	1	6	5	1

Firstly, the Node Aggregation Scheme (NAS) summarizes the information of Table 1, according to the expressions (3) and (4). Feasible lightpaths from N3.1 to N3.2 and N3.3 are: a) N3.1-N3.2, b) N3.1-N3.3 and c) N3.1-N3.3-N3.2. Therefore, according to (2) the aggregated available wavelength for each color is represented in (6). Moreover, according to (1) the aggregated delay is also represented in (7).

$$AW_1^1 = 2; AW_2^1 = 3; AW_3^1 = 4; AW_4^1 = 7, \quad (6)$$

$$D_1 = \min_{a,b,c} \left[\sum_{l \in a,b,c} D(l) \right] = \min[1,1,2] = 1. \quad (7)$$

Secondly, N3.1 bundles its information and disseminates it throughout the RA's according to a flooding mechanism proposed in [12].

Finally, Table 2 depicts the topology database in RA1 built according to the received dissemination messages. In particular, the topology database provides all the information required to compute a route from the given node to any reachable node. When a call request from node s (N1.1) to node d (RA5) reaches node N1.1, this node applies BHOR to select the lightpath based on the information represented in Table 2.

Table 2. Topology/Available Resource Database in N1.1 (RA1)

Link	λ_1	λ_2	λ_3	λ_4	D	Link	λ_1	λ_2	λ_3	λ_4	D
N1.1-N1.2	6	3	3	6	1	N1.8-N1.4	6	6	0	6	1
N1.2-N1.3	2	3	6	0	1	N1.4-RA3	6	7	7	5	1
N1.3-N1.4	6	3	0	2	1	N1.8-RA2	5	6	7	5	1
N1.2-N1.5	6	2	0	1	1	RA1-RA2	1	2	3	5	1
N1.5-N1.3	6	6	6	6	1	RA1-RA3	2	3	4	7	1
N1.5-N1.6	0	7	3	3	1	RA3-RA4	4	4	5	7	3
N1.6-N1.4	1	1	1	1	1	RA3-RA5	4	3	4	7	5
N1.1-N1.7	6	3	1	6	1	RA2-RA4	4	3	3	6	5
N1.7-N1.8	0	4	6	1	1	RA4-RA5	4	3	3	5	5

Table 3 shows possible shortest paths from node s to node d as well as the Od and the Cd values per wavelength represented by the parameter $I_i(Od, Cd)$ on each hierarchical level.

Table 3. Routing Table in N1.1 (RA1)

Route 1	I_1	I_2	I_3	I_4	DI_1	$I_i(Od, Cd)$
N1.1-N1.2-N1.3-N1.4-RA3	2	3	0	0	4	$I_1(1,2), I_2(3,3)$
RA1-RA3-RA5	2	3	4	7	1	$I_1(1,2), I_2(1,3)$
Route 2	I_1	I_2	I_3	I_4	D	$I_i(Od, Cd)$
N1.1-N1.7-N1.8-RA2	0	3	1	1	3	$I_2(1,3), I_3(1,1), I_4(1,1)$
RA1-RA2-RA4-RA5	1	2	3	5	4	$I_1(1,1), I_2(1,2), I_3(1,3)$

Once shortest paths are selected and the $I_i(Od, Cd)$ has been defined, the BHOR must select the wavelength to be used. This is performed by computing the W_h value for each wavelength on both paths according to Eq. (4). Table 4 illustrates the $W_h(I_i)$ computation. BHOR selects that wavelength minimizing the $W_h(I_i)$ value. Therefore, the first route is selected and I_1 will be used to transport the traffic.

Table 4. W_h value for Route 1 and Route 2

Route 1	$W(I_1)$	$W(I_2)$	Route 2	$W(I_2)$	$W(I_3)$	$W(I_4)$
N1.1-N1.2-N1.3-N1.4-RA3	2	4	N1.1-N1.7-N1.8-RA2	1	3	3
RA1-RA3-RA5	0.5	0.33	RA1-RA2-RA4-RA5	2	0.67	0
$W_h(I_i)$	2.5	6	$W_h(I_i)$	4.67	6.17	4.67

B. Example 2: Illustrating the BAPHOR performance

We use the same example to illustrate the BAPHOR performance. There are two routes from each node belonging to RA1 to the other routing areas and a prediction table, PT, for every route and every wavelength from RA1 to the other routing areas. For example, Table 5 shows precomputed shortest routes in N1.1. Table 6 shows the database of the node N1.1. This database has the complete topology information about RA1 as well as a PT for every route to the rest of the network.

