Hierarchical Border Gateway Protocol (HBGP) for PCE-based Multi-domain Traffic Engineering

Luca Buzzi*, Matteo Conforto Bardellini*, Domenico Siracusa*, Guido Maier*,
Francesco Paolucci$^\diamond$, Filippo Cugini‡, Luca Valcarenghi$^\diamond$, Piero Castoldi$^\diamond$
* Department of Electronics and Information, Politecnico di Milano, Milan, Italy
‡ CNIT, Pisa, Italy
$^\diamond$ Scuola Superiore Sant'Anna, Pisa, Italy
Email: {siracusa,maier}@elet.polimi.it, filippo.cugini@cnit.it, {luca.valcarenghi, fr.paolucci, piero.castoldi}@sssup.it

***Abstract*— In multi-domain multi-carrier networks the effective use of network resources shall be achieved while guaranteeing an adequate level of confidentiality and scalability. A candidate solution to perform effective multi-domain Traffic Engineering (TE) is based on a combination of (i) hierarchical routing and (ii) path computation procedures. Hierarchical routing identifies the domain sequence to cross while path computation computes the strict end to end path. In this multi-domain study we first propose a hierarchical instance of BGP (HBGP) dedicated to TE information only. Then we propose and evaluate the integration of HBGP with path computation procedures based on IETF PCE architecture. Simulation results show that the hierarchical HBGP-PCE architecture, compared to current routing solutions based on BGP only, significantly improves the overall network resource utilization. In addition, this study identifies the network scenarios in which the aforementioned HBGP-PCE features provide significant advantages. Finally, the experimental implementation of the proposed HBGP-PCE architecture in a network testbed composed of commercially available routers shows the viability of the solution in real networks.**

***Index Terms*— Multi-domain networks, Traffic Engineering, PCE, BGP.**

I. INTRODUCTION

Providers are operating in an environment that includes multiple domains, for which an efficient computation of end-to-end paths between source and destination nodes belonging to different administrative domains is frequently required. Advances in networking technologies and increasing heterogeneous demands by customers have spurred the introduction of traffic engineering (TE) techniques to efficiently manage resources and provide Quality of Service (QoS). However, multi-domain TE is complicated by the limited visibility of TE information which is usually restricted to a single domain for scalability and confidentiality issues.

Several solutions to provide multi-domain TE propose to extend the functionalities of the Border Gateway Protocol (BGP) [1], [2]. However, BGP has been designed to provide reachability information within a scope that covers the whole Internet. As a consequence, mainly because of scalability issues, such proposals have never achieved a significant consensus within providers and standardization bodies. Multi-domain TE has been addressed by the Internet Engineering Task Force (IETF) and by the International Telecommunication Unit - Standardization Sector (ITU-T) through the deployment of two architectures respectively named Path Computation Element (PCE) Architecture [3] and Automatically Switched Optical Network (ASON) [4].

PCE architecture has been proposed to perform constraint-based path computations also in multi-domain networks. Two procedures named Per-Domain (PD, [5]) and Backward Recursive PCE-based Computation (BRPC, [6]) have been proposed. In particular, the BRPC procedure has been designed for computing optimal multi-domain paths given a specific metric. Per-domain and BRPC, implemented through the PCE communication Protocol (PCEP, [5]), require the pre-computation of the sequence of domains to be crossed by an inter-domain connection. Such pre-computation is paramount to guarantee effective path computations and to efficiently exploit the overall network resources.

Three methods have been considered to provide the PCE with multi-domain information for domain sequence pre-computation. The first method resorts to the information typically announced through the main instance of BGP. However, BGP does not advertise bandwidth information and multiple alternative routes, thus providing the same route indications (i.e., deterministic sequences of domains) which may rapidly induce link congestion. The second method, recently proposed in [7], requires an exchange of information through a service plane. However, this would imply a substantial effort for implementing and standardizing a new plane. The third method resorts to the information advertised through the External Network-to-Network Interface (E-NNI) [8] defined by the Optical Interworking Forum (OIF). OIF E-NNI Routing has been introduced to address the multi-domain single-carrier scenario. It provides a hierarchical implementation of the OSPF-based routing within the ASON architecture. However, the use of a link-state routing protocol (i.e., OSPF) in a multi-domain multi-carrier scenario may not represent the optimal solution in terms of confidentiality and policy-related issues. Confidentiality may be affected by the fact that all domains obtain the detailed view of the whole network resources, i.e. including detailed info on resources belonging to non-adjacent domains. A link-state routing protocol which advertises the same information to all domains may also be an obstacle to implement routing policies due to commercial constraints (e.g., peering preferences, SLAs, and economic cost, possibly associated to selected customers/prefixes).

