Evaluation of Bandwidth-dependent TE metrics in a GMPLS Path Computation System

Gino Carrozzo

META Centre
Consorzio Pisa Ricerche
C.so Italia 116, 56125 Pisa, ITALY
Telephone: (+39) 050 915 811
Fax: (+39) 050 915 823
Email: g.carrozzo@cpr.it

Stefano Giordano, Giodi Giorgi, Franco Russo

Dept. of Information Engineering
University of Pisa
via Caruso, 56122 Pisa, ITALY
Telephone: (+39) 050 2217 511
Fax: (+39) 050 2217 522
Email: {s.giordano,g.giorgi,f.russo}@iet.unipi.it

***Abstract*— The incoming GMPLS standardization is paving the way for the implementation of new configurable traffic engineering (TE) policies for transport networks. This paper takes aim at evaluating the effects of using bandwidth-dependent TE metrics in a centralized Path Computation System (PCS), suited for handling the routing requests in a transport network with a GMPLS control plane. The results of an intensive testing campaign show an evident improvement in the utilization of network resources when such TE metrics are enabled, whatever survivability requirement is imposed on the LSP (e.g. classical 1+1 protection, pre-planned or On-the-Fly restoration, etc.). Moreover, a simple policy function is suggested as a good trade-off between the achievable performance and the computing load on CPU.**

I. Introduction

The international standardization committees (e.g. ITU-T, OIF and IETF) are all converging in the design of an integrated network with a common Generalized Multi-Protocol Label Switching (GMPLS) control plane. GMPLS will manage all the network data planes [1], providing the required automation in the computation, the setup and the recovery of circuits for next-generation Optical Transport Network (OTN). GMPLS is an extension to devices capable of performing switching in time, wavelength and space domains of the MPLS control plane architecture. The core GMPLS architecture is based on a set of extensions to protocols for routing (e.g. OSPF and IS-IS) and signalling (e.g. RSVP), just available in IP networks. In the GMPLS context, other signalling protocols have been proposed (e.g. LDP, CR-LDP). Moreover, a new link management protocol (i.e. LMP) has been designed from scratch in order to handle correctly the distinction between the data plane and the control plane; in fact, they might not share the same link connection as in the IP networks (e.g. SONET/SDH networks, DWDM networks, etc.).

From a routing perspective, the GMPLS extensions provide new information for circuit computation and they enable configurable traffic engineering (TE) policies and new recovery strategies.

In such a context, this paper takes aim at evaluating the effects of using different bandwidth-dependent TE metrics in a centralized Path Computation System (PCS) suited for transport networks with a GMPLS control plane. In details, Sec. II is focused on the requirements for the PCS in a centralized GMPLS network scenario. In Sec. III the implemented bandwidth-dependent metrics are defined, while some results of an intensive testing campaign are shown in Sec. IV, deriving conclusions in Sec. V.

II. Routing requirements for a GMPLS PCS

The upcoming standardization for GMPLS architecture is focused on protocols objects and mechanisms, while only high level requirements are proposed for traffic engineering (TE) and survivability.

Traffic engineering is fundamental for load balancing in the transport network, in order to avoid the overloading (up to saturation) of some network resources and the sub-utilization of others. Path computation systems for standard IP networks are generally based on distributed, fast and simple routing algorithms (e.g. of the Shortest Path First class -SPF-[2]), integrated into the routing protocol module. This algorithms walk along a graph derived from the real network. The graph contains only the routing-capable nodes (a.k.a. vertices) and the links between them (a.k.a. edges) with an appropriate link metric.

In the GMPLS context, a link connecting two ports of neighbouring nodes may consist of more than one consecutive physical resource (e.g. fibres), possibly crossing routingin-capable devices (e.g. regenerators, optical amplifiers, optical mux/demux, etc.). For this reason a set of properties is assigned to each link for routing purposes (e.g. TE metric, available/used bandwidths, resource colours, SRLG list, inherent protections, etc.), transforming the traditional links in traffic engineering links (TE-links). A GMPLS path calculator is expected to return Label Switched Paths (LSPs), i.e. sequences of nodes, TE-links and labels, which try to match some constraints derived from the TE information above. Once an LSP is computed, it describes univocally a unidirectional or bi-directional connection (electrical and/or optical) between a source and a destination node. The standard SPF algorithms are not suited for such a computation, as they cumulate only the link metric along the graph. A modified SPF algorithm is

needed, called Constraint-based Shortest Path First (CSPF), as routes should be the shortest among those which satisfy the required set of constraints [3].

