Routing Protocol Extension for Resilient GMPLS Multi-domain Networks

Anna Manolova, Sarah Ruepp
DTU Fotonik
Oersteds plads Building 343
2800 Kgs. Lyngby, Denmark
{anva,srru}@fotonik.dtu.dk

Ricardo Romeral, Sergio Rodriguez
Telematics Engineering Department
University Carlos III
Madrid, Spain
rromeral@it.uc3m.es

Abstract: This paper evaluates the performance of multi-domain networks under the Generalized Multi-Protocol Label Switching control framework in case of a single inter-domain link failure. We propose and evaluate a routing protocol extension for the Border Gateway Protocol, which allows domains to obtain two Autonomous System disjoint paths and use them efficiently under failure conditions. Three main applications for the protocol extension are illustrated: reducing traffic loss on existing connections by exploiting pre-selected backup paths derived with our proposal, applying multi-domain restoration as survivability mechanism in case of single link failure, and employing proper failure notification mechanisms for routing of future connection requests under routing protocol re-convergence. Via simulations we illustrate the benefits of utilizing the proposed routing protocol extension for networks employing different resilient mechanisms (both protection and restoration), as well as for networks which have not employed any resiliency technique. We show the need for differentiated failure handling for improving network performance under failure situations. Furthermore, we draw parallel between different network parameters and the efficiency of the applied notification and survivability strategies in the network.

Key–Words: multi-domain, GMPLS, BGP, AS-disjoint, performance enhancement

1 Introduction

Building the future Optical Internet will require using the entire potential of the optical networks, in particular the ability to automatically set-up and manage lightpaths (the so-called Labeled Switched Paths (LSPs)) under a framework such as the Generalized Multi-Protocol Switching (GMPLS) [1]. The ultimate goal is to perform automatic LSP establishment in a multi-domain context, where different domains (Autonomous Systems (ASes)) collaborate, allowing dynamic reservation of resources across domain boundaries. Two main challenges can be outlined: political and technical. When the political problems are solved through new inter-AS agreements and business models the technology must be ready to support the required dynamics and automation of the provisioning process.

Currently, there is no standard for inter-domain routing in GMPLS networks. Several proposals are being investigated among which is the current de facto standard for inter-domain routing in the Internet - BGP-4 [2]. In this context we can ask the question: is BGP, ready to be the next Optical Internet routing protocol? The routing requirements, taken into consideration when BGP was designed, and the requirements of the dynamic future Internet are very different. QoS support, reliability requirements and adequate support for a dynamic network environment have not been envisioned as main requirements of the routing protocol. In the future Internet though, these are paramount and necessitate extended information exchange between domains. Optical networks can transport a lot of information per second and failures heavily affect the performance of the network. Each domain should have enough information to adequately react to failures, but the current version of BGP does not provide such information.

In this paper we present an extension of the BGP protocol which allows for the computation of two AS-disjoint paths per destination. AS-disjoint paths are important for providing survivability in the multi-domain environment as well as for facilitating adequate reaction to changes in the network, which trigger BGP re-convergence. We illustrate the operation of the mechanism as well as the benefits of having two disjoint AS_PATHs for enhanced network performance under single inter-domain link failure. This work is extension of the work presented in [3].

The paper is organized as follows: Sec. II and Sec. III outline the problem and the related work in the area respectively. Sec. IV gives details on the pro-

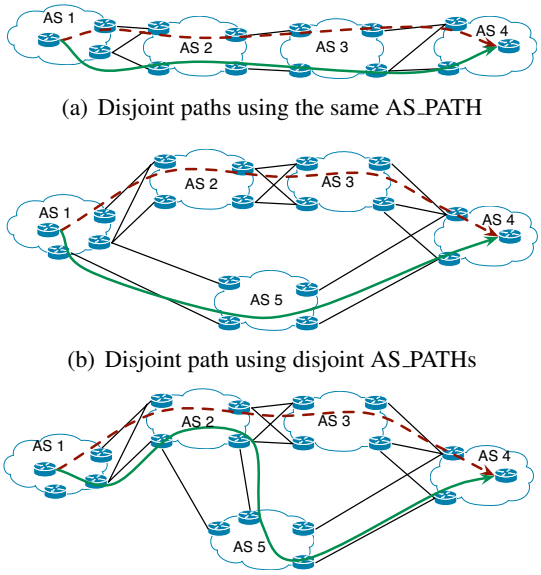
posed BGP extension. Sec. V focuses on the potential performance enhancements the extended BGP can offer to a multi-domain network. Sec. VI presents the simulation set-up and the obtained results. Conclusions are drawn in Sec. VII.