In this study, we focus on the PCE architecture capability to provide multi-domain TE [9]. However, differently from

978-1-4244-6404-3/10/$26.00 ©2010 IEEE

the OIF E-NNI Routing solution, we propose to pre-compute the sequence of domains on the basis of the information provided by a hierarchical path-vector protocol dedicated to TE information. In particular, a hierarchical instance of BGP (HBGP) is proposed to operate within a restricted set of authorized domains. This is in-line with the PCE architecture applicability, which is defined to operate within a limited set of domains with known relationships, like the peering relationships (e.g., up to 20 domains [3]). In terms of confidentiality, it restricts the multi-domain information received by a node to the network view provided by the adjacent domains and not the detailed view of the whole network resources. In addition, commercial constraints could be better supported since transit domains have the capability to potentially treat each customer/prefix in a different way. Moreover, a path-vector view of the network resources is compatible with the view provided by the PCE architecture, which is based on path information exchanged between adjacent domains.

The integrated HBGP-PCE architecture enables the implementation of multiple schemes that differently exploit the various configuration features. They include: the announcement through HBGP of single or multiple routes per each domain or network prefix, the possible announcement of inter-domain bandwidth availability information, the use of Per-domain or BRPC path computation procedures, and the possibility to perform multiple attempts along different sequences of domains. In this work we show through simulations the performance of the HBGP-PCE architecture and related schemes under different network scenarios. In addition, we present a complete experimental implementation of the proposed HBGP-PCE architecture.

II. HBGP-PCE ARCHITECTURE

A. *Hierarchical BGP (HBGP)*

The proposed HBGP-PCE architecture resorts to an additional and separated instance of BGP called Hierarchical BGP (HBGP). HBGP independently operates among a restricted set of authorized domains. One Hierarchical Routing Controller (HRC) in each domain is responsible for the HBGP message communication. HBGP provides the set of information to determine the sequence of domains to traverse for the PCE-based path computations of QoS-guaranteed services. HBGP does not affect the operations of the main BGP instance, which is still used between network Border Routers (or BGP Route Reflectors) to provide reachability information for the whole internet and to determine the multi-domain routing for best-effort traffic. HBGP only influences new path computations (in particular those referring for example to QoS-guaranteed traffic that require constraint-based multi-domain path computation) and does not affect the network operations, the forwarding of Best Effort traffic or already established LSPs. In HBGP, two main attributes are considered per network prefix. The first (mandatory) attribute is the AS_PATH as defined in current BGP specifications. It specifies the sequence of ASes through which the announcement for the prefix has passed, i.e. the sequence of domains to reach the announced network prefix. The second (optional) attribute is called AS_PATH_BW. AS_PATH_BW is introduced to specify the amount of reservable bandwidth along the whole AS_PATH. Such bandwidth is computed by taking into account just the resources on the inter-domain links that peer domains agree to make available for remote (and authorized) ASes participating in the HBGP instance. The advertisement of such information does not affect confidentiality requirements since it does not disclose intra-domain information or private inter-domain agreements. In this study, due to scalability reasons, multiple inter-domain links between the same pair of adjacent domains are considered as a single link with an available bandwidth equal to the sum of all available link bandwidths.

Two main strategies can be identified to determine the set of HBGP information to announce per network prefix. In the first strategy, hereafter called Single Route (*SR*), just one route per network prefix is stored in the forwarding table and propagated between HBGP peers. Such route is the shortest in terms of number of traversed ASes (i.e., minimum AS_PATH length). In case of multiple routes with equal number of traversed ASes, tiebreaking rules are applied (e.g., first learned, smaller IP prefix, etc.). *SR* is the strategy typically adopted by default BGP implementations. In the second strategy, hereafter called Multiple Routes (*MR*), multiple routes per network prefix are stored in the forwarding table and propagated between HBGP peers. In this study, just multiple routes characterized by the same (shortest) AS_PATH length are considered (i.e., tiebreaking rules are not applied). On the one hand, the *MR* strategy potentially allows more effective TE solutions (e.g., load balancing between alternative routes). On the other hand, *MR* determines a larger amount of exchanged and stored information compared to *SR* strategy. In this study, all prefixes belonging to the same AS are treated in the same way, i.e. HBGP information refers to the whole set of prefixes belonging to the same domain.

B. *PCE-based path computation*

When a multi-domain request for a QoS-guaranteed service arrives, the information provided by HBGP is used at the source domain to determine the sequence of domains on which the PCE-based path computation is performed. According to the considered HBGP attributes and announcement strategies, three different solutions can be identified. The first solution, called Deterministic Sequence (*D*) exploits just one sequence of domains per network prefix. Solution *D* takes place when the *SR* strategy is applied and just the mandatory AS_PATH attribute is propagated. The second solution, called Random Sequence (*R*), takes place when the *MR* strategy is applied and just the mandatory AS_PATH attribute is propagated. In this case multiple possible sequences of domains may be available per network prefix. The *R* solution identifies through random selection the sequence of domains for the PCE-based path computation. The third solution, called Bandwidth-based Sequence (*BW*), takes place when the *MR* strategy is applied and both the mandatory AS_PATH and the optional AS_PATH_BW attributes are announced. The *BW* solution identifies the sequence of domains for PCE-based path computation by selecting, among the multiple possible sequences, the one that provides the larger amount of reservable bandwidth.