In the GMPLS architecture no specification is available for the implementation of a PCS module, as this issue is considered implementation-dependent. Moreover, no preference can be derived by the standards on the choice of a centralized or a distributed implementation, in spite of theintrinsic distributed approach of the GMPLS control plane.

The architectural choice we propose in this work is for the implementation of the PCS module inside a centralized network manager (NM). This solution promises to be the most effective for a full and flexible handling of traffic engineering and survivability into the network, particularly when these requirements need to be extended to a multi-area (or multi-domain) scenario.

In our implementation the PCS acts as a path computation server for the GMPLS network, receiving from the GMPLS Network Elements (NE) the topology information and the requests for computation across a single- or multi-area/AS. The LSPs computed by PCS (if any) are communicated to the ingress GMPLS NE, triggering a standard GMPLS signalling session (e.g. via G.RSVP-TE). The communications between the NM and the NEs are carried out by means of COPS protocol with proper extensions [4]. Focusing on LSP requests with survivability requirements [5], we identified four Classes of Recovery (CoR) for the LSP requests (e.g. Gold, Silver, Bronze, Unprotected), respectively related to the request for LSPs with path protection (e.g. SDH/SONET 1+1), with Fast Restoration, with On-the-fly restoration or none of these. In our PCS different algorithms are used for the different CoRs, ranging from the optimal implementation (e.g. in case of Gold CoR) to the sub-optimal ones (e.g. in case of Silver CoR). In details, for the Gold CoR we focused on the algorithms by R. Bhandari [7] because it promised lower theoretical complexity w.r.t. other algorithms (e.g. the famous Suurballe's one), when only K = 2 disjoint shortest paths are computed. The Bhandari's algorithm implemented in this work is the optimal counterpart to the sub-optimality of the algorithm used for the Silver CoR, i.e. the Two Step Approach (TSA) algorithm. TSA is based on a Dijkstra SPF and on a simple temporary network transformation for avoiding links/nodes of the worker path. The main advantages of such an algorithm are in the easiness of implementation and in the limited complexity both of the SPF algorithm (e.g. the Dijkstra complexity in our implementation) and of the network transformation. Further details on this issue are in [4] and [6].

III. Bandwidth-dependent TE metrics

The routing engine inside our PCS module is based on an implementation of the Dijkstra SPF algorithm. In order to handle Constrained SPF computations, we added a TEconstraints validation step to the well-known Dijkstra flow [2]. One of these constraint validations is the check on the bandwidth availability of the candidate link. Moreover, in order to let the algorithm converge towards an optimal SPF solution (e.g. between those which satisfies the required TE-constraints), the metric we choose to minimize during computation is bandwidth-dependent, according to the following equation [8]:

$$total\_cost = metric_{std} + metric_{TE} + policy_x(BW_{alloc}) \quad (1)$$

where $metric_{std}$ is the standard OSPF link metric, $metric_{TE}$ is the GMPLS metric for TE purposes, BW_{alloc} is the allocated bandwidth on the TE-link and $policy_x$ (with $1 \leq x \leq 2$) is a proper traffic engineering function designed to balance the traffic load (e.g. bandwidth consumption) in the topology.

The policy functions we define for this work are detailed in Eq. 4 and Eq. 5, where BW_{free} is the free bandwidth, while BW_{thr} is a threshold with respect to the total bandwidth (BW_{tot}) of the TE-link. The constants K and t are respectively a resource cost per unit and a smoothing factor of the overall TE function.

$$f_1(x) = K \cdot \left(1 + \left(\frac{x}{BW_{tot}}\right)^2\right) \quad (2)$$

