Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 19 (2013) 1143 – 1150
Procedia Computer Science

International Workshop on Survivable and Robust Optical Networks

# Scalable Approach to Dynamic SLA Negotiation Mechanism in Protected Shared Mesh Optical Networks

Alireza Nafarieh[a*], Shyamala Sivakumar[b], William Robertson[a], William Phillips[a]

*[a]Dalhousie University, Halifax, NS, Canada*
*[b]Saint Mary's University, Halifax, NS, Canada*

**Abstract**

This paper investigates the scalability issues concerned with the dynamic SLA negotiation mechanism proposed in the previous study. The previous work presented a dynamic service level agreement negotiation mechanism considering intra- and inter-domain communications over shared mesh optical networks which may cause heavy control overheads in a dynamic environment such as the control plane of GMPLS-based networks. In this paper, two main issues regarding the dynamic propagation of information will be analyzed: i) control overhead, and ii) propagation delay. The paper employs some alternative means of communication to reduce the overheads and resolve the possible scalability issues.


Selection and peer-review under responsibility of Elhadi M. Shakshuki

*Keywords: Scalability issues; Dynamic service level agreement negotiation; Path state advertisement; OSPF-TE extensions; BGP-TE extensions*

## 1. Introduction

In some cases, the requested parameters in SLA are beyond the capacity of the network, and the connection is easily rejected or blocked. To give the customer a chance to choose another provider, or in case of having only one provider, to comply with the provider's network capacity as much as possible, an automatic, bidirectional, and dynamic mechanism for SLA parameters negotiation between service providers and customers is required. This mechanism helps service providers to control the network resource assignment in WDM networks.

*Alireza Nafarieh. Tel.: 1-902-494-2057
*E-mail address*: ali.nafarieh@dal.ca.


Selection and peer-review under responsibility of Elhadi M. Shakshuki
doi:10.1016/j.procs.2013.06.163

This has been the key motivation in [1] for proposing a dynamic mechanism to negotiate specific SLA parameters or any other metrics which may be desirable. However, the negotiation mechanism proposed in [1] may cause heavy control overheads in a dynamic environment such as the control plane of GMPLS-based networks. Employing some alternative techniques of negotiation to reduce the overheads and resolve the possible scalability issues is a wise step.

Routing within optical networks relies on knowledge of network topology and resource availability. The first step towards network-wide link state determination is the discovery, by each route, of the status of local links to all neighbours. To disseminate TE information among entire nodes of a network, the information should be propagated inside and outside the autonomous system (AS), along the path from source to destination. For intra/inter-domain TE-information dissemination, OSPF-TE opaque LSAs with newly proposed extensions have been introduced in [1]. Link attribute information should be communicated using dynamic SLA negotiation mechanisms over a multi-homed network topology. The customer side of the network is exposed to SLA information from all the service providers to which it is connected. The customer has the choice to pick the service provider that is the most suitable for satisfying the requested connection. This paper presents a scalable solution to the dynamic SLA negotiation mechanism presented in [1].

The paper is organized as follows: Section 2 identifies the motivations, objectives, and related work around dynamic SLA negotiation and propagation mechanisms. The scalable approach to traffic engineering (TE) extensions is introduced in Section 3. Scalability and overhead issues are analysed and addressed in Section 4 including network performance evaluation. Section 5 summarizes the paper discussions.

## 2. Related Work

As discussed in detail in previous work, [1], the intra-domain negotiation mechanism can propagate the link attribute as SLA parameters while an inter-domain mechanism can advertise the proposed SLA-based traffic engineering path constraints. [1] has also shown how SLA negotiation protocols together with the proposed traffic engineering attributes can improve the performance of priority-aware algorithms. The scalability issues should be considered as another implication of the new TE path attributes dissemination. Although the use of the TE attribute does not increase the number of routes, it may increase the number of update messages required to distribute the routes. In previous work [1], new sub-TLVs to BGP-TE extensions have been proposed in order to support inter-domain TE path attribute propagation while the path attributes have no impact on the BGP path calculation procedure.

