Optical Corridor Routing Protocols

Shovan Das, Paolo Monti, Marco Tacca, Andrea Fumagalli*
Open Networking Advanced Research (OpNeAR) Lab,
Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas (UTD), Richardson, TX, 75083, USA
**Tel: +1 (972) 883 6853, Fax: +1 (972) 883 2710, e-mail: andreaf@utdallas.edu*

ABSTRACT
The scalability of routing and resource advertisement protocols is a key issue in large optical networks. A commonly used solution is based on the hierarchical approach adopted in IP/MPLS-TE networks, i.e., multiple areas or autonomous systems are pre-established to contain the geographical scope of the resource advertisement protocols and the related routing information. With arbitrarily pre-established areas, routing decision might be sub-optimal. Thus, special care must be paid by the network designer to subdivide effectively the network into areas. This paper presents and discusses an alternative routing technique that attempts to improve both optimality and scalability of routing and resource advertisement protocols without requiring the use of manually pre-established areas.
Keywords: corridor routing, optical networks, self-configuring networks, path computational element, end-to-end routing.

1. INTRODUCTION

A commonly used solution to address the issue of scalability of routing and resource advertisement protocols is based on a hierarchical approach. The network designer, based on his expertise, partitions the network into multiple areas or autonomous systems (AS) to contain the geographical scope of the link state advertisement (LSA) protocol [1][2]. Information made available by the LSA protocol is used by the routing algorithm to handle *intra-area* routing. Another protocol, e.g., the border gateway protocol (BGP), distributes aggregate information that can be used by the routing algorithm to handle *inter-area* routing. A more recent improvement to BGP is based on the path computational element (PCE) concept [3]. Each PCE has a complete view of its own area. Inter-area routing is then performed jointly by a group of PCEs, each one computing an intra-area segment of the end-to-end inter-area route. While this approach is used in today's networks, it requires the network designer's expertise in choosing the areas as well as the border routers (BRs) used to interconnect adjacent areas.

This paper extends the PCE concept to investigate an alternative routing solution that attempts to improve optimality of routing decisions, while keeping the scalability properties of the hierarchical approach. The solution is based on a self-configuring network that does not require any human expert intervention. Although more general, the solution is illustrated using a wavelength routing network, e.g., routers are replaced by optical crossconnect (OXC) nodes. The key concept used here is that PCEs, areas, and BRs are dynamically identified by the routing protocol during the network lifetime. For example, they may be chosen on a per connection basis to provide best operation and results. The collection of chosen PCEs, areas, and BRs constitutes a subgraph of the graph representing the entire network. The subgraph is referred to as a *corridor*, within which the end-to-end path is computed. Due to this mechanism, the solution is referred to as *corridor routing* (CR). The CR solution is demonstrated in the paper in conjunction with a flooding technique, which is used to first identify the PCEs. Then, the chosen PCEs determine both the areas and BRs, and compute the end-to-end path in the corridor using PCE standard solutions [4].

It must be noted that routing solutions based on corridors have been previously proposed for wireless networks [5]-[7]. The commonality between the wireless solutions and the CR solution is that they are based on a subgraph, which is used to limit the search space for the end-to-end path. However, the CR solution's innovation is in the way the subgraph is built and the end-to-end path is computed, as described in more details in section 2.

Simulation results are discussed to quantify both the blocking probability due to the lack of available resources and the amount of flooding for the presented CR solution, showing both quasi optimality of the former, and ability to contain the latter. Perhaps the most significant advantage of the CR solution is its ability to self-create a two level hierarchy in the network, which in turn mitigates the scalability and convergence problem of the LSA protocol. A second advantage is its adaptability to both traffic and topology changes, as the corridors may be regularly updated during the network lifetime to best fit the current network status.

The work reported in this paper was supported in part by the National Science Foundation grant CNS-0435393, and the Italian Ministry of University (MIUR) (contract n. RBNE01KNFP).

