Dynamic grooming in IP over optical networks based on the overlay architecture

E. Salvadori, R. Lo Cigno, Z. Zsóka

Dipartimento di Informatica e Telecomunicazioni – Università di Trento, Italy
Department of Telecommunications, Budapest University of Technology and Economics, Hungary

Received 4 April 2005; received in revised form 18 January 2006; accepted 17 April 2006
Available online 23 June 2006

Abstract

This paper defines a formal framework for the definition of dynamic grooming policies in IP over optical networks. The formal framework is then specialized for the Overlay Architecture, where the control plane of the IP (Internet protocol) and optical levels are separated, and no information is shared between the two.

We define a family of grooming policies for the Overlay Architecture based on constraints on the number of hops and on the bandwidth sharing degree at the IP level, and we analyze the performances as a function of the grooming parameters in regular and irregular topologies.

Results are derived using realistic traffic models that depart from the circuit-like traffic traditionally used in grooming studies.

Keywords: IP over optical; Grooming; Overlay architecture; Routing; Elastic traffic; Traffic engineering; Simulations

1. Introduction

The IP protocol and optical transmission techniques are going to play a fundamental role in the networking scenario of the coming years — if not decades. The presence of these two pillars is an easy bet, and it is a widespread idea that most of the intermediate network management layers will gradually disappear, leaving a scenario where IP packets are carried directly on high-speed wavelength division multiplexing (WDM)-based optical connections [1]. Optical packet switching and optical burst switching are elegant long-term solutions for the natural integration of optical transmission within a packet-based IP network; however, more traditional architectures, where packets are electronically switched by routers connected by circuit-switched optical connections, are going to dominate the commercial market for a long period. GMPLS (Generalized MPLS) is also going to play its part in this scenario, enabling the use of Traffic Engineering (TE) techniques at the IP level, substantially modifying the IP behavior from pure datagram to “virtually connected”. The Label Switched Paths (LSPs) behave as logical connections carrying elastic traffic that can adapt its rate to

This work was supported by the Italian Ministry of Education and Research (MIUR) through the FIRB GRID.IT project, and by the Hungarian Italian Intergovernmental S&T Cooperation Programme for 2004–2007, project reference number I-17/03. The project is developed under the European Union E-NEXT Network of Excellence umbrella.

* Corresponding author now at: Create-Net Research Center, Via Solteri 38, Trento, Italy. Tel.: +39 0461 408400; fax: +39 0461 421157.
E-mail addresses: salvador@dit.unitn.it, elio.salvadori@create-net.it (E. Salvadori), locigno@dit.unitn.it (R. Lo Cigno), zsoka@hit.bme.hu (Z. Zsóka).

1573-4277/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.osn.2006.04.005
the network conditions. In this scenario, the interaction between routing and control of the circuit-switched optical network and that of the packet- (or label-) switched IP network is of the utmost importance for the end-to-end performance and the efficient exploitation of network resources. Traffic grooming lies at the heart of this problem.

Traffic grooming is the dynamic multiplexing capability aimed at optimizing the capacity utilization in transport systems by means of the combination of low-speed traffic streams onto high-speed (optical) channels. This problem is a variant of the well-known virtual topology design problem and has received a lot of attention in recent years (see [2] for a review).

In order to fix ideas, let us assume that, due to end-users traffic demands, a service provider dynamically generates LSP requests $f_{sd}$ between sources $s$ and destinations $d$. The nodes are Label Switched Routers (LSRs) interconnected by Optical Cross Connects (OXC) that can only switch entire lightpaths from one fiber to another. The action of mixing LSPs onto lightpaths and splitting them from lightpaths is a grooming action. The interaction between the IP and the optical levels, and the policy used in updating the virtual topology (e.g., aggressive or cautious use of optical resources) define the grooming policy within the network. In the rest of the paper, we use the terms (IP) router, LSR, and G-OXC (Grooming-capable OXC) interchangeably, although there might be some differences in specific implementations of the different devices. We do not make any specific assumption on the internal architecture of nodes (e.g., limitations on the number of ports) or on the implementation of the grooming fabric, since the focus of the paper is on the global architecture and not on specific implementations.

1.1. Static and dynamic grooming

There are two main approaches to traffic grooming: static and dynamic. Static grooming refers to some network cost minimization when the traffic matrix is known in advance. The actions undertaken by the network when a request $f_{sd}$ arrives are completely pre-defined, and the network status does not influence decisions. Dynamic grooming is a routing problem in a multi-layer network architecture, since the objective is to find the “best” path to route traffic requests arriving dynamically at grooming nodes. In this case, equivalent $f_{sd}$ requests arriving in different times may be treated differently because of changed network conditions.

The static grooming problem has been studied intensively in the past, mainly on ring and mesh topologies, and it was proved to be NP-hard ([2] provides a review of many proposed heuristics for its solution). However, the rapid increase in Internet traffic and the intrinsic difficulty in forecasting its behavior on both small and large scales [3] make static grooming non-optimal. Even very fast network reconfigurations based on on-line measurements (say 30–60 min) are useless, because the measured traffic is not a good prediction of the future traffic.

In the emerging “IP over Optical” (IPO) network scenario, dynamic grooming becomes a fundamental issue for exploiting network resources. Since the traffic matrix is not known in advance, it is not possible to perform an optimal routing, but only heuristic algorithms can be used to maximize network performance. In order to route sub-wavelength-level IP traffic over a wavelength-routed optical network, the interaction between the two routing layers (IP and optical) must be considered, and the nature of the traffic (best-effort and elastic) loading the network must be taken into account through suitable models.

1.2. Networking context and related work

Before going into detail about recent works on dynamic traffic grooming, let us clarify the networking context in which these proposals can be used. RFC 3717 defines the framework for IPO architectures [4]. In IPO networks, IP routers are attached to an optical core network and connected to other peers via dynamically established lightpaths. The core network is composed of OXC, thus the optical layer is incapable of processing directly packets, bursts or any sub-wavelength capacity: it provides point-to-point connectivity between IP routers through fixed-bandwidth lightpaths. The set of all the lightpaths established over the physical topology is known as virtual topology.