Table 5. Precomputed shortest routes

Source-destination pair	Route A	Route B
RA1-RA2	RA1-RA2	RA1-RA3-RA4-RA2
RA1-RA3	RA1-RA3	RA1-RA2-RA4-RA3
RA1-RA4	RA1-RA2-RA4	RA1-RA3-RA4
RA1-RA5	RA1-RA3-RA5	RA1-RA2-RA4-RA5

Table 7 illustrates the Od and Cd values of the first hierarchical level (into RA1) and the values of the counters of the two shortest paths selected between the source (node N1.1) and the destination (one node of the RA5, for example N5.4). In this case the $W_h(I_i)$ of each wavelength is calculated adding the value of the $W(I_i)$ of the first hierarchical level with the value of the PT counter of the route through the second hierarchical level. The BAPHOR algorithm selects the wavelength and route that minimizes $W_h(I_i)$, that is route B and I_4 .

Table 6. Database Table and Prediction Tables

<i>Link</i>	I_1	I_2	I_3	I_4	D	<i>Route</i>	I_1	I_2	I_3	I_4	D
N1.1-N1.2	6	3	3	6	1	PT Counters <i>RA1-RA2</i> (Route A)	2	1	0	3	1
N1.2-N1.3	2	3	6	0	1	PT Counters <i>RA1-RA2</i> (Route B)	2	3	1	2	3
N1.3-N1.4	6	3	0	2	1	PT Counters <i>RA1-RA3</i> (Route A)	2	3	0	2	1
N1.2-N1.5	6	2	0	1	1	PT Counters <i>RA1-RA3</i> (Route B)	1	2	0	1	3
N1.5-N1.3	6	6	6	6	1	PT Counters <i>RA1-RA4</i> (Route A)	0	0	1	3	2
N1.5-N1.6	0	7	3	3	1	PT Counters <i>RA1-RA4</i> (Route B)	0	1	3	2	2
N1.6-N1.4	1	1	1	1	1	PT Counters <i>RA1-RA5</i> (Route A)	3	1	2	0	2
N1.1-N1.7	6	3	1	6	1	PT Counters <i>RA1-RA5</i> (Route B)	0	1	3	2	3
N1.7-N1.8	0	3	6	6	1						
N1.8-N1.4	6	6	0	6	1						
N1.4- <i>RA3</i>	6	7	7	5	1						
N1.8- <i>RA2</i>	5	6	7	5	1						

Table 7. Computing the $W_h(I_i)$ values for the BAPHOR algorithm.

<i>Route A</i>	I_1	I_2	I_3	I_4	D	$I_i (Od, Cd)$	$W(I_1)$	$W(I_2)$	
<i>N1.1-N1.2-N1.3-N1.4-RA3</i>	2	3	0	0	4	$I_1(1,2), I_2(3,3)$	2	4	
<i>RA1-RA3-RA5</i>	3	1	2	0	2	I_1	3	1	
							$W_h=5$	$W_h=5$	
<i>Route B</i>	I_1	I_2	I_3	I_4	D	$I_i (Od, Cd)$	$W(I_2)$	$W(I_3)$	$W(I_4)$
<i>N1.1-N1.7-N1.8-RA2</i>	0	3	1	5	3	$I_2(2,3), I_3(1,1)$	2	3	
<i>RA1-RA2-RA4-RA5</i>	0	1	3	2	3	I_3, I_4	1	3	2
							$W_h=3$	$W_h=6$	$W_h=2$

5.2. Evaluation

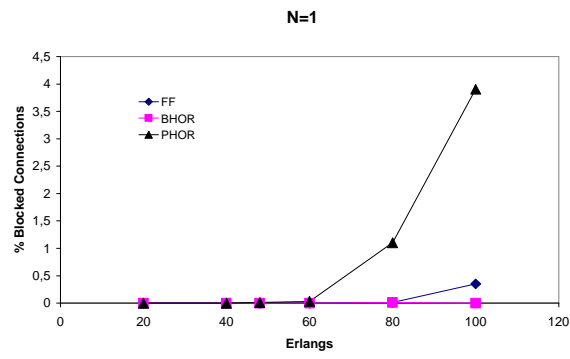
Once the proposed hierarchical network structure has been analyzed by the illustrative examples presented above, we evaluate our proposal by simulation. We assume the topology shown in Fig.1, but unlike the illustrative examples, we suppose a 5-fibre topology, with 16 wavelengths on all the fibres on all the bi-directional links. Connection arrivals are modeled by a Poisson distribution and the connection holding time is assumed to be exponentially distributed. The algorithms behaviour is measured in terms of percentage of blocked connections.