2. AN EXAMPLE OF CORRIDOR ROUTING PROTOCOL

Consider a wavelength routing network with an arbitrary large physical topology, where every node is an OXC with wavelength conversion capabilities. The network architecture is flat, i.e., all nodes are the same and may function as both PCE and/or BR.

The topology is modelled as a graph *G(N, A)*, where *N* is the set of nodes in the network and *A* is the set of directed (fiber) links connecting the nodes. It is assumed that each link in the network can accommodate *F* fibers for each direction of propagation, each one carrying *W* data wavelengths. One dedicated wavelength is reserved on each link to support out of band signaling. Each node is uniquely identified, e.g., node *i* is denoted as N_i.

Connection requests arrive randomly at the source and must be served as they are received. A connection request consists of a lightpath [8] that needs to be established between a given source (N_s) and destination (N_d). The bandwidth of the lightpath equals the wavelength bandwidth and splitting of lightpaths is not allowed. A connection request is blocked and discarded if the path connecting the source-destination pair does not have enough bandwidth to accommodate the lightpath.

An LSA protocol with limited geographical scope [9] runs on the signaling wavelength and provides each node with the information about the available wavelengths on the links that belong to the node's *virtual area* (VA). For example the node's VA can be identified by the maximum span of the transmitted optical signal without requiring signal regeneration, i.e., VA is identified by the node's transparency island [9]. For example, the node's VA can be defined as follows.

Let $C_{(i,j)}^{(k)}$ be the set of all links and nodes that can be used to establish a simple path connecting N_i to N_j with no more than *k* hops. Set $C_{(i,j)}^{(k)}$ defines a subgraph of *G(N, A)*. Let $\mathrm{VA}_i^{(k)}$ be the *k*-hop VA of N_i. $\mathrm{VA}_i^{(k)}$ is a subgraph of *G(N, A)*, defined as $\mathrm{VA}_i^{(k)} = \bigcup_{j \in N} \left(C_{(i,j)}^{(k)} \right)$.

If $N_j \in \mathrm{VA}_i^{(k)}$, then it is possible to route a lightpath, whose span is at most *k* hops, from N_i to N_j. Note that the value of *k* may be changed over time, and some VAs partially overlap, i.e., their intersection is not empty. While the VA may resemble the concept of AS partitioning into pre-defined multiple areas [10][11], a substantial difference is that pre-defined areas do not change in time, and each node belongs to one area only[1], i.e., pre-defined areas do not overlap.

The CR protocol works as follows[2]. Upon reception of a connection request, N_s first checks if N_d belongs to its VA. If so (*intra-VA* case), the shortest available path within the VA is found by N_s itself, i.e., the minimum hop path [12] considering only the links that have available wavelengths. Otherwise (*inter-VA* case), the protocol works in two steps: first the sequence of PCEs is selected and then the end-to-end path form N_s to N_d is computed. The sequence of PCEs for the source-destination pair is determined as follows. N_s broadcasts a *path discovery* message to a subset of nodes (the *bordercasting nodes* [13] or BCs) in its VA, requesting to reach N_d. BCs are chosen within a *bordercasting radius k'*, i.e., BCs are chosen to be $k' \le k$ hops away from N_s, accounting only for the links with at least one available wavelength. Each BC in turn contacts its own BCs unless N_d is inside the BC's VA. Multiple messages may eventually reach N_d, each indicating a unique sequence of BCs. The first path discovery message reaching N_d identifies the sequence of BCs chosen to be the PCEs for the source-destination pair. Let $P_{(s,d)}^{(k')} = \{N_0, N_1, ..., N_{i-1}, N_i, N_{i+1}, ..., N_z\}$ be an ordered set of nodes where N_0=N_s, N_z=N_d, and the remaining nodes are the BCs chosen to be the PCEs for the source-destination pair. Each one of these PCEs uses its own VA to compute a segment of the end-to-end path from N_s to N_d.