TABLE I
HBGP-PCE ARCHITECTURE SCHEMES

N.	Scheme	AS_PATH	AS_PATH_BW	Announcement strategy	Sequence selection	PCE-based procedure	Computation attempts
1	BGP-D-PD	Y	N	Single Route (SR)	Deterministic (D)	PD	Single
2	BGP-D-BRPC	Y	N	Single Route (SR)	Deterministic (D)	BRPC	Single
3	HBGP-R-PD	Y	N	Multiple Route (MR)	Random (R)	PD	Single
4	HBGP-R-BRPC	Y	N	Multiple Route (MR)	Random (R)	BRPC	Single
5	HBGP-BW-PD	Y	Y	Multiple Route (MR)	Bandwidth-based (BW)	PD	Single
6	HBGP-BW-BRPC	Y	Y	Multiple Route (MR)	Bandwidth-based (BW)	BRPC	Single
7	HBGP-R-PD-MA	Y	N	Multiple Route (MR)	Random (R)	PD	Multiple (MA)
8	HBGP-R-BRPC-MA	Y	N	Multiple Route (MR)	Random (R)	BRPC	Multiple (MA)
9	HBGP-BW-PD-MA	Y	Y	Multiple Route (MR)	Bandwidth-based (BW)	PD	Multiple (MA)
10	HBGP-BW-BRPC-MA	Y	Y	Multiple Route (MR)	Bandwidth-based (BW)	BRPC	Multiple (MA)

Once the sequence of domains is identified, the HBGP-PCE architecture performs the PCE-based path computation. Two standard procedures can be considered, the Per-Domain path computation (PD, [5]) and the Backward Recursive PCE-based computation (BRPC, [6]). If the PCE-based path computation fails, two cases can be considered. In the first case, the request is simply rejected. In the second case, further path computation attempts are performed (Multiple Attempts - MA). Each attempt requires the definition of a new sequence of domains (among those made available by the HBGP information) and the subsequent PCE-based path computation.

C. HBGP-PCE schemes

The HBGP-PCE architecture allows the definition of different schemes that can be derived by the combination of the aforementioned possibilities and solutions. The following schemes are here considered:

1. *BGP-D-PD*. A single attempt of Per-Domain path computation is applied along the single sequence of domains provided by BGP through the Deterministic solution (which considers just AS_PATH attributes and Single Route information per network prefix). BGP-D-PD can be implemented by exploiting just the current main BGP instance.

2. *BGP-D-BRPC*. As in *BGP-D-PD*, with the difference that BRPC is applied instead of PD.

3/4. *HBGP-R-PD/BRPC*. A single attempt of Per-Domain/BRPC path computation is applied along the randomly selected sequence of domains provided by BGP through the Random solution (which considers just AS_PATH attributes and the *MR* strategy). *HBGP-R-PD/BRPC* could be implemented by exploiting current BGP specifications and instances. However, since the use of the *MR* strategy may introduce scalability requirements, the adoption of a second hierarchical instance of (standard) BGP is recommended.

5/6. *HBGP-BW-PD/BRPC*. A single attempt of Per-Domain/BRPC path computation is applied along the least congested sequence of domains provided by HBGP information (Bandwidth-based solution, which considers both AS_PATH and AS_PATH_BW attributes and the *MR* strategy). *HBGP-BW-PD/BRPC* requires the hierarchical BGP instance implemented with the considered AS_PATH_BW attribute.

7-10. *HBGP-R/BW-PD/BRPC-MA*. The schemes described in 3-6 are here considered with multiple attempts of PCE-based path computations performed along different sequences of domains.

Table I summarizes the HBGP-PCE schemes together with the related adopted solutions.

III. HBGP-PCE ARCHITECTURE IMPLEMENTATION

The proposed HBGP-PCE architecture has been implemented in a network testbed composed of commercially available Label Switched Routers (LSRs) configured in N=5 ASes as shown in Fig. 1. One HRC and two PCEs are present in each AS. The HRC runs HBGP to exchange inter-domain route information to other HRCs belonging to adjacent domains. HRC builds and maintains the HBGP databases (i.e., HBGP-Adj-RIB-In, HBGP-Adj-RIB-Out, and HBGP-Loc-RIB) which are populated, according to the Single/Multiple Route announcement strategy, with either single or multiple routes per prefix. The first PCE is a functional element responsible for intra-domain path computations. It resorts to the Intra-domain Traffic Engineering Database (TED), e.g. the OSPF-TE TED. The second PCE, called Hierarchical PCE (H-PCE), is a functional element responsible for the computation of the sequence of domains to traverse for the multi-domain path computations. The H-PCE resorts to the HBGP-Loc-RIB database.