Since OSPF and BGP are widely used as intra-domain and inter-domain routing protocols of networks respectively, the majority of the studies modify or add some extensions to these protocols. The authors in [2] have described extensions to the OSPF protocol version 2 to support intra-area TE, using opaque link state advertisements (LSAs). In [2], different types of opaque LSAs and their associated format have been discussed. The document has talked about LSA payload details in which one of the top-level Type/Length/Value (TLV) triplets is the link TLV which describes a single link, and is constructed of a set of sub-TLVs. In [3], extensions to the OSPF routing protocol in support of carrying link state information for GMPLS has been presented. The proposed extensions in [2] and [3] can be considered as the base of any new extensions to OSPF supporting TE.

The main challenging issue investigated in this paper regarding the propagating, routing, signalling, or managing information over the entire network is control overheads that the flooded information applies to the network. An improved OSPF-TE protocol has been proposed in [4] so that rather than disseminating link state information through LSAs, a newly designed path sub-TLV called path state advertisements (PSAs) propagates path attributes. Unlike the traditional OSPF-TE, the proposed protocol in [4] has not advertised the absolute value of available link resources. Instead, it has only disseminated resources' increments or decrements to cope with control overheads issues. Although [4] has proposed a path-related extension to OSPF, it has not propagated link or path availability. In addition, inter-AS communications has not been considered in [4]. A new BGP-TE attribute which enables BGP to carry TE-information, has been presented in [5]. In [5], connection bandwidth at different priority levels and switching capability information as the attributes added to BGP for traffic engineering are presented. The idea of disseminating path-related (not domain-related) QoS-metric per destination within an extended TE-attribute has been presented in [6]. The path-related TE-attribute proposed in [6] has been a representative for the overall path from a certain node to the destination. Since the proposed mechanism in [6] has propagated TE-related information without affecting BGP path selection process, it has been considered as a good model for the mechanism presented in this paper. However, in none of the mentioned work link-availability dissemination through a dynamic mechanism has been considered. The present paper expands the PSA concept to overcome the scalability issues concerning control overheads while it considers link and path availabilities as TE constraints over inter-AS communications.

## 3. Scalable Approach to a Novel TE Extensions

***Definition 1*** Path attribute (PA) matrix: The packet routed from one AS to another AS should be routed through one of the edge routers. The routers inside an AS advertise the link attribute of the associated links into the AS. Using this information, the PA matrix is built in all decision making nodes inside an AS including edge routers. Then all the edge routers of an AS have a matrix of the form below for a network with m nodes:

$$PA_{mxm} = \begin{bmatrix} PA_{(1,1)} & \cdots & PA_{(1,j)} & \cdots & PA_{(1,m)} \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ PA_{(m,1)} & \cdots & PA_{(m,j)} & \cdots & PA_{(m,m)} \end{bmatrix}$$

If the $j^{th}$ node is one of the edge routers (ER) as shown in Figure 1, $ER_j$ will advertise the path attribute value of all routes ending at $ER_j$. This information is summarized in the $j^{th}$ column of the PA matrix. In the current operation of OSPF-TE, the label switching routers (LSR)s at each end of a TE link advertise LSAs describing the link. Unlike regular routers inside the AS that only advertise the link attributes, $ER_j$ will advertise all the information of the $j^{th}$ column of the PA matrix of the associated AS using the proposed sub-TLVs of update messages presented in Table 1, in addition to the link attribute of the external link. Here, each edge router maintains two matrices. One from its associated AS which is to be advertised to the other AS, and the other which is received from another AS informing it about the conditions on the neighbouring AS which can be flooded inside the AS. In the case of the NSFNet topology shown in Figure 2, the $j^{th}$ edge router will advertise a path attribute sub-TLV of 14 PA values including the $j^{th}$ column of the PA matrix.

Table 1 Path Attribute sub-TLV

| PA Type | PA Length |
|---|---|
| $PA_{(1,j)}$ | |
| $PA_{(2,j)}$ | |
| …… | |
| $PA_{(i,j)}$ | |
| …… | |

Table 2 Path TLV payload format

| Path Type | Path Length |
|---|---|
| PA value change Sub-TLV | |
| PA Sub-TLV (optional) | |

Table 3 PSA Sub TLV for the path attribute

| PA value change Type | PA value change Length |
|---|---|
| PA value change | |
| PA Type | PA Length |
| $PA_1$ | |
| …. | |
| $PA_m$ | |

One way of reducing link states overhead is to use the concept of PSA [4] rather than LSA. The improved OSPF-TE protocol proposed in [4] not only has disseminated link state information very effectively, but also has advertised link information if necessary. The parameters shown in Table 2 and Table 3 are proposed in this paper to be carried using the path TLV payload and PSA sub-TLV. The PA value change sub-TLV is the change of any desirable attributes for link or path in the lightpath, and the PA sub-TLV indicates the attribute, constraint or metric applied to all links.