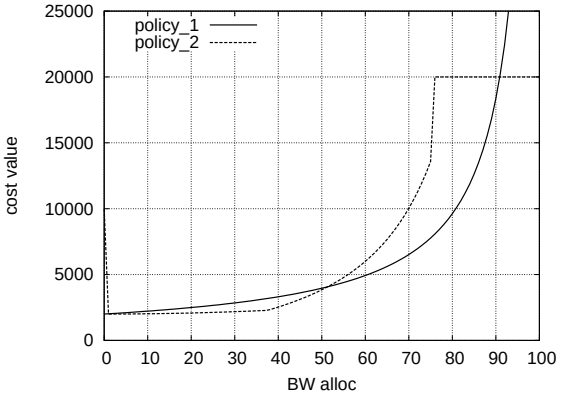
$$f_2(x) = K \cdot \left(1 - \frac{x}{1 + BW_{tot}}\right)^{-t} \quad (3)$$

$$policy_1 = K \cdot \frac{1 + BW_{tot}}{1 + BW_{free}} \quad (4)$$

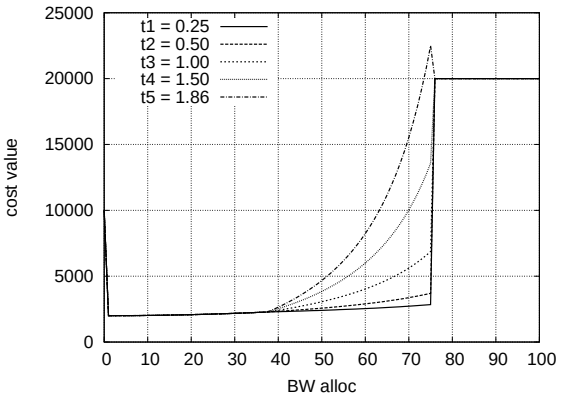
$$policy_2 = \begin{cases} 5 \cdot K & \text{if } BW_{alloc} = 0 \\ f_1(BW_{alloc}) & \text{if } 1 \leq BW_{alloc} \leq \frac{BW_{thr}}{2} \\ f_2(BW_{alloc}) - f_2(\frac{BW_{thr}}{2}) + f_1(\frac{BW_{thr}}{2}) & \text{if } \frac{BW_{thr}}{2} \leq BW_{alloc} \leq BW_{thr} \\ 10 \cdot K & \text{if } BW_{alloc} \geq BW_{thr} \end{cases} \quad (5)$$

In Figure 1-a the two policy functions are plotted w.r.t. the allocated bandwidth, with $K = 2$ for both functions, and $t = 1.5$ and $BW_{thr} = 75\%$ of BW_{tot} for the $policy_2$.

These policy functions increase the total cost of the link as allocated bandwidth increases, so as to avoid the overloading of the link and the resulting network congestion. The $policy_1$ function has been designed in order to discourage the link picking as the allocated bandwidth increases on it. On the contrary, the $policy_2$ function has been conceived in order to achieve a better flexibility in increasing the cost of the TElink. Indeed, $policy_2$ increases the link cost less than $policy_1$ till the reference bandwidth of $BW_{thr}/2$, encouraging LSP to pick resources on it. When the $BW_{thr}/2$ is reached, a "tunable" trend is configurable (ref. Figure 1-b) depending on the value for the smoothing factor, in order to encourage or discourage the link picking much more w.r.t. $policy_1$. An extra cost is assigned to totally free TE-links (e.g. $5 \cdot K$) in order to discourage their utilization in presence of just used links: this allows to fill up the most of the TE-links in a balanced way,



(a)



(b)

Fig. 1. Policy functions for traffic engineering (a) and effect of the smoothing factor t in $policy_2$ (b).

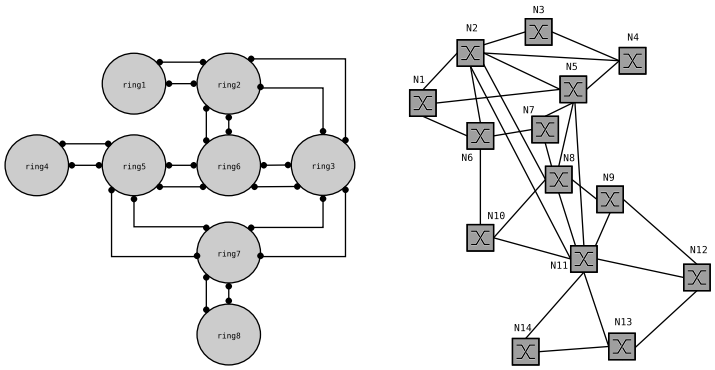
bounding the number of used links and the need for their installation (i.e. a kind of feedback to the network planning).