A. Preliminary Evaluation.

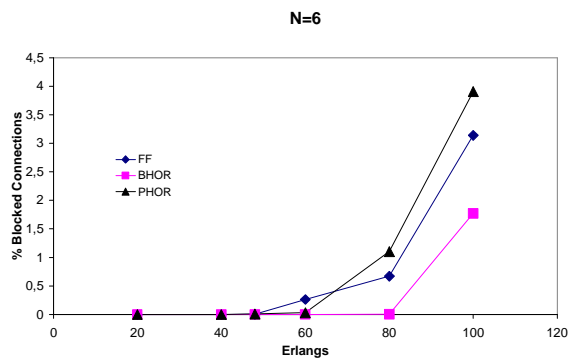
We evaluate in Fig.5 the performance of the PHOR, the First-Fit (with NAS), the BHOR (with NAS) and the PHOR in terms of the connection blocking probability. A set of simulations are carried out varying the traffic load between 0 and 100 Erlangs where the total number of connection requests is 20000 on each simulation run. All routing algorithms compute two shortest routes (A and B).

Results shown in Fig.5 are obtained for all the algorithms ranging the threshold updating N , between $N = 1$ (Fig.5.a), $N = 6$ (Fig.5.b.), $N = 10$ (Fig.5.c.) and $N = 20$ (Fig.5.d.). It is important to notice both the PHOR algorithm does not need update messages nor any aggregation scheme and the larger the N value the lower the signaling overhead. According to the results shown in Fig.5 we can conclude the following. On the one hand, from the point of view of performance, for low values of N (i.e., $N \leq 6$) the BHOR is the algorithm presenting the lower blocking probability and the PHOR is the worst. This trend changes as the value of N increases. In fact, for $N = 20$, the BHOR exhibits the worst behavior, while the PHOR is the best.

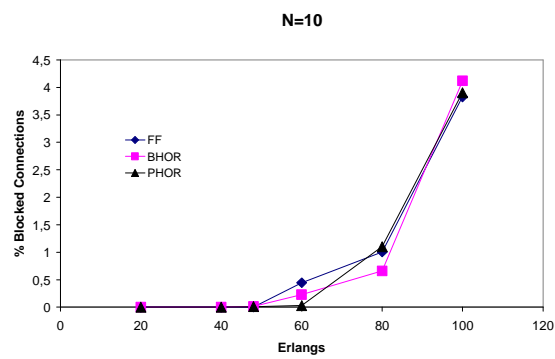
On the other hand, from the point of view of the signalling overhead and computation complexity, the PHOR is the best option since neither update messages nor aggregation schemes are required.



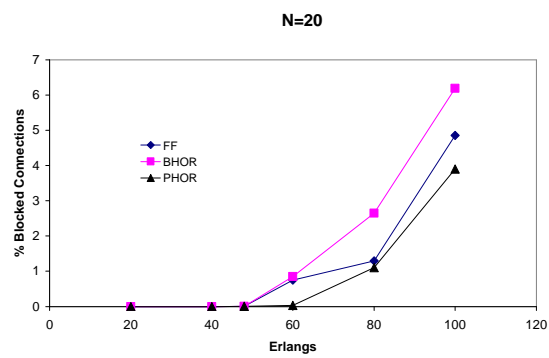
5.a. Results for N=1



5.b. Results for N=6



5.c. Results for N=10



5.d. Results for N=20

Figure 5. Connection blocking for the First-Fit, the BHOR and the PHOR algorithms.

Comparing the results obtained for the BHOR and the First-Fit algorithms in Fig.5 we can observe that while for $N = 1$ and $N = 6$ the BHOR behaves better than the First-Fit for all the traffic loads, the First-Fit exhibits better results than the BHOR for $N = 10$ and $N = 20$ with high traffic loads.

