PCE-based Inter-Domain Lightpath Provisioning

Alisson Pontes*, André C. Drummond†, Nelson L. S. da Fonseca*, Admela Jukan‡

*Institute of Computing, University of Campinas
†Dept. Computer Science, University of Brasília
‡Technische Universität Carolo-Wilhelmina zu Braunschweig
alisson@lrc.ic.unicamp.br, nfonseca@ic.unicamp.br, andred@cic.unb.br, jukan@ida.ing.tu-bs.de

Abstract—Despite the research advances in intra-domain lightpath provisioning in WDM networks, efficient and practical schemes for path computation and resource advertisement in multi-domain mesh networks still need to be developed. Few proposals in the literature address open issues such as the scalability of information exchange and the information update necessary to establish an end-to-end path [1], [2], [3], [4]. The approaches behind these solutions are based either on the Border Gateway Protocol (BGP), or the Path Computation Element (PCE) framework.

The schemes based on BGP inherit the main characteristics of this protocol [5], which is the current standard inter-domain routing protocol for IP networks. BGP provides each domain with the means for obtaining and propagating reachability information from neighboring domains and for the definition of routes to other domains. In order to distribute reachability information, different domain border routers disseminate the addresses of all the routers they can reach. When a border router receives this information, it passes it to all the routers in its domain. With the information obtained from other BGP routers, each router can define routes to routers in other domains based on a set of policies. Moreover, the BGP-based schemes also inherit the well known problems of BGP such as the inability to convey useful Traffic Engineering (TE) information, slow convergence and chattiness [2].

The growing need of traffic engineering in backbone networks has led to the proposition of the PCE architecture, standardized by the Internet Engineering Task Force (IETF) [6]. In the PCE-based approach, path computation is carried out by specialized path computation devices, called PCEs. A route computation scheme called Backward Recursive Path Computation guarantees to compute the optimal path across a specific sequence of traversed domains. The PCE architecture does not provide a detailed description of all the architectural components, but rather it describes a set of building blocks for the PCE architecture [7]. Recent community efforts in open source PCE have enabled innovation in those building blocks precisely that are relevant to a specific PCE application within a network [8]. Indeed, the PCE architecture can be viewed as a first step towards the implementation of a constraint-based path computation (traffic engineering) and new solutions needs to be developed on the top of this architecture for the development of future optical backbone control planes.

To this end, much effort still needs to be made to have a complete inter-domain service provisioning solution. As the PCE architecture lacks resources advertisement protocol, the source PCEs need to specify the sequence of domains to be traversed, which can be either administratively predetermined or discovered by some means. Furthermore, the current specification can not handle the wavelength continuity constraints [1]. Also, while work in [9] is attempting to address constraints specific to optical networks within the PCE standards, the current specifications cannot handle wavelength switched optical network specific constraints such as wavelength continuity [1]. Finally, while some prior works address the challenge for reducing signaling in an inter-domain PCE framework [10], current literature does not sufficiently address these concerns when taking connection signaling as well as inter-domain routing updates into consideration.

In this paper, we propose a PCE-based solution for inter-domain lightpath provisioning in WDM mesh networks which takes advantage of using specialized control agents as in the PCE architecture. Our solution provides a new way to compute the chain of domains which is missing in the PCE architecture. A set of policies is introduced to account for the availability of wavelengths providing a complete Routing and Wavelength Assignment (RWA) solution. The proposed scheme defines an on-demand advertising scheme combined with path computation which maintains confidentiality of intra-domain information.

Simulation experiments show the effectiveness of the proposed solution which significantly reduces the total amount of message exchanged and it also significantly reduces the overall call blocking, when compared the Optical BGP (OBGP) solution, [11], [4], [12]. More importantly however, it shows the potential of a PCE framework to be effectively deployed in multi-domain optical networks with multiple routing constraints.
II. RELATED WORK

Distributed control planes such as the Generalized Multi-protocol Label Switching (GMPLS) [13] aim at facilitating the dynamic provisioning of traffic-engineered lightpaths. The IETF has specified two approaches for the computation of multi-domain GMPLS paths: Backward Recursive PCE-based Computation (BRPC) [14] and per-domain path computation [15]. Both use the Constrained Shortest Path First (CSPF) algorithm, which prunes all the links that do not satisfy specific constraints and then executes the Dijkstra’s Shortest Path algorithm on the resulting topology [16].

The source PCE specifies the sequence of domains to be traversed which is carried in the path computation PCE Communication Protocol (PCEP) [17] request message and the BRPC computes the optimal path across the specific sequence of domains. The sequence is either administratively predetermined or discovered by some means, however, the way to determine it was left open in the standard [14], which motivated the algorithm introduced in this paper.