The virtual topology is used by the IP routers to forward the traffic to the destination. The data plane is thus realized as an overlay network of lightpaths. Considering the control plane, different architectures can be envisioned according to the assumptions about the amount of information exchanged between the IP and optical layers. RFC 3717 states three interconnection models: overlay, augmented and peer. In the peer model, IP routers and OXC are considered to be peer network elements, thus the topology and other network information are completely shared by a unified control plane. In the overlay model, each layer performs its own routing functions, since no information is exchanged between them. An intermediate architecture between these two is the augmented model, where some aggregated information from one routing instance is
Performance and trade-offs of the different policies

The peer and the augmented models are appealing because sharing the knowledge base between the two layers allows the running of an integrated routing function using, for instance, an auxiliary graph, as done in [5]. The integrated management enables better usage of the overall network resources. However, both models seem unfeasible in the near term due to the tight integration between the two levels and scalability issues regarding the amount of exchanged information. The overlay model is instead technically feasible, since it only requires the definition of a clear interface between the IP and optical levels, as well as dynamic lightpath capabilities in the optical level, which are being experimented on in laboratories and research projects [6]. Surprisingly, even if the peer and augmented models do not seem realizable in the near future, most of the dynamic grooming algorithms proposed in the literature implicitly consider such models [5,7–10], since the objective is to maximize the utilization of the multi-layer network architecture and these models better support the purpose.

In the overlay model, the absence of information exchange leads inevitably to sub-optimal network utilization [10]. Only a few simple dynamic grooming algorithms based on this model have been proposed so far [11–13], and the interaction between different routing strategies adopted in the IP and optical levels were never assessed. In [11], constraint-based routing is assumed at the IP level, while [12,13] take into account also possible limitations in the number of ports. These papers explore the two extreme policies of always privileging the optical level exploitation, or the other way around. In [10], finally, the authors analyze both the augmented and the overlay architecture, assessing the superiority of the first one in terms of performance, and proposing new algorithms that improve performance in the overlay model with respect to [11] and in the augmented model with respect to a grooming algorithm in peer models proposed in [7]. It must be noted, however, that the algorithm proposed for the overlay model, named MLH_OVLY, requires the exploration of the optical connectivity between all the possible LSR pairs in the network to decide whether to accept or block a new connection, which hardly seems feasible in large networks, at least if no information from the optical level is available at the IP level, as in the overlay model.

It must be noted that none of the works on grooming in IPO networks adopted a realistic traffic model. The traffic loading the network is always composed of CBR (Constant Bit Rate) connections characterized by the bit rate and duration only. Any realistic evaluation of algorithms to be deployed within the Internet should instead capture at least some of the basic characteristics of Internet traffic. From the routing point of view, the most important features of Internet applications are the capacity to adapt the rate to changing network conditions (elasticity) and the need to transfer a given amount of data. The holding time of a flow becomes a consequence of the network conditions and not a property of the flow itself. In a previous contribution [14], we discussed in detail the impact of realistic traffic models on dynamic grooming, showing the inherent interaction between the IP and the optical layers and its effect on the overall performance of the network. In this paper, we use only the realistic model, which we called data-based in [14]. Traffic flows in this model share the resources on a virtual topology path following the max–min fairness criterion [15], thus mimicking the ideal behavior of a bundle of TCP connections.

In this paper, we study the impact of different grooming policies in IPO networks based on the overlay interconnection model with realistic traffic. The main contributions of this paper are the following:

- In Section 2, a formal definition of dynamic grooming in a general interconnection model is given and then specialized to the case of the overlay model.
- The existing dynamic grooming proposals for overlay networks are assessed as a subset of a more general and parametric grooming policy (Section 3).
- Performance and trade-offs of the different policies are discussed and explained in both regular (ring and mesh-torus) and irregular topologies, also discussing pros and cons of dynamic versus static grooming and the impact of adopting TE techniques in the IP or optical level.
- Unfairness issues inherent in dynamic grooming and arising from different physical distances between flow end-points are discussed, and hints on the problem solution are given.

2. Problem formulation

To the best of our knowledge, a formal description of dynamic grooming has never been defined in the literature. Here we present a formal framework based on graph theory for the definition of dynamic grooming policies.

2.1. A formalism for dynamic grooming

IPO networks are based on two layers: the optical layer and the data layer. The optical layer is based on OXCs interconnected by fiber links. G-OXCs,
which support sub-wavelength traffic multiplexing onto wavelength channels, are the bridge between the optical layer and the data layer. A G-OXC is also an IP router, hence transit traffic (not terminated in the router) can be groomed with incoming traffic. The data layer is composed of routers interconnected with a virtual topology made of all the lightpaths that have been set up in the optical layer.

In the physical and data layers respectively, we define the following:

- \( \mathcal{G} = (\mathcal{N}, \mathcal{E}) \) is the physical topology, where \( \mathcal{N} \) is the set of vertices \( v_i \) (OXC) and \( \mathcal{E} \) is the set of edges \( e_{ij} \) (fiber links) connecting vertices \( v_i \) and \( v_j \).

- \( \mathcal{G}^\nu = (\mathcal{N}^\nu, \mathcal{E}^\nu) \) is the virtual topology, where \( \mathcal{N}^\nu \) is the set of vertices \( v_i^\nu \) (IP routers), \( \mathcal{E}^\nu \subseteq \mathcal{N}^\nu \), and \( \mathcal{E}^\nu \) is the set of edges \( e_{ij}^\nu \) (virtual links) connecting vertices \( v_i^\nu \) and \( v_j^\nu \). Each edge in \( \mathcal{E}^\nu \) corresponds to a lightpath in the optical layer and, since there can be more than one lightpath between any two nodes, an additional identifier \( q \) is required for uniqueness.