This is due to the special characteristics of the BHOR that uses the N when computing routes. As it is explained in subsection 5.1 the links of the routes having less than 50% of N available fibres determine the degree of obstruction of the route. In our simulations the number of fibres is 5 so that being for example $N = 6$ (the 50% (p_r) of N is 3), the links with 3 or lower available fibres contribute to the degree of obstruction. However, when $N = 10$ or $N = 20$, all the links are contributing to the degree of obstruction, since the number of available fibres is always lower or equal than 5 or 10 (computed according to p_r). In this scenario, the W factor becomes quadratic dependent with the number of hops in the route, and for this reason the algorithm tries to assign the shortest routes. Because of such an assignment, when the traffic load is high these shortest routes are heavily congested.

Because of this we propose the BAPHOR algorithm which aims to combine the benefits of the BHOR and PHOR algorithms. Fig.6 shows the results in percentage of blocked connections as a function of N , for different traffic loads. We can state that while the First-Fit and the BHOR algorithms behave worse than the PHOR algorithm (not affected by the N value) for high values of N , the best algorithm is the BHOR for low values of N .

B. Global Evaluation.

In order to evaluate the BAPHOR algorithm performance we compare the blocking probability produced by the BHOR (NAS), PHOR and BAPHOR algorithms when traffic load varies from 0 to 100 Erlangs. The results for the BHOR and the BAPHOR algorithms are presented for 4 different values of N , $N = 1$ (Fig.7.a), $N = 6$ (Fig.7.b), $N = 10$ (Fig.7.c) and $N = 20$ (Fig.7.d). Note that the PHOR does not vary with the N value since it does not need any update messages. We can observe that the BAPHOR algorithm better tolerates high values of N than the BHOR. This is because the routing decision is carried out also including prediction issues. Moreover we can see that up to 50 Erlangs all the algorithms has similar performance. Instead, from 50 Erlangs on the

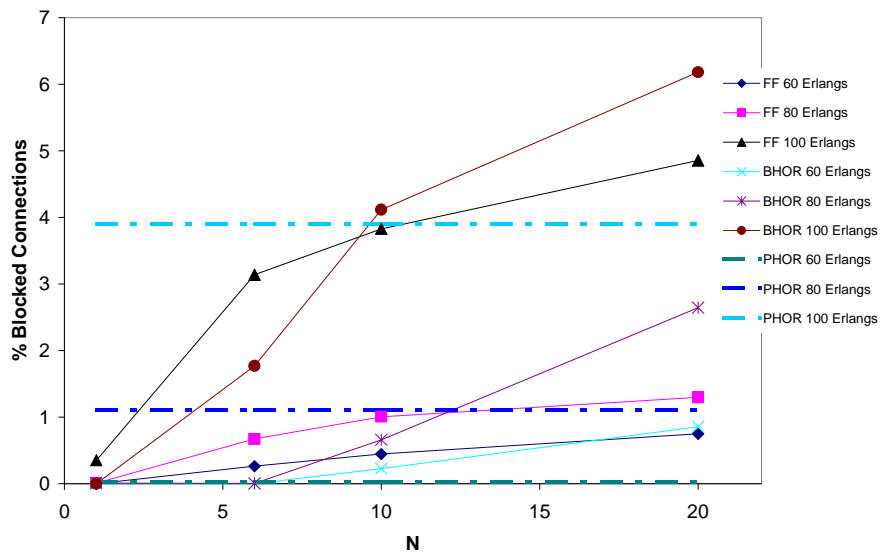
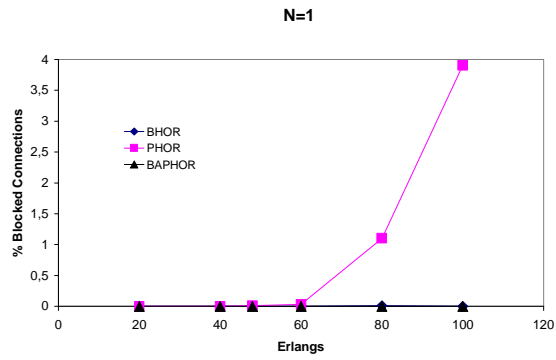
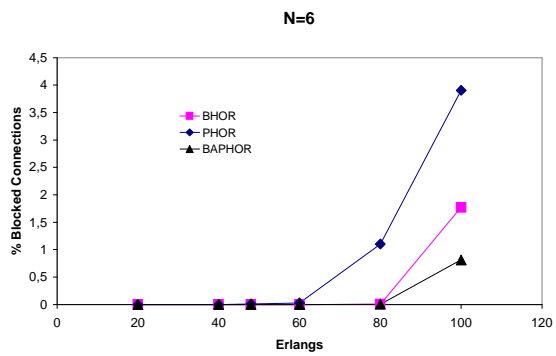