As a special case discussed in [4], the "PA value change Sub-TLV" field, in the path TLV payload presented in Table 2, is the increment or decrement of bandwidth in the lightpaths, and the "PA Sub-TLV" field represents the wavelengths used in all links. Theoretical analyses and simulation results in [4] have shown that the OSPF-TE's control overheads could be reduced between 3 and 7 times compared to the conventional flooding mechanism. Based on the graphs presented in [4], the blocking probability of the proposed PSA concept has also been reduced, and the performance of optical networks has been improved significantly.

The standard OSPF-TE LSAs are propagated periodically to advertise the total link resources or when a link is down or restored. These standard OSPF-TE messages can advertise all detailed information about TE links. Otherwise for intra-domain communications, only source LSR floods PSAs, and for inter-domain negotiations, only edge routers flood PSAs [4]. Assuming that not all LSRs flood PSAs, this mechanism can cope with the scalability issues by minimizing the number of flooded advertisements. In addition, to keep the control overhead small, PSAs notify PA changes on a lightpath, rather than notify how much resources are available on these links.

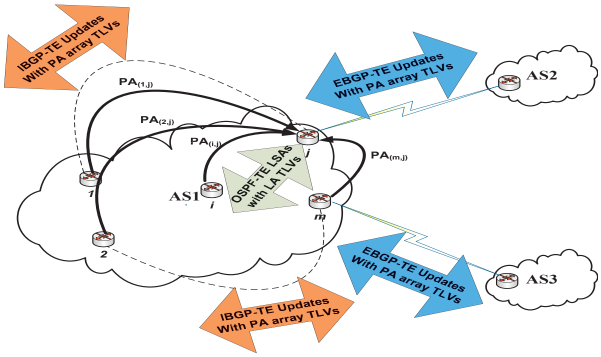


Figure 1 Inter-AS dissemination dynamic SLA-related packets

## 4. Performance Analysis

The network performance is evaluated from three different and important point of view: i) the control overhead reduction employing the PSA concept by which it is shown that the mechanism introduced in this paper is easily scalable; ii) an analysis of the propagation delay of the flooded information over several ASs by which it is shown that the updated messages are fresh; iii) the performance improvement that the mechanism proposed in this paper brings to the network.

### *4.1 Control overhead analysis*

Since control overheads of update advertisements is the main concern, only network resources consumed by OSPF-TE messages are considered. Different types of LSA flooding mechanisms are compared in [4]. In period-based update [4], all LSRs advertise their LSAs for a specific constant interval considered here 100µs. In threshold-based update [4], when the variance of a link's state is below a threshold, LSA is not advertised. Otherwise, this LSA is flooded. In immediate PSA update [4], the proposed PSAs have been advertised immediately. As the results shown in [4], the control overhead of conventional OSPF-TE has been compared to different schemes of update advertisements including the proposed PSA method. As the results in [4] show, the control overheads of the conventional OSPF-TE are about eight times more than the proposed PSA method. When threshold time is considered 3, the control overheads of threshold-based update are about 3 times as much as the proposed mechanism in [4]. The graph in [4] comparing simulation results of different types of LSA flooding also clearly shows the superiority of the PSA concept over LSA.

*4.2 Propagation delay analysis*

The new link/path attributes should be flooded over the network after any changes to the requested connections including establishment and release. Depending on how frequently these changes take place and how long the flooded information may take to be propagated, the validity of the information received by provisioning algorithms should be investigated. That is, an analysis comparing the maximum arrival rate of the connection with the propagation delay of the flooded advertisements should be studied. To clarify this issue, the following considerations including a numerical analysis are presented here.