The virtual topology varies in time as lightpaths are set up and torn down.

In the rest of the paper, the superscript \( \nu \) is used to specify vertices, edges and paths belonging to the virtual topology. When the superscript is absent, we refer to the physical topology. Notice that \( \mathcal{G}^\nu \) has nothing to do with the auxiliary graph defined in [5], which is an abstract representation of both levels, assuming complete sharing of the two control planes.

Each edge \( e_{ij} \) in \( \mathcal{G} \) is assigned a vector of properties \( \bar{w}_{ij} \) describing any static or dynamic (possibly vectorial) metrics pertaining to the physical or traffic-related characteristics of the link. Similarly, a property vector \( \bar{w}_{ij}^\nu \) is assigned to each edge of \( \mathcal{G}^\nu \).

In the optical layer, a path \( \pi_p(v_i, v_j) \) (or simply \( \pi_p \)) of length \( n \) is defined as a sequence of \( n \) distinct edges \( e_{ij} \) joining \( v_i \) and \( v_j \), where \( v_i, v_j \in \mathcal{N}, e_{ij} \in \mathcal{E} \), and \( \pi_p(v_i, v_j) = \{e_{ij}, e_{ik}, \ldots, e_{jk}\} \). The value of \( p \) is unique in \( \mathcal{G} \) and identifies the path explicitly. This identifier is required since more parallel paths may exist between the nodes \( v_i \) and \( v_j \). The path \( \pi_p \) is a lightpath, and corresponds to a specific wavelength if no wavelength conversion is considered.\(^1\)

Let \( \models \) be the operator that maps a lightpath in the physical topology onto an edge of the virtual topology: \( e_{lm}^\nu \models \pi_p \) if the path \( \pi_p \) joins the two vertices \( v_i^\nu, v_k^\nu \in \mathcal{N}^\nu \). In the data layer, a path \( \pi_i \models \pi_i^\nu(v_j, v_k^\nu) \) is a sequence of \( n \) distinct edges \( e_{ih}^{\nu_i, q} \in \mathcal{E}^\nu \), with \( t \) again a unique identifier to distinguish multiple parallel paths.

Fig. 1 presents an example to better understand the formalism used. The maximum number of wavelengths per link is \( W = 2 \). Let us suppose that the following five paths have been set up in the optical layer:

\[
\begin{align*}
\pi^{(1)}_1 & = \{e_{12}, e_{24}\}, \\
\pi^{(2)}_1 & = \{e_{23}, e_{34}\}, \\
\pi^{(2)}_2 & = \{e_{12}, e_{23}\}, \\
\pi^{(2)}_3 & = \{e_{24}, e_{34}\}, \\
\pi^{(2)}_4 & = \{e_{13}\}.
\end{align*}
\]

The superscript \((i)\) is used to specify the corresponding wavelength.\(^2\)

The corresponding set of edges in the data layer is:

\[
\mathcal{E}^\nu = \{e_{14,1}^\nu, e_{24,2}^\nu, e_{13,3}, e_{23,4}^\nu, e_{13,5}^\nu\}
\]

where:

\[
\begin{align*}
e_{14,1}^\nu & \models \pi^{(1)}_1, \\
e_{24,2}^\nu & \models \pi^{(2)}_2, \\
e_{13,3} & \models \pi^{(2)}_3, \\
e_{23,4}^\nu & \models \pi^{(2)}_4, \\
e_{13,5}^\nu & \models \pi^{(2)}_5.
\end{align*}
\]

The identifiers \( p \) and \( q \) are conceptually different, as they are defined in different logical planes but, as seen in the example, they can assume the same value for simplicity.

For each pair \( v_i, v_j \), a set \( \mathcal{P}^i_j \) is defined as the set of all paths existing between \( v_i \) and \( v_j \): \( \mathcal{P}^i_j = \{\pi_p(v_i, v_j) \mid v_i, v_j \in \mathcal{N}, i \neq j\} \). Similarly, a set \( \mathcal{P}^\nu_{i,j} \) is defined as the set of all paths existing between \( v_i^\nu \) and \( v_j^\nu \): \( \mathcal{P}^\nu_{i,j} = \{\pi_i(v_i^\nu, v_j^\nu) \mid v_i^\nu, v_j^\nu \in \mathcal{N}^\nu, i \neq j\} \).

Given some Routing and Wavelength Assignment (RWA) algorithm \( \Lambda \) in use in the optical layer, a cost \( c^\Lambda(\pi_p) \) is assigned to a path \( \pi_p \) by using a combination of the properties \( \bar{w}_{ij} \) of its links. The algorithm \( \Lambda \) selects the minimum cost path \( \pi_{ij} = \Lambda(\mathcal{P}^i_j) \) available, which we assume to be unique, possibly breaking ties with random choices. If no path is available, \( \Lambda \) returns \( \emptyset \), identifying the empty path.

Let us illustrate the path cost assignment and routing mechanism by using the adaptive routing FPLC

\(^1\) In the rest of the description, we will keep this assumption, but extension to the general case of wavelength conversion capability is trivial.

\(^2\) This notation is used here only to help the reader, and it will not be used any further in the following, since this information can be embedded in the path characteristics.
(Fixed-Paths Least-Congestion) presented in [16]. The cost \( c^{A}(\pi_{p}) \) is calculated as:

\[
c^{\text{FPLC}}(\pi_{p}) = \max_{e_{ij} \in \mathcal{E}} C(e_{ij}) - a(\pi_{p}) + \sum_{e_{ij} \in \pi_{p}} \frac{1}{|\mathcal{N}| + 1}
\]

where \( C(e_{ij}) \) is the number of optical channels (wavelengths) on the edge \( e_{ij} \) and \( a(\pi_{p}) \) is the counting function of the wavelengths that are concurrently available on each link of the whole path \( \pi_{p} \), i.e., the number of contiguous wavelengths on it. The number \( |\mathcal{N}| \) of vertices in \( \mathcal{G} \) provides a smaller-than-unit weight to break ties based on the number of hops.