Packets H1-H5 are reported in Fig. 1 as an example of HBGP message exchange referred to prefix *R100* belonging to AS5. Message H1 is exchanged between HRC5 and HRC3. It encloses AS5 in the AS_PATH and it specifies the AS_PATH_BW attribute value as the available inter-domain bandwidth between AS5 and AS3. The same occurs to message H2 between HRC5 and HRC4 referred to the inter-domain link between AS5 and AS4. HRC3 (/HRC4) updates and propagates the H1 (/H2) information to HRC2 in message H3 (/H4). HRC3 (/HRC4) appends AS3 (/AS4) to the AS_PATH. In addition, it selects the minimum value between the bandwidth included in message H1 (/H2) and the available inter-domain bandwidth between AS3 (/AS4) and AS2. Then, HRC2 merges the received information in message H5 and sends it to HRC1. In particular, HRC2 appends AS2 to the two AS_PATH attributes and updates the two AS_PATH_BW attributes by enclosing the minimum value between the received AS_PATH_BW value and the bandwidth value on the inter-domain link between AS2 and AS1.

Fig. 1 also shows the PCEP Packets P1-P10 exchanged, upon connection request, to complete the multi-domain path computation. Message P1 is a PCEP PCReq message sent from a PCC, e.g. the AS1 Network Management System (NMS), to PCE1. The considered multi-domain PCEP request has as end-points source *R110* belonging to domain AS1 (IP: 110.110.110.110) and destination *R100* belonging to AS5 (IP: 100.100.100.100). PCE1 then behaves as PCC and forwards the P1 PCEP PCReq message to H-PCE1 (message P2). H-PCE1 resorts to the HBGP-Loc-RIB database to select the sequence of domains to traverse. If scheme *HBGP-BW-PC/BRPC(-MA)* is applied, two sequences are available: AS2-AS3-AS5 and AS2-AS4-AS5. They both have the same AS_PATH length, but they may present different AS_PATH_BW values. If both values indicate null reservable bandwidth, HPCE1 replies with a NO_PATH object. Such information is immediately forwarded to the PCC to communicate the failure in the multi-domain path computation. If at least one AS_PATH_BW value indicates a non-null value, HPCE1 selects the sequence of domains with the highest AS_PATH_BW value (e.g., AS2-AS4-AS5 in the experiment) and encloses it within the PCEP PCRep message P3 to PCE1. PCE1 then starts the standard Per-domain or BRPC path computation (messages P4-P9) along the provided sequence. The final ERO is then returned to the PCC through PCEP PCRep message P10.

In HBGP-PCE architecture implementation, the bandwidth information on the inter-domain links can be dynamically retrieved by HRC from either the OSPF-TE TED database or from the border LSRs. The former method requires the use of the Inter-AS-TE Link State Advertisement (LSA) recently proposed in [10]. In our experiment, the utilized LSRs do not support such extension. Thus, the latter method is adopted: a proprietary User-to-Network Interface (UNI) is used to periodically download the RSVP reservable bandwidth information in the form of a short XML file. For example, the file exchanged from border *R60* to HRC3, referred to a single inter-domain link, has a size of only 3 KB. The H5 message sent from HRC2 to HRC1 (referred to AS_PATH AS2-AS3-AS5) has a size of 57 bytes, as shown in Fig. 2. In the experiment, all the components have been internally developed. The HBGP protocol has been implemented in C++ exploiting the Socket TCP libraries. The PCEs and PCEP implementations are derived from the *SPaCE* PCE architecture implementation presented in [9]. Fig. 2 shows the overall message exchange when the *HBGP-BW-BRPC* scheme is applied. In particular, packet n.8 shows the H5 HBGP update message. Packets n. 28-66 include the PCEP messages running the BRPC-based path computation procedure. Packets n.34, 38, 40, and 54 correspond to messages P1-P4, while packet n.59 and 64 correspond to messages P9 and P10.