1) Only the proposed link and path attributes are propagated over the decision making nodes rather than all nodes. The complexity of the data structure exchanged between the nodes is a function of the network size. That is, having a large number of decision making nodes in the network makes the propagated packets huge. The practical approach to this issue is employing the concept of the separation of the control plane from data plane in GMPLS networks. The decision making nodes in the control plane of such networks use the same protocols (OSPF and BGP) and sometimes the same infrastructure as the data plane uses. However, the number of the nodes propagating the information is much fewer than the number of the nodes in the data plane. In the case of NSFNet topology studied in this paper, the decision making nodes are not 14 nodes, but much fewer.

2) Benefiting from the PSA concept discussed in [4], for intra-AS communications, only source nodes flood PSAs, and for inter-AS negotiations, only edge routers flood PSAs. Since not all nodes flood PSAs, the number of the flooded advertisements is minimized. In addition, to keep the control overhead of the advertised packets small, the changes to the desirable attributes are advertised as increments or decrements values, rather than notifying the exact values of the available resources on links or paths.

3) The information regarding paths and related TE metrics among different ASs is exchanged using EBPG Update packets. The decision making nodes within an AS will advertise path attribute values for each destination to their neighbours that are located in other ASs benefiting from IBGP connections. The IBGP runs among routers within the same AS. According to IBGP with a TE extension, when a decision making node in an AS receives the TE path attribute for a destination from another AS, it will send these externally learned paths to internal nodes. Accordingly, the BGP routing table is extended to keep the TE-related information regarding inside and outside of ASs.

4) Assuming an average of 500km as the nodes distance, it takes 2.5ms for PSAs to be propagated on fiber traveling between two adjacent nodes. Since PSAs are advertised as the source-based packets and may take the data plane infrastructure to travel from one decision making node to another, and over NSFNet topology it may travel on average 2.8 nodes (this number is shown in AWPC performance in Figure 4 for a no protection scheme), the average propagation delay would be 7ms.

Therefore, for the average events rate (including connection arrival or release) of less than 145 connections per second, the provisioning algorithms residing in the decision making nodes will be provided by fresh and valid information. The arrival rate considered for simulation purposes in this paper is on average 40 connections per unit of time. Assuming that the only events causing information to be flooded are establishment or release of such connections, and the connections duration on average is one unit of time, on average 80 events per unit of time may trigger the flooding, which keeps the decision making nodes and the algorithms in the safe side in terms of receiving fresh and valid information.

Employing the above considerations, the fresh and valid information will always be available for the algorithms and decision making nodes in the network.

### *4.3 Network performance analysis*

In this section, the blocking probability (BP), the availability satisfaction ratio (ASR), and the average assigned wavelengths per connection (AWPC) of the conventional shared path protection (CSPP) scheme are compared to the standard shared path protection (SSPP) and no protection (NP) schemes. BP denotes the percentage of blocked connection requests over all arriving requests. ASR represents the percentage of provisioned connections whose availability requirements are met over all provisioned connections. AWPC shows the average number of assigned wavelength per connection. Algorithms 1 and 2 introduce and compare the SSPP and CSPP schemes, respectively. Unlike the SSPP, the CSPP scheme considers links' availability as a constraint in the path calculation process. Unlike the SSPP and CSPP schemes, the NP scheme provides no backup paths for any established primary paths.

**Algorithm 1 SSPP scheme**

Input: connection request $\{C_n(s,d)\}$
Output: primary-backup pair of paths

1. $n \leftarrow 1$
2. Serve the $n^{th}$ connection request $\{C_n(s,d)\}$
3. Apply Dijkstra's algorithm [7] based on free wavelengths on the links to find the primary path
4. If primary path exists
    Assign free wavelengths to primary path & Apply Dijkstra's algorithm [7] based on free wavelengths on the links to find the link-disjoint backup path
    If backup path exists
        Assign new wavelengths to backup path if there is no sharable wavelengths
    Else
        Block the request & Go to Step 6
   Else
    Block the request & Go to Step 6
5. Return the primary-backup pair of paths
6. $n \leftarrow n+1$ & Go to Step 2