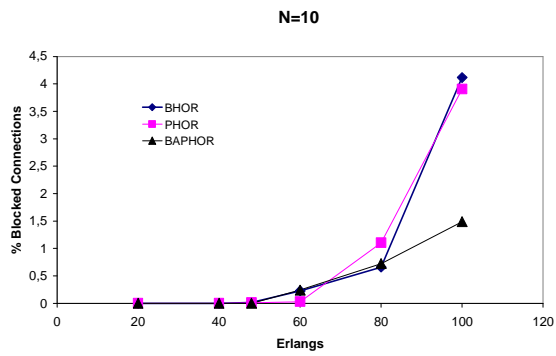
Figure 6. Connection blocking for the First-Fit, the BHOR and the PHOR algorithms depending on the updating frequency



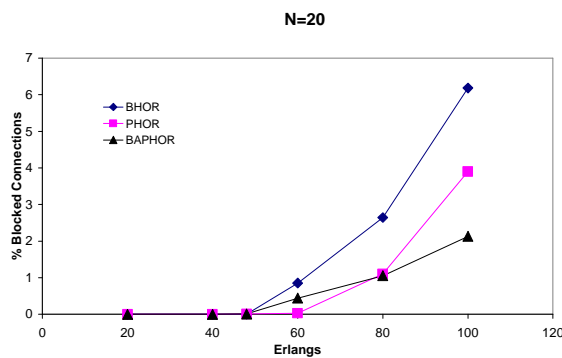
7.a. Results for N=1



7.b. Results for N=6.



7.c. Results for N=10.



7.d. Results for N=20.

Figure 7. Connection blocking for the BHOR, the PHOR and the BAPHOR algorithms.

connection blocking strongly depends on the value of N . However, after analyzing all the graphs included in Fig.7 we can conclude that the lower connection blocking is obtained by the BAPHOR algorithm. This is justified because the BAPHOR algorithm combines the benefits of both the PHOR algorithm, i.e., prediction issues, and the BHOR algorithm, i.e., load balance and congestion reduction.

Finally, Fig.8 shows the connection blocking behavior for the BHOR, PHOR and BAPHOR algorithms as a function of the value of N . While the BAPHOR algorithm behaves similarly than the BHOR algorithm for low values of N , the BAPHOR (and also the PHOR algorithm) better tolerates high values of N for high traffic loads compared to the BHOR algorithm.

6. Conclusions

In this paper we propose and evaluate a hierarchical network architecture for optical transport networks mainly focusing on routing concerns aiming to address scalability issues. We decompose the hierarchical routing problem into three main issues: the aggregation process used to reduce disseminated information throughout the network; the update policy used to keep network state databases perfectly updated; and the lightpath selection process. After introducing an in-depth description of a hierarchical network approach, proposing an aggregation scheme and an update policy, we focus on the routing problem, so proposing three different routing algorithms, namely the Balanced Hierarchical Optical Routing (BHOR), the Prediction Hierarchical Optical Routing (PHOR) and the Balanced Prediction Hierarchical Optical Routing (BAPHOR). The PHOR and the BAPHOR algorithms are inferred from the Prediction-based routing mechanism, already proposed by the authors in optical networks. The main skill of the PBR is to provide source nodes with the capability of taking routing decisions without using the traditional routing information hence removing the required flooding of update messages.