The entry BRs (Figure 1) for a given PCE (N_i) are the optical nodes that belong to the subgraph defined as $B_{(i,entry)}^{(k)} = \mathrm{VA}_{i-1}^{(k)} \bigcap \mathrm{VA}_{i+1}^{(k)}$, where N_{i-1} and N_{i+1} are the predecessor and the successor of N_i in $P_{(s,d)}^{(k')}$. The exit BRs for N_i are the entry BRs of the successor of N_i, i.e., $B_{(i,exit)}^{(k)} = B_{(i+1,entry)}^{(k)}$. For the last PCE in $P_{(s,d)}^{(k')}$, $B_{(z-1,exit)}^{(k)} = \{N_d\}$. The nodes in $P_{(s,d)}^{(k')}$, their VAs and BRs constitute a subgraph or *corridor*. For a given source-destination pair and a bordercasting radius *k'*, the *corridor* is defined as $C_{(s,d)}^{(k,k')} = \bigcup_{N_i \in P_{(s,d)}^{(k)}} \mathrm{VA}_i^{(k)}$. Proceeding backward from N_{z-1} to N_1, each PCE in $P_{(s,d)}^{(k')}$ computes the multipoint to point tree (MP2P) of the shortest available paths to connect its entry BRs to N_d, and passes the paths onto its predecessor in the sequence. The MP2P tree is computed using the backward recursive PCE-based computation procedure (BRPC) defined in [4]. N_s concatenates the results of all the PCEs by performing its own MP2P tree computation and determining the shortest available end-to-end path to N_d. This is the end-to-end path chosen to establish the requested connection.

After choosing the path, N_s can reserve one wavelength on each link of the path using some known reservation protocol, e.g. RSVP [14][15].

[1] With the exception of area border routers.

[2] Many CR protocol variants are possible, but they are not within the scope of this paper.

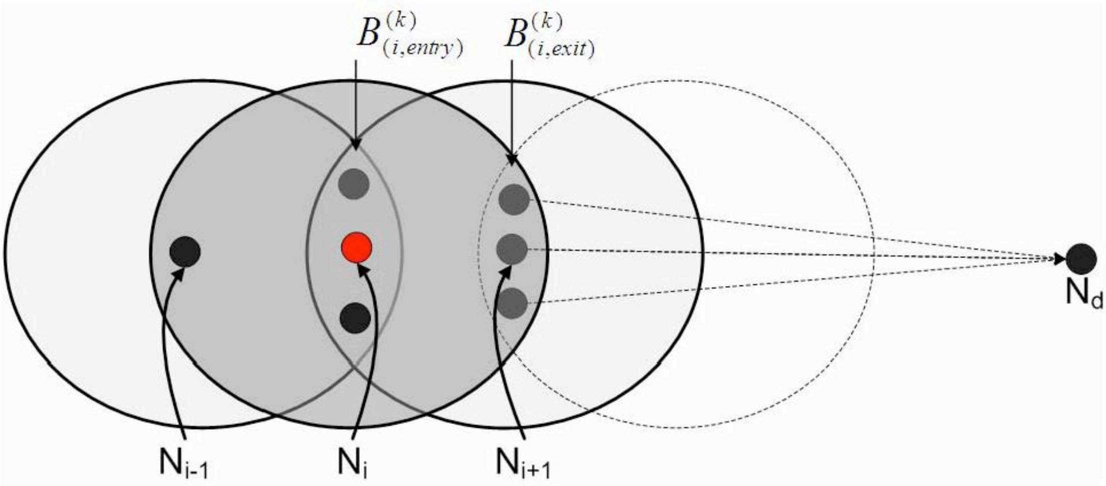

Figure 1. Entry and exit BRs.

The sequence $P^{(k')}_{(s,d)}$ is saved at N_s, and used for future connection requests that are intended for N_d, thus containing the number of times a path discovery message is broadcast across the network. However, if the stored $P^{(k')}_{(s,d)}$ can no longer provide an end-to-end path to reach N_d (e.g., lack of available wavelengths, or network failure) a path discovery message is broadcast by N_s once again to identify a new sequence of PCEs (and a new corridor) to be used.