IV. Performance studies

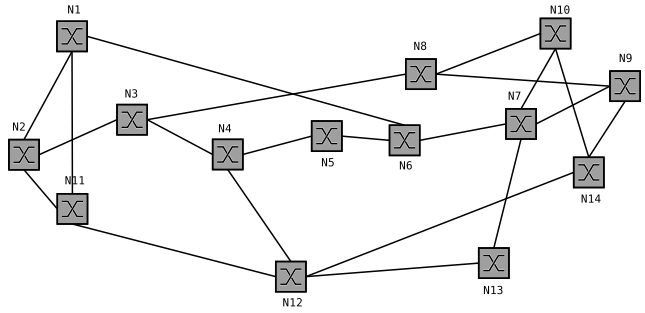
This section is aimed at highlighting the performances achieved when processing LSPs requests in case of the different bandwidth-dependent TE metrics proposed in Sec. III. The computational environment for all the tests (NM) is based on an Intel Celeron 500MHz PC with Linux Slackware 8.1 OS.

Measures have been collected on different topologies with increasing meshing degrees:

- an interconnected rings topology, with 64 nodes in 8 rings and meshing degree 2.56 (ref. Figure 2-a);
- a NSFNET topology, with 14 nodes and meshing degree 3.00 (ref. Figure 2-c);
- a simple Manhattan topology with 49 nodes and meshing degree 3.43 (ref. Figure 3-a);
- an Italian topology, with 14 nodes and meshing degree 4.00 (ref Figure 2-b);
- a half-meshed Manhattan topologies with 25 nodes and meshing degree 4.48 (ref. Figure 3-b);
- three meshed Manhattan topologies with 16, 25, 36 nodes and meshing degrees 5.25, 5.76, 6.11, respectively (ref. Figure 3-c).

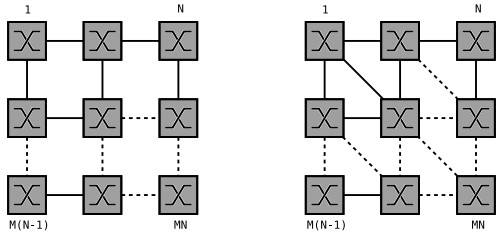


(a) (b)

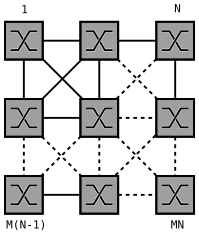


(c)

Fig. 2. (a) interconnected rings topology; (b) Italian topology; (c) NSFNET topology.



(a) (b)



(c)

Fig. 3. Manhattan topologies: (a) simple; (b) half-meshed; (c)meshed.

All these topologies have been modelled with generic nodes, configurable as SDH 4/4 Cross-Connects. Nodes have been assumed to be fully connectable, i.e. any of their ingress port may be cross-connected to an egress one. Adjacent nodes have been connected by a TE-link with random values for its TE-information (e.g. TE metric, available/used bandwidths, resource colours, SRLG list, etc.). All the TE-links have been assumed to be bi-directional and each configured with 4 STM-

64 ports VC4-multiplexed. A large number of LSP computations (req_path) have been requested on each topology (e.g. up to a connection request from each node towards all the others), trying to establish a stress condition for the algorithm operations. All the requests have been configured for bi-directional LSPs, with four possible values for the bandwidth:

- VC4 (e.g. 95.5% of req_path at 139 Mbit/s ca.);
- VC4-4c (e.g. 3.0% of req_path at 556 Mbit/s ca.);
- VC4-16c (e.g. 1.0% of req_path at 2224 Mbit/s ca.);
- VC4-64c (e.g. 0.5% of req_path at 8896 Mbit/s ca.).

For each topology have been observed: the number of computed LSPs ($comp\_path$); the mean TE-link usage in ($link\_utilization$); the number of totally disjoint pairs of LSPs (disjoint), in case of LSP requests with recovery requirements.