**Algorithm 2 CSPP scheme**

Input: connection request $\{C_n(s,d)\}$
Output: primary-backup pair of paths

1. $n \leftarrow 1$
2. Serve the $n^{th}$ connection request $\{C_n(s,d)\}$
3. Modify the link cost of the graph using the following equation

$$C_{Pij} = \begin{cases} \infty, & \omega_{ij} = 0 \\ -ln(A_{ij}), & \omega_{ij} > 0 \end{cases}$$

*where $C_{Pij}$ is the cost of the link (i,j) for the primary path and is a function of the link availability, $\omega_{ij}$ is the number of free wavelengths in the link (i,j), and $A_{ij}$ is the availability of the link*

4. Apply Dijkstra's algorithm [7] to find the primary path
5. If primary path exists
    Assign free wavelengths to primary path & Modify the link cost of the graph using the following equation

$$C_{Bij} = \begin{cases} \infty & \omega_{ij} = 0 \\ -ln(A_{ij}) * \omega_{ij} & \omega_{ij} > 0,\ \omega_{rsvd} > \omega_B \\ -ln(A_{ij}) * \omega_{ij} + 1 & \omega_{ij} > 0,\ \omega_{rsvd} \leq \omega_B \end{cases}$$

*where $C_{Bij}$ is the cost of the link (i,j) for calculating backup path, $\omega_{rsvd}$ is the number of the reserved wavelengths on the link for all shared backup paths, $\omega_B$ is the number of required wavelengths in case one of the links forming the primary path fails*

    Apply Dijkstra's algorithm [7] to find the link-disjoint backup path

If backup path exists
Assign new wavelengths to backup path if there is no sharable wavelengths
Else
Block the request & Go to Step 7
Else
Block the request & Go to Step 7
6. Return the primary-backup pair of paths
7. $n \leftarrow n+1$ & Go to Step 2

*4.4 Simulation environment*

The topology selected for the simulation is NSFNet shown in Figure 2 with 14 nodes and 21 bidirectional fiber links of the same physical distance. The links have wavelength conversion capability with 8 wavelengths per link. To simulate a high-risk-network [8], the link availabilities are uniformly distributed between 0.99 and 0.9995 which is assigned per link prior to running the simulation. The average availability of different networks and topologies for various types of protection [9] shows the assumption made for link availability values are reasonable. Connection availability requests are uniformly distributed between 0.99 and 1.0. A Poisson process with arrival rate of $\beta$ is considered for the arrival process of connection requests. The holding time of the connections follows an exponential distribution with the mean value of 1. For simulation purposes, $\beta$ is ranging from 20 to 70 to simulate the offered load [10] of 20 to 70 Erlangs. For the sake of simplicity, it is assumed that all the primary paths whose backup paths share the resources are totally link disjoint and the failure of primary links at the same time is very unlikely. To achieve a 95% confidence interval for the illustrative results, $10^5$ connection requests are simulated in every experiment which may introduce a maximum error of $3x10^{-3}$.

*4.5 Simulation results*

Figure 3 shows how employing the negotiation mechanism proposed in this paper affects the performance of the standard shared mesh protection scheme. Since the CSPP scheme benefits from the negotiation mechanism and uses link availability as a constraint in the path computation, it makes a significant improvement on the number of the connections whose availability requests are satisfied. Clearly, the NP scheme has the worst ASR since it does not protect the paths.