3. PERFORMANCE

This section reports the potential gain in using the CR protocol presented in section 2 when compared against some benchmark routing protocols. Two metrics are reported: blocking probability and bordercasting frequency. The bordercasting frequency is defined as the ratio between: *(i)* the total number of times bordercasting has to be performed and *(ii)* the total number of connection requests considered. The benchmark routing protocols are described next.

3.1 Benchmark routing protocols

Three benchmark routing protocols are considered. The first protocol (*fixed-SP routing*) neither requires path discovery messages nor LSA protocols. The path connecting each source-destination pair is the shortest path [12], computed on $G(N, A)$ regardless of link wavelength availability.

The second protocol (*ideal routing*) is a centralized solution, whereby each connection request is routed along the shortest path, computed taking into account what are the available wavelengths at the moment.

In the third protocol (*memory routing* (*mem-R*)) the sequence of nodes used to connect a source-destination pair is found using bordercasting, i.e., set $P^{(k')}_{(s,d)}$ is defined by the first path discovery message reaching the destination. The end-to-end path from the source to the destination is obtained concatenating the shortest paths (computed taking into account each link wavelength availability) between adjacent nodes in set $P^{(k')}_{(s,d)}$ .The end-to-end path is then stored at the source. When the most recently computed path for a given source-destination pair cannot accept more connections due to lack of wavelengths, bordercasting is repeated. This solution is used to evaluate the effect of the width of the corridor (determined by both the VA size (k) and the bordercasting radius (k')), as opposed to simply using the shortest available path found by the bordercasting procedure.

3.2 Simulation results

The CR protocol presented in section 2 is tested using the Pan American optical network shown in Figure 2. The network comprises 79 nodes and 102 bidirectional links. Each link accommodates $F = 1$ fiber for each direction of propagation and each fiber supports $W = 3$ data wavelengths. Connection request arrivals constitute a Poisson process with arrival rate λ. Source and destination of each request are randomly chosen with a uniform distribution. Each established connection is held for a time equal to a random variable with exponential distribution and parameter $\mu = 1$. The value of the bordercasting radius is $k' = 3$. The total number of connection requests considered is 100,000. For simplicity, race conditions are ignored, i.e., each request is completely handled before the next arrival.

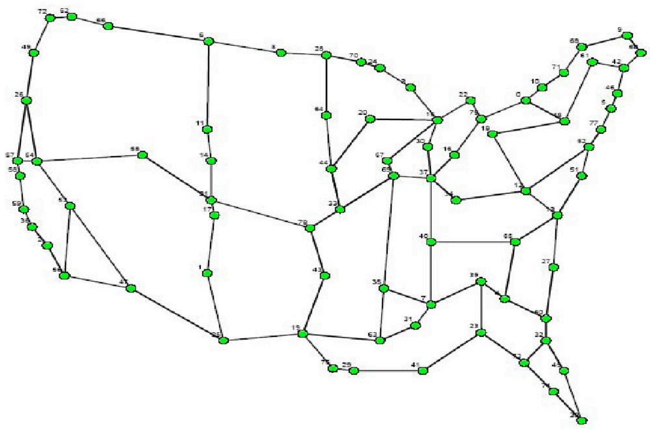

Figure 2. The Pan American optical network.

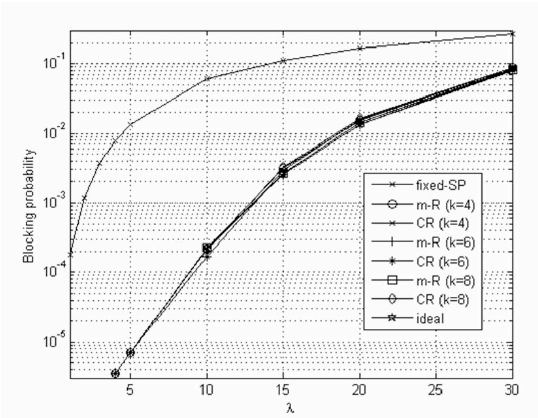

Figure 3. Blocking probability as a function of the network load (λ). k' = 3.