The results in Figure 4 and in Table I show the higher link utilization achieved when $policy_1$ or $policy_2$ are used. The trends in Figure 4-b/c show that greater improvements are obtainable for lower meshing degree: this is a common case for currently operative networks.

Moreover, computations with TE policies achieve higher scores because of the fair utilization of the network resources. These advantages are much evident in case of Silver CoR, because of the worst resource picking due to the suboptimality of algorithm adopted. In order to summarize the performances achieved in case of Gold and Silver CoRs, a Global Performance Factor (GPF) has been defined as:

$$GPF = \frac{disjoint}{req\_path} \cdot \frac{comp\_path}{req\_path} \tag{6}$$

in which the first term is related to the algorithm's effectiveness in creating maximally disjoint paths, while the latter represents the algorithm's effectiveness in computing valid paths, according to the resource availability on TE links. In Figure 5 the GPF is drawn at the different meshing degrees, showing a mean higher performance in case of utilization of the TE policies, both for Gold and for Silver CoRs (ref. Table 2 for numerical details).

TABLE I

GAIN IN LINK UTILIZATION ADOPTING THE NEW TE POLICIES.

	Bronze CoR	*Silver CoR*	*Gold CoR*
$policy_1$ vs. no_policy	+2.66%	+0.76%	+1.75%
$policy_2$ vs. no_policy	+2.75%	+0.80%	+2.01%

V. CONCLUDING REMARKS

The results shown above highlight how an evident improvement in the utilization of the network resources is achievable when applying TE policies in path computation, whatever CoR is required. Moreover, the advantages of defining complex policy functions (e.g. Eq. 5) do not pay for the introduced CPU complexity (e.g. $policy_2$ improves $policy_1$ performance for less than 0.25% in case of link utilization, and it worsens $policy_1$ performance in case of GPF). Therefore, a simple policy function such as the one described in Eq. 4 is suitable

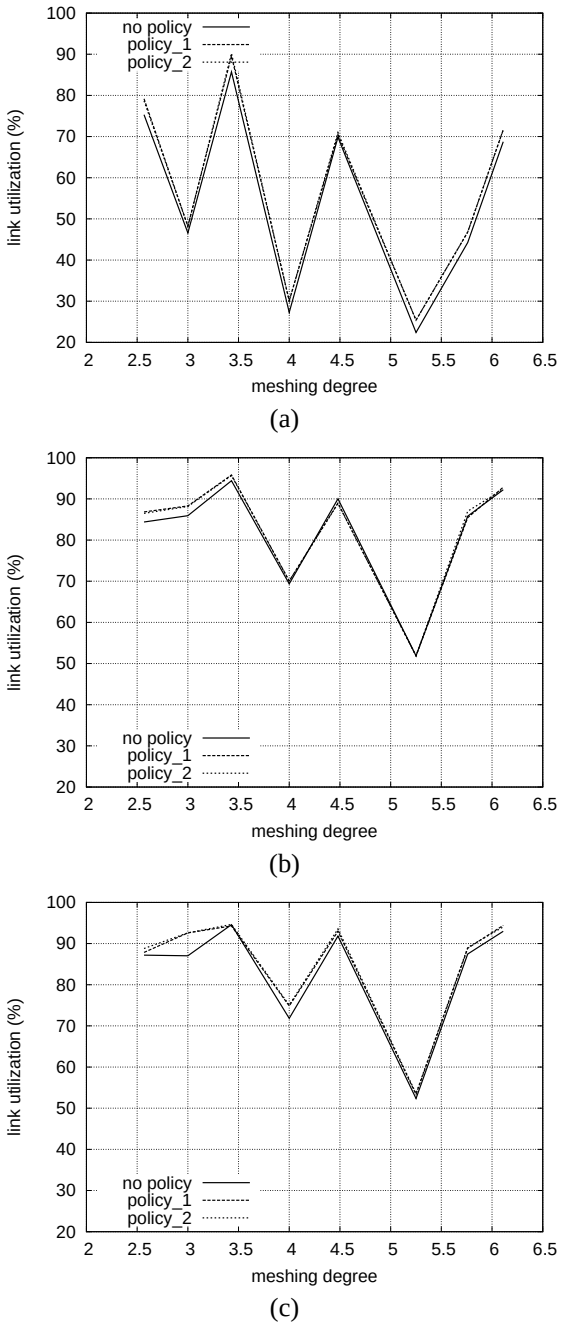
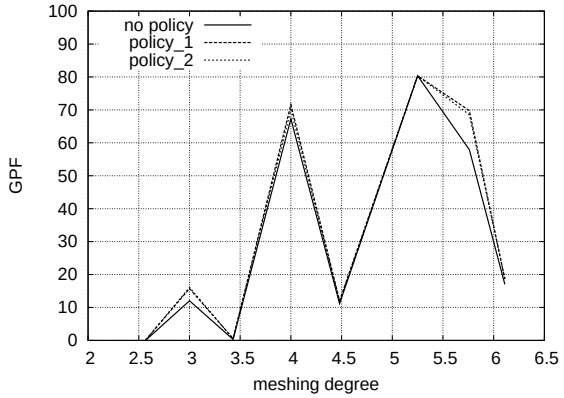