In the virtual topology \( \mathcal{G}^{v} \), we can operate with the same approach. Given an IP-based routing algorithm \( \mathcal{O} \), the minimum cost path between any two nodes is selected as \( \bar{\pi}_{ij}^{v} \triangleq \mathcal{O}(\mathcal{P}_{ij}^{v}) \).

As an example, the cost function used for the Minimum Distance (MD) routing presented in [17] is:

\[
c^{\text{MD}}(\bar{\pi}_{ij}^{v}) = \sum_{e^{v}_{jk,q} \in \pi_{ij}^{v}} \frac{1}{B(e^{v}_{jk,q})}
\]

where \( B(e^{v}_{jk,q}) \) is the bandwidth that is available on link \( e^{v}_{jk,q} \) for an arriving data-flow request.

A final important set of paths is the set \( \mathcal{P}_{\text{pre}} \) of pre-established lightpaths in \( \mathcal{G} \). \( \mathcal{P}_{\text{pre}} \) contains a set of permanent lightpaths that guarantees the connectivity of the virtual topology regardless of the traffic pattern and lightpath establishment dynamics. The corresponding set of edges in the virtual topology is called \( \mathcal{E}_{\text{pre}}^{v} \). In a previous work [14], we observed that grooming policies that use the optical resources aggressively may, from time to time, and specially at low/medium loads, build logical topologies that are not completely connected — a situation that is extremely critical in operation.

Let us consider the simple example in Fig. 2, where \( W = 2 \). If the six lightpaths highlighted in the left-hand side of the figure are set up to allow traffic relations among those node pairs, the resulting logical topology is disconnected: two logical disjoint networks of three nodes (A, C, E and B, D, F) are defined as shown in the right-hand side of the figure, and no more virtual links can be established. This situation can be broken only if one of the existing lightpaths is torn down. Including a pre-defined spanning tree or any other topology that ensures connectivity in the data layer avoids this problem. The same phenomenon was also observed in [18]. The set \( \mathcal{E}_{\text{pre}}^{v} \), if present, is always taken into account in the data layer \( \pi_{ij}^{v} \) paths computation.

We can therefore define a generic grooming policy as a procedure \( G(A, \mathcal{O}, \Delta) \) where \( \Delta \) is a set of criteria defining the interaction between the optical and IP levels, determining the collaboration between \( A \) and \( \mathcal{O} \), and defining \( m \), the number of times that \( A \) is invoked to set up multiple lightpaths. For instance, \( \Delta \) defines multiple \( v_{ji}, v_{ki} \) pairs for the lightpath set up, where \( \{v_{ji}, v_{ki} \in \mathcal{V}, 1 \leq i \leq m\} \), and can be different from the nodes \( v_{s}, v_{d} \) corresponding to the source–destination pair \( v_{s}^{v}, v_{d}^{v} \) of the incoming flow \( f_{sd} \). The RWA algorithm \( \Lambda(\mathcal{P}_{\mathcal{G}}^{b_{ki}}) \) is performed \( m \) times (once for each \( i \) to open multiple lightpaths), while the route selection \( \mathcal{O}(\mathcal{P}_{\mathcal{G}^{v}}^{sd}) \) is performed only once, since a non-ambiguous routing algorithm always returns the same path for the same \( v_{s}^{v}, v_{d}^{v} \) couple once \( \mathcal{P}_{\mathcal{G}^{v}}^{sd} \) is fixed. If present, admission control criteria (at both the optical and IP levels) can be integrated in \( \Delta \), which is normally expanded as a set of if-then-else clauses.
The set of criteria $\Delta$ is influenced by the integration level of the IP and optical control planes (overlay, peer or augmented architectures). Indeed, before restricting research to the overlay case, some comments are due on the mapping of Eq. (3) to peer or augmented models.

In case of peer architecture, the data layer and optical layer control planes are merged together, which means, in the ideal case, that an auxiliary routing graph can be defined as in [5]. In such an extreme case, we do not need two separate routing functions $\Lambda$ and $\Omega$, since, given a single graph, we need a single routing function to operate on it. However, this is an extreme case. Most often, even if the two control planes are unified, it will not be possible to define an auxiliary graph\(^3\) where a single routing function can operate.

When the single auxiliary graph cannot be built in a peer architecture model, the set of criteria $\Delta$ operates concurrently on data-layer and optical-layer parameters, also allowing the use of simple optimization heuristics to define the best possible strategy for optical/virtual routing mix. Ideally, even ‘what-if’ analysis could be performed, running $\Omega$ on several resulting from different choices at the optical layer. However, we deem that this is far too complex to apply in dynamic situations, where new flows and paths must be set up on demand in a few hundreds of milliseconds.

In an augmented architecture, the data layer has some additional information about optical resources, but not the entire knowledge of the optical layer topology. For instance, the optical network can periodically signal to the IP level the cost of opening new lightpaths between routers. This information can be used in defining the criteria $\Delta$ to direct the optical layer properly in opening the ‘best’ lightpath to support the new flow based on some cost function.

As we discuss in the following, in the case of an overlay model, since the two layers do not exchange any information, the most reasonable assumption for $\Delta$ is that $m = 1$, $j = s$ and $k = d$, i.e., a single lightpath can be established, and only between the entry and exit node of the current flow. In the rest of the paper, we consider this assumption.

So far, we have discussed only the acceptance of new flows, but the grooming policy also defines how to release optical resources when lightpaths are unused. Releasing a lightpath between $v_i$ and $v_j$ means re-computing the set $\mathcal{P}_{G_v}^{ij}$ in the optical layer and consequently deleting an edge $e_{ij}^v$ from the virtual topology and re-computing all the sets $\mathcal{P}_{G_v}^{lm}$ that included $e_{ij}^v$.