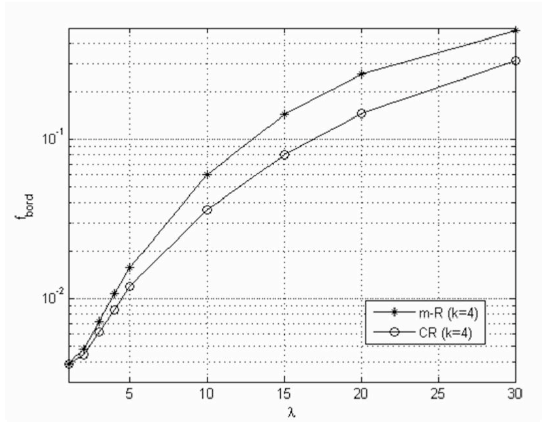

Figure 4. Bordercasting frequency (f_{bord}) as a function of the network load (λ). k = 4,k' = 3.

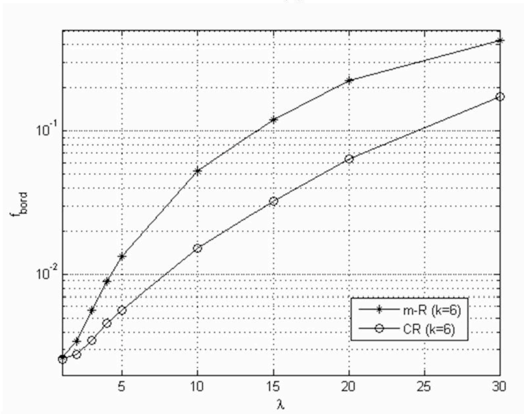

Figure 5. Bordercasting frequency (f_{bord}) as a function of the network load (λ). k = 6,k' = 3.

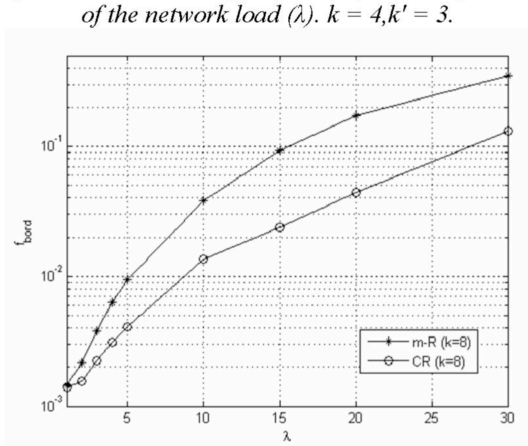

Figure 6. Bordercasting frequency (f_{bord}) as a function of the network load (λ). k = 8,k' = 3.

Figure 3 shows the value of the blocking probability as a function of the network load λ. Three values for the VA's size are considered: $k = 4$, $k = 6$ and $k = 8$. The CR protocol performance is compared against the three benchmark routing protocols described in section 3.1. The figure shows that the CR protocol is able to perform as well as the *ideal routing* protocol.

Figure 4, Figure 5, and Figure 6 show the value of the bordercasting frequency as a function of the network load and different values of the VA's size. The bordercasting frequency is computed only after that 16,000 connection requests have been received, to discard the cold start transient. The figures confirm the ability of the CR protocol to contain the number of times bordercasting has to be performed when compared to the *memory routing* protocol.