Fig. 4. Link utilization: (a) Bronze CoR - i.e. On-The-Fly LSP; (b) Silver CoR - i.e. Fast Rest. LSPs; (c) Gold CoR - i.e. 1+1 Prot. LSP.

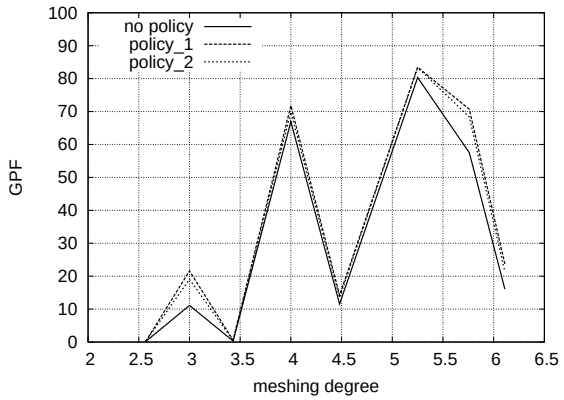
TABLE II

GPF GAIN ADOPTING THE NEW TE POLICIES.

	Silver CoR	*Gold CoR*
$policy_1$ vs. no_policy	+2.86%	+5.28%
$policy_2$ vs. no_policy	+2.49%	+4.01%



(a)



(b)

Fig. 5. GPF at different TE policies (a) for Silver CoR; (b) for Gold CoR.

and effective for a good tradeoff between the achievable performance (i.e. in terms of load balancing and blocking probability of the computation) and the computational load on CPU.

ACKNOWLEDGEMENTS

This work has been supported in part by the Italian Ministry of Education, University and Research through the project TANGO (MIUR Protocol No. RBNE01BNL5).

REFERENCES

[1] E. Mannie (Editor) et al., *Generalized Multi-Protocol Label Switching (GMPLS) Architecture*, draft-ietf-ccamp-gmpls-architecture-07.txt, Internet Draft, Work in progress, May 2003.

[2] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, *Network Flows - Theory, Algorithms and Applications*, Prentice Hall, 1993.

[3] J. Strand et al., *Issues for Routing in the Optical Layer*, IEEE Communications Magazine, pages 81-87, Feb. 2001.

[4] *Traffic models and Algorithms for Next Generation IP networks Optimization (TANGO) Project*, http://tango.isti.cnr.it/.

[5] P. Lang (Editor) et al., *Generalized MPLS Recovery Functional Specification*, draft-ietf-ccamp-gmpls-recovery-functional-00.txt, Internet Draft, Work in progress, Jan. 2003.

[6] G. Carrozzo et al., *A Pre-planned Local Repair Restoration Strategy for Failure Handling in Optical Transport Networks*", Photonic Network Communication, vol. 4 (no. 3-4), pages 345-355, Jul./Dec. 2002.

[7] R. Bhandari, *Survivable Networks - Algorithms for Diverse Routing*, Kluwer Academic Publisher, 1999

[8] B. Awerbuch et al., *Throughput-Competitive On-Line Routing*, IEEE Symposium on Foundations of Computer Science(FOCS), 1993, pp. 32-40.