In the per-domain approach, path computation methods are usually defined based on auto-discovery mechanism such as that of OBGP. The complete path is obtained by concatenating segments that are computed for every domain. However, the resulting path is usually obtained from outdated information in routing information bases. Thus, to guarantee a good level of confidence, the path computation must take advantage of well designed routing and resource advertisement protocols.

Both schemes of path computation fail to handle the Wavelength Continuity Constraints (WCC) mainly due to the lack of network resources availability (NRA) information. Indeed, to handle the previously mentioned issues, domains must be able to exchange information on both, reachability, and resource availability. As a consequence, these multi-domain routing schemes do not have all the information needed for the establishment of multi-domain lightpaths.

Some previous work tried to address these issues. Casellas et al. [1] extends the BRPC algorithm to handle the end-to-end WCC; Francisco et al. [4] provided the first implementation of OBGP (as far as we know), specifying requirements and the necessary extensions of BGP to create OBGP; Yannuzzi et al. [2], [3] reinforced the need of advertising aggregated path-state information for lightpath provisioning in WDM networks, and defined a process for the aggregation of wavelength availability information. Greco et al. [18] developed a security model for PCE to be applied for inter-domain networks, both in a cascaded and alliance based peering models.

A relevant missing point in some of these proposals is the triggering event for resources/virtual topology advertisement update messages which is critical to the signaling load experienced by the network. In the proposal in [10], the authors propose a mechanism for pre-reserving inter-domain resources and triggering updates when resource levels reach specific thresholds in order to reduce inter-domain signaling overhead while reducing inter-domain connection blocking. However, as the proposal is not evaluated in conjunction with a routing protocol typically used for evaluating a domain chain, the issue of increased signaling load based on the current set of triggering events on the PCE remains unsolved.

III. INTER-DOMAIN LIGHTPATH COMPUTATION

The proposed scheme for inter-domain service provisioning uses the Multiple PCE Path Computation with Inter-PCE Communication Architecture [7], which means that there are at least one PCE per domain, that can perform inter-domain routing based on information in its own Traffic Engineering Database (TED). Multi-domain paths are computed by means of the information exchanged among the PCEs.

The proposed solution introduces a novel algorithm for information dissemination and computation of domain chains, and it implements a selection policy to better balance the lightpath load. The messages exchanged in our solution include both reachability and resource availability information which makes possible to select a potential wavelength to establish an end-to-end lightpath that is being explored during the computation of the domain chain. Moreover, we use a backtracking technique to cope with outdated TED information which enhances the capability of computing the domain chain. Backtracking is a scheme similar to the RSVP crank-back scheme [19] implemented when path computation failure is returned from the point of failure to allow new path computation attempts to be made, avoiding the blocked resources.

A novel message dissemination scheme triggers the exchange of resource availability update messages only when a backtrack event occurs. This dissemination involves only the previous domain, allowing the domain to choose for routes based on updated information without flooding the entire network. Finally, the proposed wavelength selection scheme allows the end-to-end path establishment to be based on the distribution of available wavelengths in the network, yielding a balanced routing solution.

All the PCEs run a path vector with path caching protocol [20], which means that each PCE has in its TED one or more defined paths (domain chains) from it to all other domains in the network. Actually, a PCE can hold in its TED, at most, a number of paths to a given destination equal to the number of edge nodes. The flexibility to be able to choose among multiple paths, by the use of path caching increases the protocol reliability. Each TED entry contains a list of available wavelengths and the output border node to the next domain in the chain of domains; an output border node is a node which has a link to a neighboring domain. When a PCEP request for multi-domain lightpath arrives at the source PCE, it must first choose available routes to the destination. The first criterion in this selection determines that paths with the highest number of available wavelengths should be chosen. If there is more than one choice, then the path with the shortest domain chain should be chosen. At last, if more than one choice exists, BGP tie-breaking rules should be followed. After selecting a path, the PCE randomly picks one wavelength among the available ones for that path. These criteria yields a balanced distribution.
of resources and consequently increases resources availability by avoiding the formation of bottlenecks.

After defining the domain chain and wavelength, the source PCE signals the local network to allocate resources necessary to support the call and forwards another PCEP request to the next PCE along the domain chain. The neighbor PCE which received the PCEP request perform the same procedure considering the wavelength chosen by the source PCE and the output border node of previous domain. The procedure continues until the PCE of the destination domain is reached and an end-to-end lightpath can be established. The final step is left to an end-to-end computation path procedure and a resource reservation setup protocol such as BRPC and RSVP [21].