In an overlay model, the optical layer shall release a lightpath as soon as it is empty, maybe with some time-out to avoid a too aggressive resource release. In augmented and peer architectures, there are clearly many more options. For instance, it is reasonable to suppose that, in an augmented architecture where the optical layer control plane signals lightpath costs to the data layer one, the data layer signals some information about its traffic, which allows for some form of optimization. $\Delta$ can also be delegated the task of defining which lightpaths to tear down and when.

2.2. Detailing $G$ for overlay architectures

In IPO networks based on the overlay architecture, the control planes of the optical and IP levels are separated. Each time an incoming request $f_{sd}$ needs to be routed, there are only two possible options: (i) route it over the current virtual topology $G_v$ invoking $\Omega$ or (ii) set up a new lightpath $e_{jk,q}^v \models \pi_{jk}$ invoking $\Lambda$ and route the request over the new virtual topology $G_v \cup e_{jk,q}^v$ (invoking $\Omega$ in a second phase).

Fig. 3 specifies the procedure (3) for an overlay IPO network, without detailing the criteria set $\Delta$. Notice

---

\(^3\) For instance, in the case of realistic networks with tens of nodes and tens (hundreds) of network elements per node, the auxiliary graph can be too complex to operate upon.
that policies privileging the use of already established lightpaths can always resort to invoke \( A \), either because no \( \hat{\nu}_{sd} \) was found or because the result of \( \Omega \) is refused for any reason.

3. Grooming policies

In the scenario depicted above, independently of the definition of \( A \) and \( \Omega \), the set of rules \( \Delta \) defines how and when to invoke \( A \). Although many criteria can be envisaged, based on any network measure available, the simplest rule is based on the number of IP hops between \( s \) and \( d \). In other words, \( \Delta \) defines as a rule the invocation of \( A \) only if there is no path \( \pi^\nu_{ij} \) available that does not exceed \( K \) hops from \( s \) to \( d \). We call this policy Hop Constrained Grooming, \( HC(\cdot) \). Additionally, \( HC \) can include rules for refusing a logical path \( \pi^\nu_{ij} \) based on any suitable measure. The formal definition of the policy is shown in Fig. 4.

The dynamic grooming policies studied in [12] and named “optical-layer-first” and “IP/MPLS-layer-first” are simply the extreme cases for \( K = 0 \) and \( K = \infty \) and have been studied in the simple case of circuit-like traffic. It is interesting to observe that these two policies, which may seem opposite to each other, are indeed just particular cases of a single policy. This fact highlights the importance of a proper formal framework for the description of dynamic grooming, as defined in this paper.

With a realistic elastic model of Internet traffic, the definition of congestion is not trivial, since it cannot refer directly to the amount of resources already allocated. Bandwidth overbooking is normal practice, and we assume that a new lightpath is opened only when accepting the flow \( f_{sd} \) on the virtual topology would result in assigning it a bandwidth smaller than some given amount. In order to fix ideas, let us assume that flow requests arrive at the network with three attributes: a peak transmission rate \( B_M \); a desired average bandwidth expressed as a fraction \( th_s \) of \( B_M \); and a minimum requested rate expressed as a fraction \( th_s \) of \( B_M \), with \( th_s \leq th_o \). The grooming policy tries to find a route that provides more than \( \tau_o = th_o \cdot B_M \) bandwidth for the flow. Besides, if at any time the bandwidth assigned to flow \( f_{sd} \) falls below \( th_s (f_{sd}) \cdot B_M (f_{sd}) \), then the flow will close and is counted as a “starved” flow, because the network was not able to guarantee its correct completion. \( B_M, th_o \) and \( th_s \) may be included in some SLA (Service Level Agreement) at the IP/optical interface (see [19] for initial works on optical-SLA). We thus define a starvation probability and not a blocking probability, since the adaptive and elastic nature of Internet traffic does not allow the easy definition of strict admission procedures, even assuming that flow requests are not single IP flows but large aggregations.

A network operator running the \( HC(\cdot) \) grooming procedure requires new lightpaths based on the two parameters \( K \) and \( \tau_o \) defined above, which can be seen as a threshold under which new optical resources are added to enhance the virtual topology.

The set of criteria \( \Delta \) for \( HC(\cdot) \) can be specified as \( \delta(K, \tau_o) \) defining the if-then-else rules of the generic grooming policy in Fig. 3. This leads to the grooming procedure described in Fig. 4, and compactly written as \( HC(A, \Omega, \delta(K, \tau_o), f_{sd}) \).

With reference to Fig. 4, \( H(\pi^\nu_{ij}) \) returns the number of hops in virtual topology path \( \pi^\nu_{ij} \) and

\[
b(\pi^\nu_{ij}) = \min_{e^{\nu}_{jk,q} \in \pi^\nu_{ij}} B(e^{\nu}_{jk,q})
\]

where \( B(e^{\nu}_{jk,q}) \) is the bandwidth that is available on link \( e^{\nu}_{jk,q} \) for the incoming flow.

4. Results and discussion

A theoretical performance analysis is not feasible due to the complexity of the system, thus we resort to simulations for the performance evaluation, exploring both regular and irregular topologies.

After describing the simulation tool that we use in Section 4.1, defining the relevant performance parameters in Section 4.2, and introducing the networking scenarios that we consider in Section 4.3, we devote the rest of the section to results. Results are organized in three sections, each one discussing...
a specific topic. Section 4.4 is devoted to the general behavior as a function of the topology and networking scenario, Section 4.5 discusses fairness issues, and Section 4.6 introduces the broad problem of the interaction of grooming algorithms with traffic engineering techniques used at either the data or optical layers, also discussing the use of routing algorithms that are not shortest-path based.

4.1. The simulation tool

The implementation of grooming policies in a packet-level simulator such as ns-2 is not convenient for efficiency reasons. Starting from existing tools in our research groups, we have developed a simulator that is capable of handling the layered topological structure of IPO networks as well as several different $\Lambda$ and $\Omega$ functions.