4. CONCLUSIONS

Corridor routing (CR) is a solution to cope with the scalability problem of routing and resource advertisement protocols in large networks. For illustration purpose, the concept of CR was applied to wavelength routing networks, but can be easily extended to other traffic engineering solutions, e.g., IP/MPLS-TE. As already mentioned, the first advantage of the CR solution is its ability to self-create a two level hierarchy in the network, which in turn mitigates the convergence problem of the LSA protocol. A second advantage is its adaptability to both traffic and topology changes, as the corridors are regularly updated during the network lifetime to take advantage of available network resources. A third advantage is its adaptability to the actual transmission impairment of the optical layer, as the CR protocol can dynamically adjust the size of the virtual area [9], e.g., to best fit the chosen optical transmission rate. A forth advantage is a new dimension that the CR solution can exploit, i.e., the width of the corridor. The width of the corridor is determined by both the LSA protocol scope (e.g., the virtual area size) and the hop distance between the PCEs chosen for a given connection request. Thus, it is possible to trade convergence time of the LSA with the number of PCEs involved to process a newly generated connection request.

It must be finally noticed that the presented study on the CR solution is based on a simplistic scenario, i.e., unprotected connections, single class traffic, and absence of race condition when reserving wavelengths. The CR solution may provide additional advantages, e.g., computation of (link/edge disjoint) secondary path, QoS provisioning, and reduction of race condition thanks to the local computation of the path segments performed by each PCE. These and other potential advantages of the CR solution will have to be carefully assessed in future studies.

REFERENCES

[1] J. Moy: RFC 2328 – OSPF version 2, April 1998.

[2] D. Katz: RFC 3630 – Traffic engineering (TE) extensions to OSPF version 2, September 2003.

[3] J.P. Vasseur, J.L. Le Roux: Path computation element (PCE) communication protocol (PCEP), Networking Group Internet Draft, March 2007.

[4] J.P. Vasseur, R. Zhang, N. Bitar, J.L. Le Roux: A backward recursive PCE-based computation (BRPC) procedure to compute shortest inter-domain traffic engineering label switched paths, Networking Group Internet Draft, March 2007.

[5] C. Vogt, C. de Waal, M. Gerharz: Corridor routing in mobile wireless ad-hoc networks, in *Proc. 3rd Mediterranean Ad Hoc Networking Workshop*, Bodrum, Turkey, Jun 2004.

[6] V. Subraminian, R. Arumugam, A. A. Minai: Self-organization of connectivity and geographical routing in large-scale sensor networks, in *Proc. 4th International Conference on Complex Systems*, Nashua, NH, USA, June 2002.

[7] P. Samar, M. R. Pearlman, Z. J. Haas: Independent zone routing: an adaptive hybrid routing framework for ad hoc wireless networks, *IEEE/ACM Trans. on Net.*, vol. 12. no. 4 pp. 595-608, August 2004.

[8] I. Chlamtac, A. Ganz, G. Karmi: Lightpath communications: a novel approach to high bandwidth optical WAN's, *IEEE Trans. on Commun.*, vol. 40. no. 7. pp. 1171-1182, July 1992.

[9] S. Das, R.R. Tabrizi, P. Monti, M. Tacca, A. Fumagalli: A link state advertisement protocol for optical transparency island, in *Proc. IEEE HPSR 2007*, New York, NY, USA, May-June 2007.

[10] R. Rastogi, Y. Breitbart, M. Garofalakis, A. Kumar: Optimal configuration of OSPF aggregates, *IEEE/ACM Trans. on Net.*, vol. 11. no. 2. pp. 181-194, April 2003.

[11] A. Shaikh, D. Wang, G. Li, J. Yates, C. Kalmanek: An efficient algorithm for OSPF subnet aggregation, in *Proc. IEEE ICNP*, Washington, DC, USA, 2003.

[12] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein: *Introduction to Algorithms* (Second Edition), Cambridge, MA: MIT Press, 2001.

[13] Z. J. Haas, M. R. Pearlman, P. Samar: The bordercast resolution protocol (BRP) for ad hoc networks, IETF, MANET Internet Draft, July 2002.

[14] R. Braden: RFC 2205 - Resource reservation Protocol (RSVP) - version 1 functional specification, September 1997.

[15] D. O. Awduche, RFC 3209 - RSVP-TE: extensions to RSVP for LSP tunnels, December 2001.