Evaluation results show the benefits obtained in the network performance in terms of blocking probability when computing lightpaths based on these three proposed algorithms. Simulation results lead to conclude that

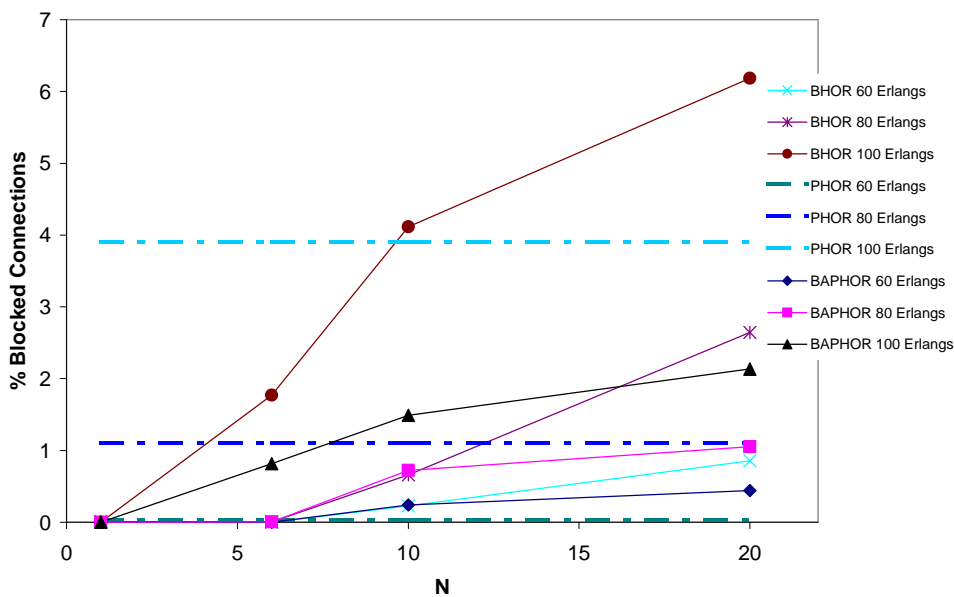


Figure 8. Connection blocking for the BHOR and the PHOR algorithms depending on the updating frequency

the best network performance is obtained by the BAPHOR algorithm which combines the hierarchical approach and the prediction algorithm.

Acknowledgements

This work was partially funded by the TRIPODE MCyT (Spanish Ministry of Science and Technology) under contract FEDER-TIC2002-04344-C02-02 and the CIRIT (Catalan Research Council) under contract 2001-SGR00226.

References

- [1] ITU-T Rec.G8080/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)", November 2001.
- [2] P.Ashwood, et al., "Generalised MPLS – Signalling Functional description", draft-ietf-mpls-generalised-signalling-09.txt, February 2003.
- [3] Wesam Alanzar, "Requirements for Generalized MPLS (GMPLS) Routing for Automatically Switched Optical Network (ASON)", draft-ietf-ccamp-gmpls-ason-routing-reqts-04.txt, May 2004.
- [4] G.7715 ITU-T Rec. G.7715/Y.1306, "Architecture and Requirements for the Automatically Switched Optical Network (ASON)," June 2002.
- [5] G.7715.1 ITU-T Draft Rec. G.7715.1/Y.1706.1, "ASON Routing Architecture and Requirements for Link State Protocols," Nov. 2003.
- [6] H.Zang, L.Sahasrabudde, J.P.Jue, S.Ramamurthy, B.Mukherjee, "Dynamic lightpath establishment in wavelength-routed WDM networks", IEEE Communications Magazine, pp. 100-108, September 2001.
- [7] K.Lu, J.P.Jue, G.Xiao, I.Chlamtac, T.Ozugur, "Intermediate-Node Initiated reservation (IIR): A New Signaling Scheme for Wavelength-Routed Networks", IEEE JSAC, vol.21, n°:8, October 2003.
- [8] S.Shen, G.Xiao, T.H.Cheng, "Evaluating the impact of the link-state update period on the blocking performance of wavelength-routed networks", OFC 2004.
- [9] X.Masip-Bruin, S.Sánchez-López, J.Solé-Pareta, J.Domingo-Pascual, E.Marín-Tordera, "Hierarchical Routing with QoS Constraints in Optical Transport Networks", in Proc. of Networking'04, Athens, Greece, May 2004.
- [10] S.Sengupta, D.Saha, S.Chaudhuri, "Analysis of Enhanced OSPF for Routing Lightpaths in Optical Mesh Networks", in Proc. of IEEE ICC2002.
- [11] E.Marín-Tordera, X.Masip-Bruin, S.Sánchez-López, "New Routing Approach for Hierarchical Optical Transport Networks", Technical Report, March 2005 <http://www.ac.upc.edu/~dacsecre/reports/2005/10/hierarouting.pdf>.
- [12] X.Masip-Bruin, S.Sánchez-López, J.Solé-Pareta, J.Domingo-Pascual, D.Colle, "Routing and Wavelength Assignment under Inaccurate Routing Information in Networks with Sparse and Limited Wavelength Conversion", in Proc. IEEE GLOBECOM 2003, San Francisco, USA, December 2003
- [13] E.Marín-Tordera, X.Masip-Bruin, S.Sánchez-López, J.Solé-Pareta, J.Domingo-Pascual, "A New Prediction-Based Routing and Wavelength Assignment Mechanism for Optical Transport Networks", in Proc. of QofIS'04, Barcelona, September 2004.
- [14] Smith J.E., "A study of branch prediction strategies", In Proc. of 8th International Symposium in Computer Architecture, Minneapolis 1981.