The description of the tool, named GANCLES, goes beyond the scope of this paper, and we refer the interested reader to the GANCLES web-site [20].

Presently, GANCLES includes different TE techniques at both the optical level and the IP level. The grooming policies that we propose in this paper are all implemented, and there is the possibility of choosing different $\mathcal{P}_{\text{pre}}$ sets as well as disabling this feature and studying the phenomenon of logical topology disconnection. The tool is downloadable and freely available to the community.

4.2. Performance indices

Among the possible performance indices, we have selected the following, which blend both user-perceived performance and network operation costs.

$P_s$ — Starvation probability. This is the probability that a flow closes during its life because it is not receiving service with acceptable quality. A flow $f_{sd}$ closes and drops the network if its instantaneous throughput falls below $th_s \cdot BM$.

$T$ — Average normalized throughput of completed flows.

$$T = \frac{\sum_{v_i^s \in \mathcal{N}^s} \sum_{v_i^d \in \mathcal{N}^d} \sum_{f_{sd} \in F_{sd}} \hat{T}(f_{sd})}{\sum_{v_i^s \in \mathcal{N}^s} \sum_{v_i^d \in \mathcal{N}^d} \sum_{f_{sd} \in F_{sd}} 1}.$$  

The average throughput of a flow is computed by dividing its transmitted data by its holding time. $\hat{T}(f_{sd})$ is the average throughput of flow $f_{sd}$ normalized with reference to its maximum requirement $BM$; $F_{sd}^c$ is the set of non-starved flows observed on the $s \rightarrow d$ traffic relation. Notice that, in a resource-sharing environment, $T$ is not the average resource occupation divided by the number of flows.

$R_c$ — Routing table change rate. Each time a new lightpath $\pi_p$ is established, some of the sets $\mathcal{P}_G^{\nu_i}$ must be re-computed. The rate of such changes is a good measure of the joint grooming and routing cost within the network.

$U_d$ — Distance unfairness index.

$$U_d = \frac{\max_{0<r<|\mathcal{N}|} T' - \min_{0<r<|\mathcal{N}|} T'}{T}$$

measures if the resource assignment is fair with respect to physical distance. $T'$ is the average throughput calculated for node-pairs with hop distance $r$ in the physical topology. $U_d$ ranges from zero to $\infty$; any value larger than 1 indicates strong unfairness, because some flows take at least twice as many resources as others. In general, a larger $U_d$ implies larger performance differences. This parameter is evaluated only for regular topologies and uniform traffic, since in other cases it can be influenced by factors external to the grooming policy.

The goal of a grooming algorithm is to maximize $T$ while minimizing $P_s$, $R_c$, and $U_d$.

4.3. Networking scenarios

We use three different topologies:

- $R8$ is an 8-node ring;
- $MT16$ is a 16-node mesh-torus network with a connectivity of four, i.e., each node is connected to the adjacent four in a regular, closed lattice mapped on the surface of a bi-dimensional torus (or a doughnut, in popular terms);
- $NSF$ is the well-known NSF-net topology with 14 nodes and 21 fiber links [21]. In this topology, we assume that only six OXCs have grooming capabilities, as shown in Fig. 5.

The number of wavelengths, $W$, is one of the parameters of major interest for investigating the generality and scalability of solutions with respect to the amount of available resources. Each wavelength has a capacity of $g = 20$ Gb/s in our study. A data-layer traffic source is connected to each G-OXC, generating requests with $BM = 10$ Gb/s following a Poisson process; each flow transfers data whose amount is randomly chosen from an exponential distribution with average 12.5 GB; $th_s$ is 0.1 in all simulations and $\tau_o = 3$ Gb/s. Dynamically opened lightpaths are immediately torn down if they are not used. All
simulations are run until performance indices reach a 99% confidence level over a ±1% confidence interval around the point estimate. Estimations are carried out using the batch means technique. Results are plotted versus the total load, \( L \), offered to the network.

Unless otherwise stated, \( A \) is the FPLC algorithm described in Section 2.1 with first-fit wavelength assignment and \( R \) is the standard fixed shortest-path (FSP) algorithm. A uniform traffic pattern is simulated, i.e., when a new flow request is generated, the source and destination are randomly chosen with the same probability.

As was mentioned in Section 2.1, the virtual topology connectivity can be guaranteed with pre-established lightpaths that are never closed. We consider two possibilities for populating \( P_{\text{pre}} \):

- A Minimum Spanning Tree (MST) of lightpaths, which connects all the G-OXCs by using the minimal amount of optical resources (for implementation issues, we refer the reader to [20]).
- A ‘Physical-Topology’ (PT) of lightpaths, which simply sets up the lowest-order wavelength among each pair of adjacent nodes. This approach is meaningful only if all the network nodes are G-OXCs, because lightpaths between non-grooming OXCs would be unused.

We applied the solution PT for the regular topologies \( MT16 \) and \( R8 \) and MST for \( NSF \), since MST in rings and mesh-tori is extremely inefficient.

### 4.4. Grooming policy behavior on different topologies

Our aim is to assess the behavior and performance of grooming policies with elastic traffic. As we discussed in Section 3, the two reference algorithms known in the literature, named “optical-layer-first” and “IP/MPLS-layer-first”, are particular cases of the \( HC \) parametric policy that we introduce in this paper. These two policies have been studied in [12] for the simple case of circuit-like traffic. Comparing ‘our’ proposal with other proposals thus means comparing \( HC \) policies with different parameters; namely, \( K = 0 \) corresponds to “optical-layer-first” and \( K = \infty \) to “IP/MPLS-layer-first”.