IV. SIMULATION RESULTS

The performance of the proposed HBGP-PCE architecture has been evaluated through simulations. The reference multi-domain network is a PAN-EUROPEAN topology composed of six ASes: GEANT2 and five National Research and Education

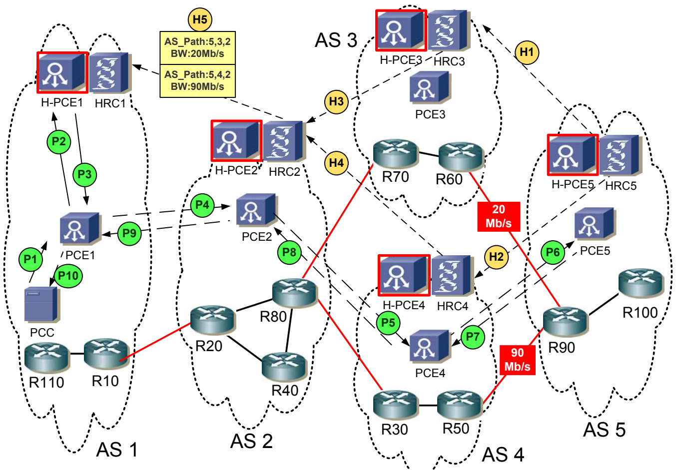
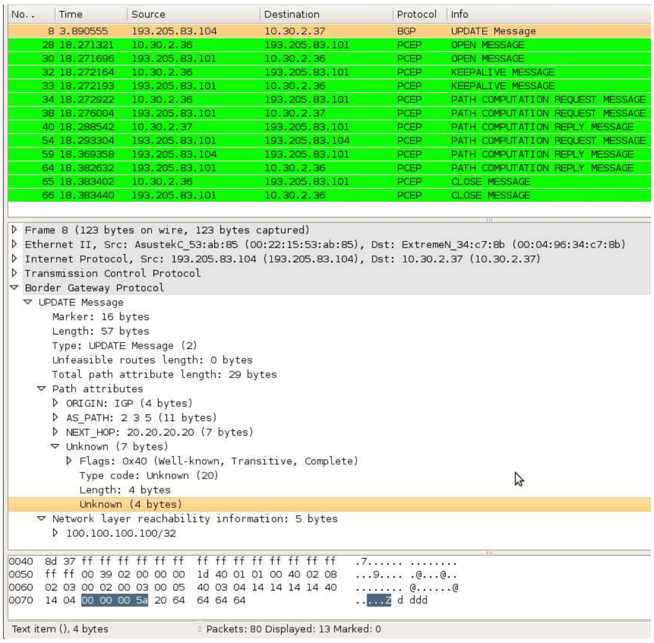


Fig. 1. HBGP-PCE architecture: experimental testbed.



```
No. .   Time       Source           Destination      Protocol  Info
      8 3.890555   193.205.83.104   10.30.2.37       BGP       UPDATE Message
     28 18.271321  10.30.2.36       193.205.83.101   PCEP      OPEN MESSAGE
     30 18.271696  193.205.83.101   10.30.2.36       PCEP      OPEN MESSAGE
     32 18.272164  10.30.2.36       193.205.83.101   PCEP      KEEPALIVE MESSAGE
     33 18.272193  193.205.83.101   10.30.2.36       PCEP      KEEPALIVE MESSAGE
     34 18.272922  10.30.2.36       193.205.83.101   PCEP      PATH COMPUTATION REQUEST MESSAGE
     38 18.276004  193.205.83.101   10.30.2.37       PCEP      PATH COMPUTATION REQUEST MESSAGE
     40 18.288542  10.30.2.37       193.205.83.101   PCEP      PATH COMPUTATION REPLY MESSAGE
     54 18.293304  193.205.83.101   193.205.83.104   PCEP      PATH COMPUTATION REQUEST MESSAGE
     59 18.369358  193.205.83.104   193.205.83.101   PCEP      PATH COMPUTATION REPLY MESSAGE
     64 18.382632  193.205.83.101   10.30.2.36       PCEP      PATH COMPUTATION REPLY MESSAGE
     65 18.383402  10.30.2.36       193.205.83.101   PCEP      CLOSE MESSAGE
     66 18.383440  193.205.83.101   10.30.2.36       PCEP      CLOSE MESSAGE

▷ Frame 8 (123 bytes on wire, 123 bytes captured)
▷ Ethernet II, Src: AsustekC_53:ab:85 (00:22:15:53:ab:85), Dst: ExtremeN_34:c7:8b (00:04:96:34:c7:8b)
▷ Internet Protocol, Src: 193.205.83.104 (193.205.83.104), Dst: 10.30.2.37 (10.30.2.37)
▷ Transmission Control Protocol
▽ Border Gateway Protocol
  ▽ UPDATE Message
      Marker: 16 bytes
      Length: 57 bytes
      Type: UPDATE Message (2)
      Unfeasible routes length: 0 bytes
      Total path attribute length: 29 bytes
    ▽ Path attributes
      ▷ ORIGIN: IGP (4 bytes)
      ▷ AS_PATH: 2 3 5 (11 bytes)
      ▷ NEXT_HOP: 20.20.20.20 (7 bytes)
      ▽ Unknown (7 bytes)
        ▷ Flags: 0x40 (Well-known, Transitive, Complete)
          Type code: Unknown (20)
          Length: 4 bytes
          Unknown (4 bytes)
    ▽ Network layer reachability information: 5 bytes
      ▷ 100.100.100.100/32

0040  8d 37 ff ff ff ff ff ff  ff ff ff ff ff ff ff ff   .7...... ........
0050  ff ff 00 39 02 00 00 00  1d 40 01 01 00 40 02 08   ...9.... .@...@..
0060  02 03 00 02 00 03 00 05  40 03 04 14 14 14 14 40   ........ @......@
0070  14 04 00 00 00 5a 20 64  64 64 64                  .....Z d ddd

Text item (), 4 bytes        Packets: 80 Displayed: 13 Marked: 0
```