2 Disjoint path computation in multi-domain networks

The multi-domain disjoint path computation problem stems from the fact that no one in the multi-domain network has the complete network graph in order to run the Suurballe algorithm [4] for disjoint path computation. In some scenarios, especially multi-AS ones, it is not possible to obtain the complete graph of the network without flooding the network with sensitive information, which is unacceptable because of the strong privacy protection policies between the ASes.

We divide the methods for solving the multi-AS disjoint path computation problem in three categories (see Fig. 1). The first one uses standard BGP information or manual configuration to find the AS_PATH to a destination and tries to compute two disjoint paths along the obtained AS_PATH, i.e. it shares ASes. Solutions from this category necessitate sharing of information between the domains or employing novel protocols and/or new extensions to standard protocols as in PCE [5], PPRO [6] and ARO [7]. A drawback is that the applied optimization in these mechanisms can be done only within one AS_PATH, and this limits their efficiency. Furthermore, an AS failure or disconnection between two ASes on the path cannot be recovered using such approaches. The second category of solutions provides two AS-disjoint paths between the source and the destination. After that, two LSPs can be established via standard signalling protocols, e.g RSVP-TE, one along each AS_PATH. The third category provides partially AS-disjoint paths, i.e. only part of the paths share the same ASes, but this type of solution requires sharing of more information in order to obtain the solution and suffers from the same drawbacks as the solutions in the first category.

This paper proposes a solution within the second category. We design an algorithm for AS-disjoint path selection using BGP extensions. The mechanism provides two AS-disjoint policy-compliant paths between any two routers in a multi-domain network where ASes are multi-homed. Employing the proposed solution in connection-oriented networks does not exclude the usage of other advanced schemes for optimal path computation such as the PCE approach. BGP and PCE are completely interoperable since PCE elements need an AS_PATH in order to calculate an optimal



(a) Disjoint paths using the same AS_PATH

(b) Disjoint path using disjoint AS_PATHs

(c) Mixed solution

Figure 1: AS-disjoint path solution categories.

path for each LSP request. Furthermore, calculating two disjoint paths using different AS_PATHs reduces the complexity of the joint path computation process since the complex disjointness calculations need to be performed only for the source and the destination domains. We envision the BGP as a complementary routing protocol which provides a higher level path specified by AS numbers, whereas PCE can be used internally in each domain for optimal path computation.

3 Related work

Several proposals for BGP modifications for multipath dissemination can be found in the literature. Kushman et al. [8] have proposed R-BGP, a modification of BGP that allows to send to downstream neighbors alternative paths per destination. Two BGP peers exchange the standard AS_PATH and an alternative AS_PATH, called a *failover path*. Using this solution the number of lost packets decrease significantly during BGP re-convergence. There are two drawbacks though. First, the forwarding process in all involved routers must be changed, because a decision of which path (normal or failover) to be used is taken on a per packet basis. Second, in order to use the failover path an extra BGP session via the failover path with the neighbor must be established. There are several differences between R-BGP and our solution. First, R-BGP is for packet switched networks, whereas our proposal is for connection oriented networks. In GMPLS networks source routing is a funda-

mental feature, thus the head-ends of all connections require an extended view of the overall path, not only the next hop. In our solution the BGP speakers disseminate alternative AS_PATHs to all their neighbors, thus disseminating alternative AS_PATHs to all BGP speakers in the network, not only to the downstream neighbors. Second, the Kushman et al. [8] proposal supports protection only against link failures between neighboring domains. Our proposal is more general and offers survivability support for link failure, node failure and even entire AS failure.

Other proposals, as Bhatia et al. [9] or Walton et al. [10], try to eliminate the BGP route oscillations sending extra information in the BGP UPDATE messages. Walton propose to send several paths, not necessarily disjoint, for the same destination using a *Path Identifier* attribute. Bhatia propose to use *Multiple-Hop* Capability to report to a BGP peer more than one Next-Hop for the same reachable destination. The goal of these proposals is to reduce or eliminate the well-known BGP route oscillations. Our proposal, on the other hand, does not seek to eliminate route oscillation during BGP re-convergence, but rather to eliminate (or minimize) the effect of the osculations on the operation of the network. In this paper we illustrate how providing two AS-disjoint paths per destination can be used for survivability support in connection-oriented networks as well as for enhancing network performance under BGP re-convergence for future connection requests.

4 Obtaining disjoint AS_PATHs with BGP

BGP is a path vector protocol, which means that the created routing table contains the destination, the next hop towards the destination and the path to reach the destination. These are distributed via BGP UPDATE messages between BGP peers which contain: Network Layer Reachability Information NLRI (i.e. the destination), AS_PATH (sequence of ASes to be traversed on the way to the destination) and the next_hop. Interior routing information is not shared across domain borders via BGP, so different ASes have no interior information about other ASes. Limiting the shared information is done for scalability purposes as well as to avoid disclosure of sensitive information to other domains. Thus, the received BGP information by an AS is aggregated as much as possible and just one AS_PATH to a destination is chosen and further distributed to other BGP peers[1]. BGP peers choose paths according to a special decision procedure described in RFC 4271 [2]. In this paper paths chosen under the standard BGP operation are referred to as *primary AS_PATHs*.