The first set of results shows how the grooming policy \( HC \) behaves on both regular and arbitrary topologies when varying \( K \) and \( W \). The impact of \( \tau \), studied in [22], thus it has not been considered in the following. When considering regular topologies with all G-OXCs and uniform traffic, the load can also be expressed as the relative traffic \( \rho \) offered to each network node normalized to the total capacity of its egress links. Given the total number of nodes \( |N| \) in the network, the connectivity degree \( D \), the number of wavelengths per fiber \( W \) and their data rate \( g \), we have \( \rho = \frac{L}{|N|DWg} \). \( \rho \) cannot be defined if the topology is not regular or if some nodes do not have grooming capabilities, but it offers a means of comparison between different topologies and different values of \( W \).

We first consider the impact of \( W \) on \( R8 \). Fig. 6 and Fig. 7 show the impact of using different values of \( K \) on \( T \) and \( P_s \) for \( W = 4 \) and \( W = 8 \), respectively. On the bottom \( x \)-axis we use \( L \) and on the top \( x \)-axis we use \( \rho \).

In both cases, \( K = 1 \) ensures the best performance. The performance spread increases with \( W \); results for \( W = 12 \) confirm this result, as shown in the following Fig. 11, but are not included here for space reasons. This behavior comes from the aggressive use of optical-layer resources with \( K = 0 \). Setting up lightpaths even when not needed, the optical layer becomes overcrowded with lightpaths, which leads to the blocking of lightpath set-up requests when congestion is impending, resulting in a poorly connected virtual topology, reduced throughput \( T \), and increased starvation \( P_s \). Increasing \( K \) above 1, the performance tends to be similar to \( K = \infty \). This is due to the small average distance between nodes, but it also confirms that the best way to use optical resources is trying to build a fully connected mesh in the virtual topology. Although difficult to formally prove and even to show, we have observed that, by using \( K = 1 \), the grooming algorithm tends to build a full-mesh virtual topology when \( W \) provides enough resources, independently from the physical topology. In our opinion, this is the main reason why setting \( K = 1 \) guarantees better performance.

Although our study concentrates on the behavior of the grooming policies when elastic IP traffic is considered, we also present some results using guaranteed CBR traffic for the sake of completeness. We use the \( R8 \) topology and assume the presence of a straightforward CAC function that refuses connections.
if the required throughput $B_M$ cannot be guaranteed. 

Comparing the blocking probability $P_b$ on the left plot of Fig. 8 with $P_s$ on the right plot of Fig. 6, the difference between the policies with different $K$ values is slightly larger for the CBR traffic than for the elastic traffic. This relation changes when more resources ($W = 8$) are available in the network, but the relative merit of the grooming policies remains the same in all cases. The right plot of Fig. 8 and the right plot of Fig. 7 show the case $W = 8$. The blocking probability that we
Fig. 9. Per-flow average throughput $T$ (left plot) and starvation probability $P_s$ (right plot) for MT16 with $W = 8$.

Fig. 10. Per-flow average throughput $T$ (left plot) and starvation probability $P_s$ (right plot) for NSF topology with $W = 4$.

obtain is intolerable, but we kept the same offered load for the sake of comparison. Additional results with CBR guaranteed traffic were presented in [22].

Fig. 9 shows $T$ and $P_s$ for a MT16 topology with $W = 8$. The behavior is similar to that observed in R8, confirming that the relative merit of grooming policies is not related to the topology.

The analysis of the behavior in different network topologies is completed with the NSF topology. Fig. 10 shows $T$ and $P_s$ when $W = 4$. In this case, $K = 0$ shows a higher throughput, but this is due to the higher $P_s$ and not to a better performance: the lower number of flows coexisting in the network increases the throughput for the surviving ones.

Summarizing this first set of results, we can claim that $K = 1$ ensures the best combination of $T$ and $P_s$ among the policies considered in the paper. It is not easy to define a single metric that combines $P_s$ and $T$ so that the “best” solution has a higher score, because $P_s$ and $T$ have different “perceptual” effects. Starved flows contribute marginally to $T$ variations, so including them in $T$ would not change the average value significantly. On the other hand, $P_s$ is considered to be unacceptable even when it is as low as $10^{-2}$. In other words, the best way to combine the two metrics is by ordering them: the best solution is the one that gives higher $T$, provided that it has a lower $P_s$. $K = 1$ is the best solution, since it yields the lowest $P_s$ (over two orders of magnitude smaller than the case for $K = 0$) and its throughput performance is only 10% smaller.

Restricting the analysis to $K = 0,1$ and regular topologies, we analyze the effect of $W$ on performance with the normalized load $\rho$. Fig. 11 compares the behavior on R8 with $W = 4, 8, 12$. The figure clearly shows that, increasing $W$, the choice of $K = 1$ is indeed the best one, keeping the efficiency basically constant as the amount of resources and the traffic increases, while for $K = 0$ the performance decreases drastically. As explained at the beginning of the current section, for
Fig. 11. Per-flow average throughput $T$ (left plot) and starvation probability $P_s$ (right plot) for $R8$ with varying $W$ and $K = 0, 1$.

Fig. 12. Per-flow average throughput $T$ (left plot) and starvation probability $P_s$ (right plot) for $MT16$ with varying $W$ and $K = 0, 1$.

$K = 0$ the aggressive use of resources in the optical layer leads to a poorly connected virtual topology.

Fig. 12 presents a similar set of results for $MT16$ and $W = 2, 4, 8$. These results confirm the conclusions above.4 Most interesting is that, for low $W$, the performances are similar, and in both topologies they tend to diverge as $W$ increase.

So far, we have considered user-level performance. An important cost factor for evaluating grooming policies is how frequently the routing tables of the virtual topology need to be updated due to changes induced by opened and closed lightpaths. Fig. 13 shows the impact of $K$ on $R_c$, for both $R8$ and $NSF$ (similar results have been obtained on $MT16$). In both cases, an aggressive use of the optical resources ($K = 0$) leads to a very costly change rate, while instead more “conservative” grooming policies ($K > 0$) allow lower rates, thus guaranteeing a much safer stability in the data layer. It is interesting to notice that, after rising quickly with the load, $R_c$ converges to very low values, independently from the topology or from $K$, as the load increases and the virtual topology becomes stable (lightpaths are closed with very low probability). Allowing a lightpath to be idle for some time-out period, while waiting for possible new flows before closing it, would obviously reduce this cost, but also prevent the use of the wavelengths for other traffic.