Fig. 2. HBGP and PCEP message exchange.

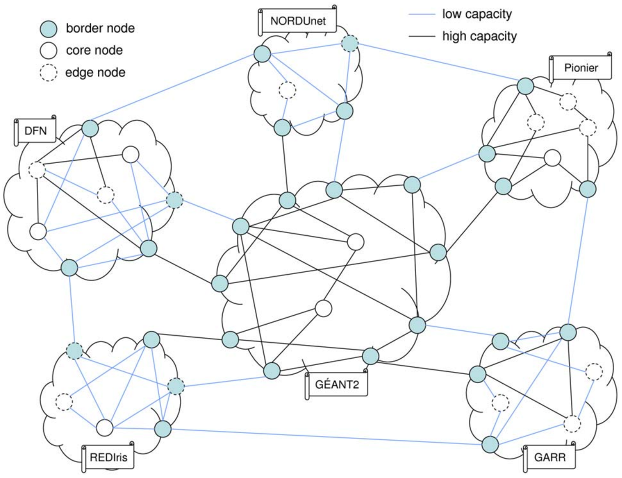


Fig. 3. PAN-EUROPEAN reference network.

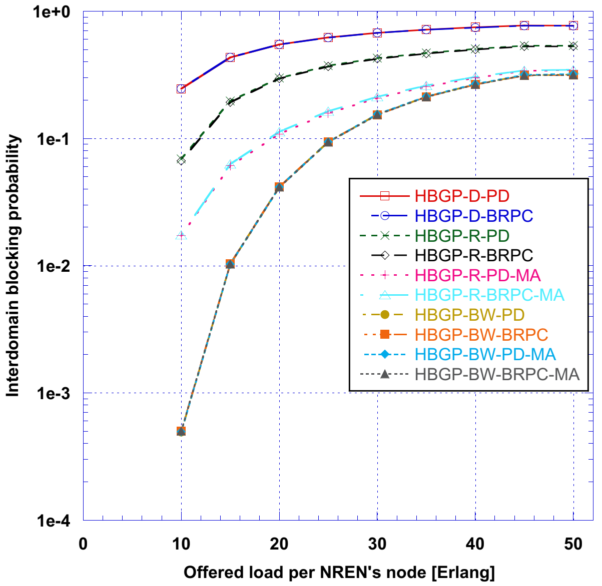


Fig. 4. Blocking probability vs offered load; first network scenario.

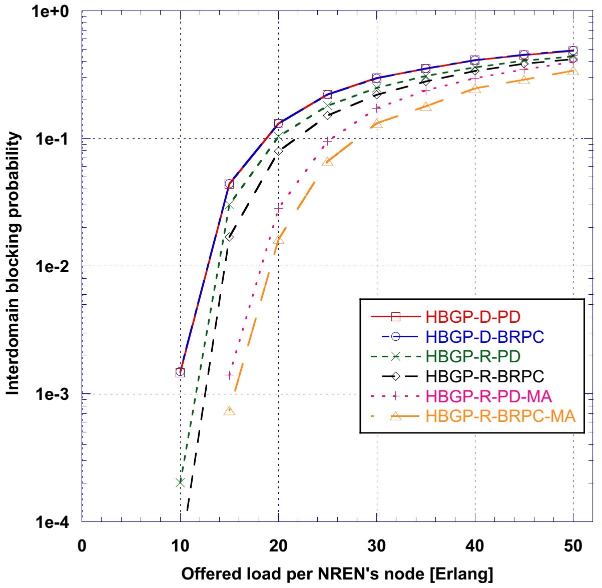


Fig. 5. Blocking probability vs offered load; second network scenario.