Our proposed mechanism is a concurrent modified BGP decision procedure which obtains a disjoint AS_PATH to the primary AS_PATH, referred to as *secondary AS_PATH*. This secondary path can be used for resilience purposes, load balancing or routing of LSP requests during BGP protocol re-convergence.

The proposal necessitates three new Routing Information Bases (RIBs)[2]: Adj-RIB-Disj-In, Loc-RIB-Disj and Adj-RIB-Disj-Out (in practice a secondary route can be identified by a flag in the existing RIBs). The proposed extended BGP decision procedure constitutes of three phases as follows:

1. When a BGP entity receives an UPDATE message for a secondary AS_PATH from a peer, the route is added to the Adj-RIB-Disj-In and a preference is assigned. Upon a route addition or change in the Adj-RIB-In or Adj-RIB-Disj-In, phase two is triggered.

2. For the destination under update, select the best route disjoint to the one in the Loc-RIB from all available routes to that destination in all Adj-RIB-In and Adj-RIB-Disj-In. The selected route is included in the Loc-RIB-Disj; this RIB keeps all the BGP secondary routes used locally. If this implies a route change in Loc-RIB-Disj, apply phase 3.

   If no disjoint route exists this means there is trap in the topology of ASes[3] or the existing disjoint path is not policy-compliant. This can be solved by changing the primary path by manual configuration of the local preferences or by adjusting the local policies.

3. After a change in the Loc-RIB-Disj, the new route undergos a policy filtering process and is included in the selected Adj-RIB-Disj-Out; UPDATE messages are sent further.

Note that the received secondary routes are in Adj-RIB-Disj-In not in Adj-RIB-In and thus, they are not selectable as primary routes by the normal BGP decision procedure. This is a desirable behavior since secondary routes might create loops if the usual hop-by-hop routing of the standard BGP is used. Due to

[1]Note that aggregation of destinations is a common practice in BGP, in which case only one *AS_PATH* is distributed per aggregated destination.

[2]Please refer to [2] for standard BGP operation description and terminology.

[3]Theoretically in a 2-node-connected network there always exist 2 disjoint paths between nodes unless there is a trap in the topology.

the specifics of the proposed algorithm and the hop-by-hop routing paradigm, enforced by the BGP protocol, source routing is needed in order to use the secondary paths [11]. This is necessary because in some cases two neighboring ASes choose each other as next_hop for their secondary paths and in other cases they choose other neighboring ASes. This is topology dependent and cannot be predicted. Thus, in order to use the secondary path, the responsible border node must apply source routing for forwarding LSP requests on the secondary path. Possible solutions for source routing on the inter-AS level with BGP are proposed in [11] and [12].

Considering the operation of the proposed BGP enhancements the scalability of the protocol is not seriously harmed. The amount of stored data is at most twice the amount of data stored in BGP speakers under standard BGP operation since there are only two paths per destination.

On Fig. 2 an example of how the proposed BGP modification works is shown. Subfigure a) shows the RIB's content for destinations in AS 5 in the other ASes and the selected AS_PATH (marked with solid line orange background). These paths are obtained using the standard BGP process. As subfigure b) shows, all ASes which can select disjoint AS_PATH to the primary path among all available paths in their Adj-RIB-Ins do that (paths marked with dashed line purple background), and send an UPDATE with the disjoint AS_PATH information to their peers (solid purple lines). ASes which receive new disjoint AS_PATH information include it in the Adj-RIB-Disj-In and the selection mechanism is activated (subfigure c)). The new selected route is sent further until all ASes have a disjoint AS_PATH, just as shown in subfigure d).