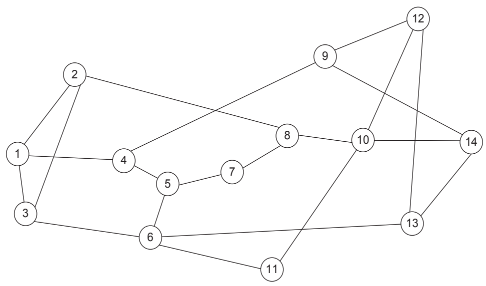


Figure 2 NSFNet network topology

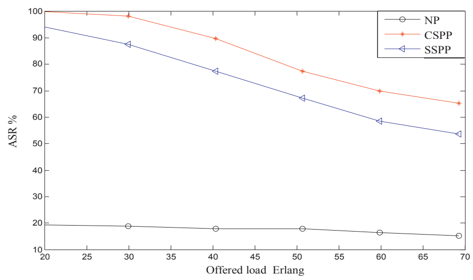


Figure 3 Availability satisfaction rate

The CSPP scheme has better wavelength utilization than the SSPP scheme as shown in Figure 4 since the CSPP decreases the average number of assigned wavelengths to each path by 25% on average compared to the SSPP. However, the NP scheme has a lower wavelength usage in this case since it does not reserve any resources for backup path. Since the routing and wavelength assignment process for both the CSPP and SSPP schemes is the same, the CSPP is not expected to have a significant improvement in blocking probability compared to SSPP, as shown in Figure 5. Blocking probability of the NP is expected to be low since wavelength usage of this scheme is the lowest which keeps the network resources free for future connection requests.

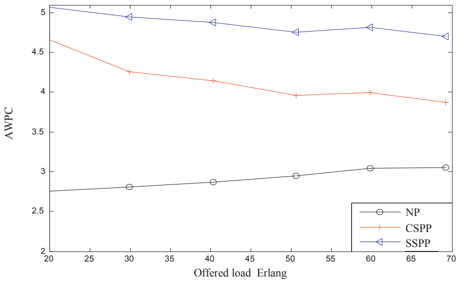


Figure 4 Average wavelength usage per connection

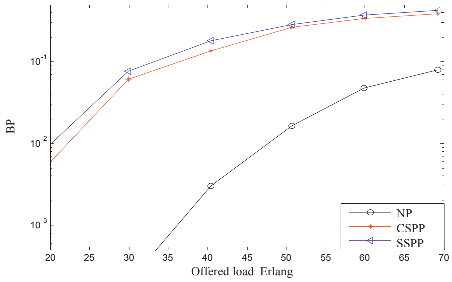


Figure 5 Blocking probability

## 5. Conclusion

This paper has presented a dynamic SLA negotiation mechanism for shared mesh optical networks. The proposed TE extensions applied to OSPF and BGP protocols consider both intra and inter domain communications. Link attributes as an SLA parameter have been negotiated via intra-domain mechanism and any new proposed SLA-based TE path attributes can be advertised through the inter-domain negotiation mechanism. The paper has shown how an SLA parameter negotiation mechanism together with the proposed TE metric can improve the performance of different protection schemes or algorithms. Since the proposed mechanism in this paper may cause heavy control overheads when disseminating link/path attributes, an alternative means of negotiation has been employed to reduce the overheads and resolve the possible scalability issues. The network performance has been evaluated from two different and important points of view: the control overhead reduction, and the network performance improvement.

## References

[1] A. Nafarieh, B. Robertson, W. Phillips, and S. Sivakumar, Dynamic SLA Negotiation Mechanism in Support of Priority-aware Algorithms in Shared Mesh Optical Networks, vol. 2, pp. 324-328, ICINC 2010

[2] D. Katz, K. Kompella, and D. Yeung, Traffic Engineering (TE) Extensions to OSPF Version 2, RFC 3630, September 2003

[3] K. Kompella, and Y. Rekhter, OSPF Extensions in Support of GMPLS, RFC 4203, October 2005

[4] Y. Yin, and G. Kuo, An Improved OSPF-TE in GMPLS-Based Optical Networks, Workshop on High Performance Switching and Routing, pp. 241-145, 2005

[5] H. Ould-Brahim, D. Fedyk, Y. Rekhter, BGP Traffic Engineering Attribute, IETF RFC 5543, May 2009

[6] A. Manolova, S. Ruepp, J. Buron, and L. Dittmann , On the Efficiency of BGP-TE Extensions for GMPLS Multi-Domain Routing, 2009

[7] E.W Dijkstra, A note on two problems in connection with graphs, Numeriche Mathematik, vol. 1, pp. 269-271, 1959

[8] A. Nafarieh, S. Sivakumar, W. Phillips, and B. Robertson, Dynamically provisioned priority-aware algorithms in shared mesh optical networks, Springer LNICST, vol. 74, pp. 290-307, 2012

[9] J. Segovia, E. Calle, and P. Vila, Availability analysis of GMPLS connections based on physical network topology, in ONDM, 2008, pp. 1-6

[10] A. Leon Garcia and I. Widjaja, Communication Networks Fundamental Concepts and Key Architectures: Mc Graw Hill, 2004.