The proposed solution can be better understood by means of an example. Let us consider the network, shown in Figure 1 composed of five domains, each with a PCE. PCE2 is aware of intra and inter-domain links of domain D2 and, thus, knows about the availability of resources to reach D1. Update messages received by PCE3 from PCE2 are analyzed and stored in the TED of PCE3, which now has updated information on D1. These messages contain information on: i) the destination domain (D1), ii) the input border nodes from D2 and iii) the available wavelengths between the input border node of D2 and the input border node of the destination domain (D1); an input boarder node is a node which has a link to previous domain in the chain. In the same way, PCE3 can furnish updating information to the TED of PCE4 extending the path to D1 and defining the available wavelengths through the intersection of available wavelengths in the path D1-D3 and D3-D4. The TED of PCE5 can be finally updated in the same way from PCE3 and PCE4.

At the end of this process, PCE5 should have a vision of the whole topology shown in Figure 2, which includes two possible paths to reach D1, one via D3 (dashed line) and other via D4 (continuous line). Moreover, it does not have any information on the internals of any other domain, which is an important measure that guarantees administrative independence.

The protocol is summarized in Algorithm 1 and in Algorithm 2.

The most relevant contribution of the proposed scheme is the resource availability dissemination procedure used to update the TED of the PCEs. Unlike BGP-based solutions, which uses frequent update messages, the proposed solution employees on-demand notification, the update messages are triggered only by backtrack events [16].

If a PCEP request message arrives at a PCE and the target wavelength is no longer available in that domain a backtrack event occurs and a message is sent back to the previous

---

**Algorithm 1 Inter-Domain Lightpath Computation**

**Require:** Each PCE runs a path vector algorithm with path caching.

**Require:** Each TED path entry has a list of available wavelengths attached.

**Ensure:** A multi-domain end-to-end lightpath.

1: The source PCE receives a call request to establish an end-to-end lightpath. It refers to its TED to select to which neighbor it will forward the PCEP request message;
2: Use path selection policy (Algorithm 2);
3: Randomly picks one wavelength from the list of available wavelengths of the chosen path;
4: Send messages to the intra-domain nodes to allocate resources necessary to support the call from the source node to the chosen output border node;
5: repeat
6: PCE receives a message;
7: if It is a backtrack message then
8: Update its TED;
9: The PCE refers to its TED to select to which neighbor it will forward the PCEP request message;
10: if There is no path to the destination then
11: Send back a backtrack message piggybacked with a state update message;
12: Use path selection policy (Algorithm 2);
13: Send messages to the intra-domain nodes to allocate resources necessary to support the call from the input border node to the chosen output border node;
14: until The PCEP request reaches the destination domain OR n backtracks were triggered;
15: if The PCEP request reaches the destination domain then
16: perform end-to-end path computation and resource reservation setup;
17: else
18: The call is blocked;
19: de-allocate resources not used;

---

**Algorithm 2 Path Selection Policy**

**Require:** Each TED path entry has a list of available wavelengths attached.

1: Randomly picks one wavelength from the list of available wavelengths of the chosen path;
2: Send messages to the intra-domain nodes to allocate resources necessary to support the call from the source node to the chosen output border node;
3: until The PCEP request reaches the destination domain OR n backtracks were triggered;
4: if The PCEP request reaches the destination domain then
5: perform end-to-end path computation and resource reservation setup;
6: else
7: The call is blocked;
8: de-allocate resources not used;
Algorithm 2 Path Selection Policy

Require: Given a destination domain and all the possible paths (domain chains) to reach it.
Ensure: Outputs a single path to the destination domain.

1: if There is more than one output neighbor domain then
2: Choose the one that has more available wavelengths in the path between it and the destination domain;
3: if There is more than one output neighbor domain then
4: Choose the shortest path (domain chain hops);
5: if There is more than one output neighbor domain then
6: Run BGP tie-breaking rules.

The proposed approach provides a lightweight solution by combining path computation and resource advertisements. As a result, the proposed solution addresses some of the main limitations of the BGP-based solution, such as: (i) the inability to convey useful traffic engineering information, achieved by the use of the PCE architecture; (ii) lack of multipath routing, made possible by the path caching scheme which allows the source PCE to pick more than one path if desirable; and (iii) slow convergence and chattiness, addressed by the backtrack messages which carries update information triggered only if necessary during the establishment of an end-to-end lightpath.