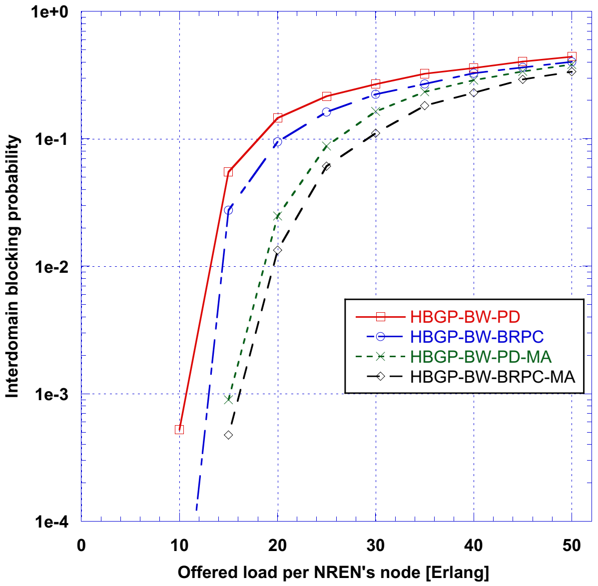


Fig. 6. Blocking probability vs offered load; second network scenario.

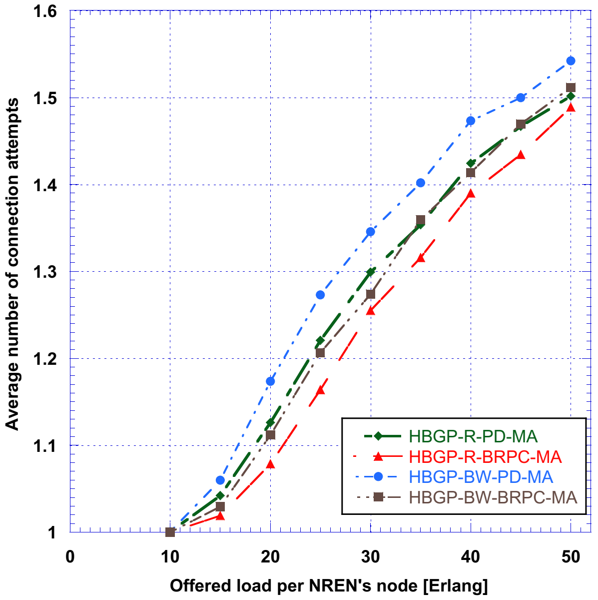


Fig. 7. Average number of path computation attempts vs offered load; second network scenario.

Networks (NRENs) [11]. Two network scenarios are considered. The first scenario, taken from [11] considers two different types of links: high capacity (*C1*=64 Gb/s) and low capacity (*C2*=16 Gb/s) (see figure 3). In the first scenario, connection requests of capacity *c*=1 Gb/s are dynamically generated with uniform distribution among all node pairs. Simulation results, carried out with an event-driven tool developed in our labs, have a confidence level of 95% and a confidence interval of 1%.

Fig. 4 shows the blocking probability of inter-domain connections (i.e., with source and destination belonging to different ASes) as a function of the offered load. *BGP-D* schemes yield the worst results: the deterministic selection of AS sequences determines high congestion along few inter-domain links while other resources remain underutilized. *HBGP-R* schemes significantly improve the blocking probability performance by randomly distributing the network load along multiple equal cost (in terms of AS_PATH length) sequences of domains. *HBGP-R-MA* schemes further improve the blocking probability performance thanks to the possibility to carry out multiple attempts along different sequences of domains. Bandwidth-aware schemes (i.e., *HBGP-BW-PD/BRPC(-MA)*) achieve the best performance: the AS_PATH_BW attribute allows to optimally balance the traffic along the sequences of domains (least fill routing is adopted). In this case, the *MA* option does not provide any significant improvement in blocking performance. The main reason is that, in the considered scenario, blocking is mainly determined by lack of available bandwidth just on inter-domain links. This also explains the reason that PD and BRPC achieve the same results in all the considered schemes. A second simulation scenario is then considered: intra-domain traffic load within GEANT2 has been increased of a factor of 4. In addition, all inter-domain links have been considered with high capacity.

Fig. 5 and 6 show the blocking probability performance of the HBGP-PCE architecture schemes in this second network scenario. Fig. 5 confirms that the worst blocking probability results are achieved by *BGP-D* schemes. *HBGP-R* schemes, in this scenario, only slightly improve the blocking probability performance. This is motivated by the fact that one of the possible sequences of domains is typically congested within the transit domain GEANT2. Thus, when the random choice selects such sequence, the blocking typically occurs. This also motivates the improvement achieved by *HBGP-PD/BRPC-MA* schemes which significantly lower the blocking probability. Fig. 6 show the performance of bandwidth-aware HBGP schemes. Results highlight that also in this case the adoption of multiple attempts significantly improve the blocking performance. The comparison between the results provided in Fig. 5 and Fig. 6 show that the least fill routing performed on the basis of the inter-domain link bandwidth availability does not allow to achieve good blocking probability results. For example, results show that *HBGP-BW-PD* scheme achieves similar performances compared to *BGP-D* schemes. In this scenario, however, the two considered PCEP-based computation scheme achieve different outcomes. In particular, BRPC-based solutions allow to improve the blocking probability performance in all the non-deterministic schemes.