4.5. Fairness issues with elastic traffic

One of the major drawbacks of dynamic network management, as it is well known from Internet experience, is the unfair behavior of the network...
towards longer connections (both in terms of physical distance and in terms of logical hops). In [23], the problem is addressed in the context of IPO networks (for the first time, to the best of our knowledge) with a circuit-like traffic model. In general, resource sharing exacerbates the unfairness, because longer connections compete with more connections with respect to short ones.

One of the interesting questions is whether the physical topology and available number of wavelengths have an impact on the problem. Fig. 14 reports the results for $W = 4$ (left plot) and $W = 8$ (right plot) for $R8$ and $MT16$. The hop constraint $K$ is set to $0, 1, \infty$.

On the one hand, it is clear that none of the parameter setting can guarantee perfect fairness, and that the physical topology does not have a major impact. On the other hand, increasing $W$ does help to keep a certain degree of fairness, especially if the grooming policy tends to build a logical topology that approaches a full mesh ($K = 1$). Instead, the very aggressive use of optical resources operated with $K = 0$ gobbles resources to build unnecessary parallel paths between adjacent node pairs, exacerbating the unfairness at high loads.

In conclusion, it seems that to achieve fairness without per-flow state in the network, a grooming policy should try to build a logical topology that is as close as possible to a full mesh with a resource allocation that is matched to the load offered by the node pairs. This issue, however, needs further research.

4.6. Single-layer TE and comparison with static resource assignment

So far, we have limited the discussion to the performance comparison of dynamic grooming policies,
keeping the routing algorithms fixed and as simple as possible. However, two major questions remain open: how dynamic grooming compares with optimal, static resource assignment, and how a dynamic grooming policy interacts with different routing algorithms in either the optical layer or the IP layer. This second question is extremely important for assessing the coexistence of dynamic grooming with non-shortest path data-layer routing.

For the sake of simplicity, and also because of the complexity of optimal static resource assignment in general topologies, we limit the study to the R8 topology with $W = 8$. For such a network, it is possible to build up a full mesh in the virtual topology using the method presented in [24]. In the case of uniform traffic, this is an optimal solution of static resource assignment.

The static grooming acts as a reference. Here we consider three possible solutions, all based on HC with $K = 1$ and $\tau_o = 3$ Gb/s.

(A) Both $\Lambda$ and $\Omega$ are FSP, and in $\Lambda$ the wavelength assignment is first fit; this combination is named “NO TE”, since the routing algorithm is the simplest possible at both levels.

(B) $\Lambda$ is FPLC with first fit and $\Omega$ is FSP; this is the combination that we have used in previous results and is named TE($\Lambda$).

(C) $\Lambda$ is FSP with first fit and $\Omega$ is the MD algorithm described in Section 2.1; this solution is named TE($\Omega$).\(^5\)

\(^5\) We deem that the use of TE techniques at both layers without exchanging information, thus adopting an augmented or peer model, will lead to “conflicts” in the use of resources, creating instabilities in the network; however, this issue requires additional investigation.

Fig. 15 reports $T$ and $P_s$ for the four grooming solutions described above. In order to enhance the differences and make comparisons easier, $T$ in the left plot is normalized with respect to the static grooming case, so that this solution throughput is constant and equal to 1. As expected, the static, optimal assignment performs generally better than the others, although TE($\Omega$) achieves higher throughput with lower $P_s$ at light-to-medium loads, but shows a sharp degradation as the load increases. This behavior was first observed in [25] in the context of static routing with alternate paths. In [26], the same behavior was analyzed in the context of dynamic, IP-based routing, so that it is not surprising to find it also in the context of dynamic grooming in IPO networks.

In order to complete the picture, we report results for the case of time-varying traffic. We expect that the HC grooming will adapt dynamically to the traffic change and improve performance above what is achievable with a static solution unable to adapt the resource assignment. In the scenario that we have just analyzed, we modified two nodes with distance 4 on R8, so that they periodically increase the volume of traffic that they generate, mimicking the behavior of some ISP node subject to “flash crowd” phenomena. The traffic increase from low to high is five times. The on and off periods are distributed exponentially, with an average of one hour. The behavior of the two servers is independent. Fig. 16 reports $T$ and $P_s$ for the two servers in this scenario. As expected, the relative merits of static and dynamic grooming are reversed, with TE($\Omega$) remaining the best choice among dynamic grooming. It must be noted that, even at high loads, where $T_{rel}$ decreases below one, $P_s$ remains lower than the reference static grooming.
5. Conclusions

This paper has introduced a formal description of dynamic grooming policies, clearly defining the limits between grooming in overlay architectures and grooming in peer or augmented architectures, where there is total or partial integration of the optical and IP control planes.

A family of grooming policies, based on constraints on the number of hops and bandwidth available at the virtual topology level, has been defined and analyzed in different regular and irregular topologies, discussing parameter setting, and the impact of the number of available wavelengths per fiber on the grooming policy. It has been shown that dynamic grooming policies previously presented in the literature are particular cases of the family that we defined, and that it is possible to define grooming parameters that lead to good performance regardless of the topology and that scale well with the amount of optical resources. Fairness issues in dynamic grooming were discussed, proposing guidelines for solving them.

The impact of traffic engineering techniques applied at the optical or IP level were also discussed. Results highlight that, using constraint-based routing techniques in the IP level, which is characterized by quicker dynamics, ensures better performance. The well-known problem of performance degradation at high loads when constraint-based routing is used is also present in IPO networks.

This paper, with the formal definition of grooming policies and the simulation tool that it offers to the community for studying grooming in generic IPO networks, opens up the possibility of planning and provisioning IPO networks based on the overlay architecture, as well as studying other architectural solutions.

References