EVA MARIN-TORDERA obtained the MS degrees in Physics in 1993 and in Electronic Engineering in 1998 both from the Barcelona University. Actually she is pursuing her PhD at the Technical University of Catalonia where she works as an assistant professor. She has published many papers in national and international conferences. Her main interests focus on QoS provisioning, traffic engineering, and optical networks. She is actively participating in the IST projects E-NEXT, e-Photon/One and the national projects TRIPODE, SAM.



XAVIER MASIP-BRUIN received the M.S. and Ph.D. degrees from the Technical University of Catalonia, Barcelona, Catalonia, Spain, both in Telecommunications Engineering, in 1997 and 2003 respectively. He is currently an Associate Professor of Computer Science with the Technical University of Catalonia. His current research interests lies in the field of Broadband Communications, QoS management and provision and traffic engineering. His recent work has focused on QoS provisioning both in IP/MPLS and Optical Transport Networks. His publications include around 40 papers in national and international refereed journals and conferences. Since 2000 he has participated in many research projects: in IST projects E-NEXT, NOBEL and EuQoS; and in Spanish research projects SABA, SABA2, SAM and TRIPODE



SERGIO SANCHEZ-LOPEZ got his BS degree in telecommunication engineering in 1989 from Polytechnic University of Catalonia, his MS degree in electrical engineering in 1996 from University of Barcelona and he got his PhD in Telecommunications Engineering in 2003 from Polytechnic University of Catalonia, Spain. He is an associate professor in the Computer Architecture Department at Polytechnic University of Catalonia since 1992. His research interests include ATM, MPLS and Optical Networks. He has been involved in the IST projects LION, NOBEL, EuQoS and e-Photon/One.



JOSEP SOLE-PARETA was awarded his Master's degree in Telecommunication Engineering in 1984, and his Ph.D. in Computer Science in 1991, both from the *Universitat Politècnica de Catalunya* (UPC). In 1984 he joined the Computer Architecture Department of UPC. Since 1992 he has been an Associate Professor with this department. He is co-founder and member of the Advanced Broadband Communications Centre of UPC (<http://www.ccaba.upc.es>). His current research interests are in Broadband Internet and High Speed and Optical Networks with emphasis on traffic engineering, traffic characterisation, traffic management, MAC protocols and QoS provisioning. He has participated in many ACTS IST European project devoted to Optical Networking, such as SONATA, LION, DAVID and The action COST 266. Within the VI Framework Program he is participating in NOBEL (IP-project), in e-Photon/One (Network of Excellence) and in the Action COST 291.



JORDI DOMINGO-PASCUAL is Professor of Computer Science and Communications at the Universitat Politècnica de Catalunya (UPC) in Barcelona. There, he received the Engineering Degree in Telecommunication (1982) and the Ph.D. Degree in Computer Science (1987). Since 1983 he is lecturer at the Computer Architecture Department. His research topics are Broadband Communications and Applications. Since 1988 he has participated in RACE projects (Technology for ATD and EXPLOIT), in several Spanish Broadband projects (PLANBA: AFTER, TR1 and IRMEM), ACTS projects (INFOWIN, MICC, IMMP) and IST projects (LONG, E-NET, EuQoS, E-NEXT). Since 1995 he is researcher at the Advanced Broadband Communications Center of the University (CCABA). Currently is in charge of the i2CAT next generation Internet infrastructure project (GigaCAT).