The improvement provided by the schemes exploiting multiple attempts along different AS sequences require PCEs to repeat the path computation many times, which may significantly increase the overall connection set up time. In Fig. 7 we show the average number of attempts per connection request. As expected the average number of attempts per connection request is proportional to the offered traffic. Results show that the average number of attempts at reasonable loads does not exceed the value of 1.5.

V. CONCLUSIONS

A hierarchical BGP protocol instance, possibly enhanced with inter-domain link bandwidth availability attributes, has been proposed to provide the PCE architecture with additional information to be used in multi-domain path computations. Several schemes have been evaluated, considering the HBGP announcement of different set of information, either Per-domain or BRPC path computation procedures, single or multiple attempts along different sequences of domains.

Simulation results have shown that, under network scenarios where congestion is typically induced by lack of resources on inter-domain links, the proposed HBGP-PCE solutions, compared to solutions exploiting current BGP information only, significantly improve the overall blocking probability performance. In this case, the knowledge of inter-domain link bandwidth information guarantees the best performance. In addition, results highlight that the adoption of multiple attempts along different sequences of domains and the use of the BRPC procedure do not provide significant improvements. On the other hand, under network scenarios where congestion is typically induced by lack of intra-domain resources, schemes exploiting inter-domain bandwidth information do not provide significant advantages. In this case, however, the announcement of alternative equal cost routes provided by HBGP, the adoption of multiple attempts along different sequences of domains and the use of BRPC-based procedures allow to significantly improve the overall blocking probability performance compared to solutions exploiting just BGP information.

The proposed HBGP-PCE architecture has been implemented in a network testbed composed of commercially available routers configured in five ASes. The implementation successfully includes PCE prototypes and PCEP Per-domain and BRPC path computation procedures. In addition, the HBGP-PCE architecture testbed includes an implementation of the HBGP protocol instance which also encompasses the exchange of the introduced AS_PATH_BW attribute.

ACKNOWLEDGMENT

The work was partially supported by the BONE-project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Program.

REFERENCES

[1] X. Masip-Bruin, M. Yannuzzi, R. Serral-Gracia, J. Domingo-Pascual, J. Enriquez-Gabeiras, M. A. Callejo, M. Diaz, F. Racaru, G. Stea, E. Mingozzi, A. Beben, W. Burakowski, E. Monteiro, and L. Cordeiro, "The EuQoS system: a solution for QoS routing in heterogeneous networks," *Communications Magazine, IEEE*, vol. 45, no. 2, pp. 96–103, 2007.

[2] A. Manolova, S. Ruepp, J. Buron, and L. Dittmann, "On the efficiency of BGP-TE extensions for GMPLS multi-domain routing," *Optical Network Design and Modeling, 2009. ONDM 2009. International Conference on*, pp. 1–6, 2009.

[3] A. Farrel, J. P. Vasseur, and J. Ash, "A Path Computation Element (PCE)-based architecture," *Internet Engineering Task Force (IETF), RFC 4655*, Aug 2006.

[4] *G.Imp8080 : Implementers Guide for ITU-T Rec. G.8080/Y.1304 (05/2005) - Architecture for the automatically switched optical network (ASON).* International Communication Union ITU-T, 2007.

[5] J. Vasseur and J. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)," *Internet Engineering Task Force (IETF), RFC 5440*, Mar 2009.

[6] J. Vasseur, R. Zhang, N. Bitar, and J. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths," *Internet Engineering Task Force (IETF), RFC 5441*, Apr 2009.

[7] R. Douville, J.-L. L. Roux, J.-L. Rougier, and S. Secci, "A service plane over the PCE architecture for automatic multidomain connection-oriented services," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 94–102, 2008.

[8] *External Network-Network Interface (E-NNI) OSPF-based routing - 1.0 (Intra-Carrier) implementation agreement.* Optical Internetworking Forum, OIF, Jan 2007.

[9] F. Cugini, F. Paolucci, L. Valcarenghi, P. Castoldi, and A. Welin, "PCE communication protocol for resource advertisement in multi-domain BGP-based networks," *Optical Fiber Communication 2009. OFC 2009.*, pp. 1–3, 2009.

[10] M. Chen, R. Zhang, and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering," *Internet Engineering Task Force (IETF), RFC 5392*, Jan 2009.

[11] Multi-partner European test beds for research networking, "Preliminary definition of a reference architecture for an intelligent optical network supporting advanced application in research environments," *MUPBED, Deliverable D1.1*, 2007.