IV. PERFORMANCE EVALUATION

In this section, the effectiveness of the proposed scheme is assessed and compared to that of Optical BGP [11], [4], since effort has been made to adapt BGP to the GMPLS architecture [22]. For that, the WDMSim simulator [23] was extended to allow simulation involving multi-domains. Topologies used in the simulations were the NOBEL-EU (Fig. 4.) and the NEWYORK (Fig. 5.) mesh topologies, which description can be found in the library of test instances for Survivable fixed telecommunication Network Design (SNDlib) [24]. The NOBEL-EU topology was defined in the COST 266 European project [25], and it has been used for protocol evaluation. It is composed of 28 domains and 41 inter-domain links, resulting in a mean domain connectivity of 2.93. Nodes were chosen to include some of the main Internet exchange points. The NEWYORK network represents a telecommunication network in the greater New York area, but its origin is not known due to non-disclosure agreements. It has 16 domains and 49 inter-domain links, resulting in a mean domain connectivity of 6.12. These two topologies have different number of nodes and connectivity.

The nodes in each domain are fully connected and there are as many inter-domain links as the number of nodes at the border. In this way, blocking due to unavailability of intra-domain paths are avoided. The Dijkstra algorithm is used to define intra-domain paths.

Connection requests are uniformly distributed among all pair of nodes in the network. The network load varied from 20 to 200 erlangs and each simulation involved 100,000 connection requests. Confidence intervals with 95% confidence level were derived using the method of independent replication. At least 10 replication were generated. The blocking probability and the signaling overhead were assessed in the simulations. The signaling overhead considers all control signaling generated by the inter-domain routing protocol to discover and
maintain resource availability. In the proposal, these messages are triggered at backtracks. The MinRouteAdvertisementIntervalTimer parameter in the OBGP protocol, that determines the amount of time that must elapse between two eBGP advertisements, was set to 30 seconds, as suggested in [5]. Moreover, every changes detected triggers an advertisement for iOBGP.

Figures 6 and 7 show the blocking probability for the NOBEL-EU and NEWYORK topologies, respectively. The blocking probability given by the OBGP increases sharply as the load increases which did not happen with the blocking probability generated by the proposed scheme. The difference between the curves increases significantly with load increase in both topologies. For the NOBEL-EU, it is almost 13% while for the NEWYORK it is almost 4%. For the NEWYORK topology which have higher node connectivity, the increase in the blocking probability produced by the proposed scheme is quite low even under high loads facing around 0.1%, while the blocking produced by OBGP is 3.5%.

The proposed scheme balances the load by choosing inter-domain links which have the highest number of wavelengths available. Moreover, by updating the routing tables only at backtrack times leads to more stable routing tables.

In the OBGP protocol, whenever a request cannot be forwarded to the next domain in the path establishment procedure, it is blocked and updating messages are flooded in the network. OBGP does not attempt to recover from a failed trial which increases blocking.

Figures 8 and 9 show the signaling overhead for the NOBEL-EU and NEWYORK topologies respectively. The difference in the number of signaling message sent is quite striking. While OBGP generates 7 to 8 times more message for the NOBEL-EU topology, the number of message sent by OBGP in the NEWYORK topology (higher number of nodes and connectivities) is five orders of magnitude higher than that generated by the proposed scheme. This large difference is mainly due to the on-demand updating scheme of our proposal which is in clear contrast with OBGP flooding updating scheme.
Fig. 9. Routing advertisements as a function of the load for the NEW YORK topology.

V. CONCLUSION

This paper introduced a novel PCE-based scheme for the provisioning of lightpaths in multi-domain WDM optical mesh network scenarios. The proposed domain chain computation procedure preserves intra-domain confidential informations. On-demand dissemination of information on reachability and resource availability allows a lightpath establishment procedure which includes backtracking as a solution for alternative attempts of those which failed. A proposed RWA algorithm balances the load based on the number of available wavelengths per path which avoids the formation of bottlenecks and consequently decreases blocking. Both blocking and signaling overhead are significantly lower than those produced by OBGP.

Our future work includes a more detailed performance analysis considering the impact of intra-domain connection load in the network and the various aspects of signaling involving the backtracking mechanisms. We should also investigate what would happen in case of failures in the network when path vector updates are sent only when backtracking occurs.

REFERENCES


[34] Survivable fixed telecommunications network design library. [Online]. Available: http://www.ikr.uni-stuttgart.de/


[38] Survivable fixed telecommunications network design library. [Online]. Available: http://www.ikr.uni-stuttgart.de/


[40] Survivable fixed telecommunications network design library. [Online]. Available: http://www.ikr.uni-stuttgart.de/