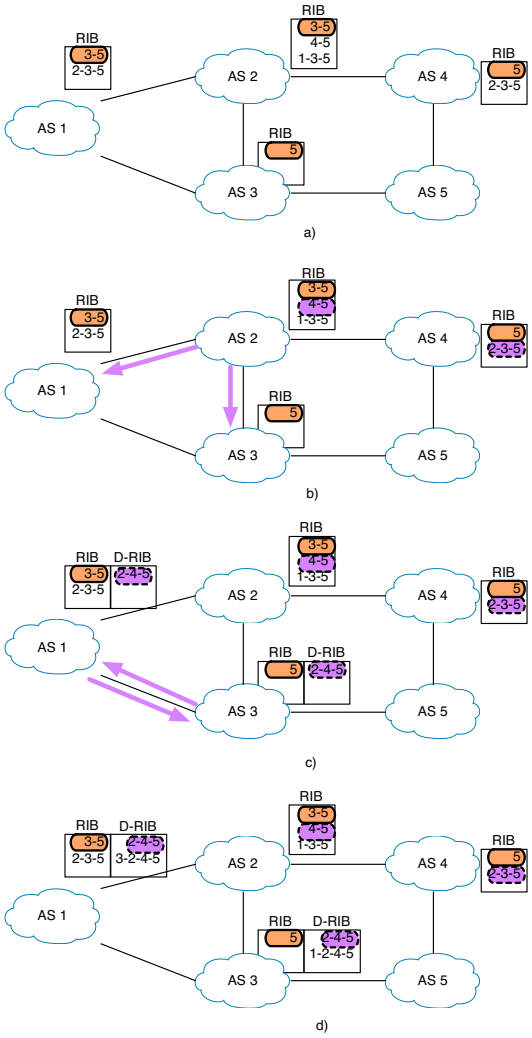


Figure 2: Modified BGP work example.

5 Network performance enhancement

Obtaining two disjoint paths per destination is beneficial not only for survivability in a dynamic multi-domain environment, but also for load balancing and network performance enhancement in case of failures when no resiliency mechanisms are applied. In our work we focus on three performance aspects. First we analyze the benefit of having two disjoint paths per destination with respect to the loss of traffic. Since BGP protocol re-convergence takes significant time [13], this results in high loss of traffic on existing connections and thus degraded network performance. Then, we focus on applying connection restoration for the affected LSPs. Utilizing the pre-selected disjoint paths we can restore the affected connections at the time of failure without being affected by the BGP re-convergence delay or route oscillations. The third aspect we analyze is the performance of the dynamic multi-domain network in terms of blocking of future connection request when no resiliency mechanisms have been applied. During the BGP re-convergence some nodes loose visibility of destinations or loops are created. This increases the LSP connection request blocking. Thus, the applied failure notification mechanism becomes paramount for proper network operation.

5.1 Failure recovery

There are two main approaches for providing resilience in a network - protection and restoration. Protection is the process of establishing a backup path, disjoint to the primary path, before the failure occurs. Restoration is the process of re-establishing an affec-

ted connection after failure using an alternative path. The total recovery time for the different approaches can be approximately given by the following equations [14]:

$$T_{recovery}^{Protection} = T_d + T_n + T_{sw}, \quad (1)$$

$$T_{recovery}^{Restoration} = T_d + T_n + T_{setup} + T_{sw}, \quad (2)$$

where T_d is the time to detect the failure, T_n is the time to notify the node, responsible for the failure recovery, T_{setup} is the time to set up the new LSP and T_{sw} is the time to switch over the traffic to the new established path.

Improving network performance under link failure means minimizing the time to recover from the failure. In case of protection, the time to recover is typically much shorter than in case of restoration because in the latter case the node initiating the backup path setup must compute an alternative path at time of failure. If no specific disjoint path computation mechanism is used then T_{setup} typically includes the BGP re-convergence time. If an AS-disjoint alternative path is available, the T_{setup} can be drastically reduced. Thus, in both cases (protection and restoration) the availability of AS-disjoint primary and secondary paths is clearly beneficial for resilience support in the multi-domain network.

A similar approach for path protection, based on pre-computed backup paths, is presented in previous works [15] and [16]. The authors of [15] use pre-computed paths for MPLS protection and focus on the delay parameter, whereas we focus on blocking probability under restoration. Moreover, our work is focused on dynamic failure recovery in multi-domain connection-oriented networks, whereas [15]and [16] focus on single domain cases and on pre-planned protection strategy.

5.2 Failure notification strategies

In case of a link failure it is paramount to inform the proper network elements in order to minimize the impact of the failure through proper failure notification. In a multi-domain scenario there is still no consensus whether a failure should be signalled all the way to the head-end of an affected connection or if it shall be handled locally. For single domain operation the head-end of the connection decides the protection method [17]. For multi-domain networks though it is not clear due to the diverse policies applied in the ASes and their capabilities for survivability support. In order to evaluate this we use the extensions of the BGP protocol proposed in this paper and we analyze the blocking ratio of connection requests after an inter-domain link failure using the following notification strategies:

- *No notification*: In this case the BGP protocol re-converges without notifying anybody of the failure. All LSP requests which cannot be routed due to lack of visibility or routing loops in this period are dropped.

- *Local notification*: In this case only the border nodes of the domains which detect the failure are notified. The border nodes then route the upcoming LSP requests using the secondary paths obtained by the BGP modification proposed in this paper. If a routing loop occurs (in case a domain uses its upstream neighbor for the backup path) the requests are dropped at the upstream node[4]. No BGP re-convergence is performed.

- *Head-end notification*: In this case the head-ends of the connections are notified that they must use their corresponding secondary paths, obtained using the proposed AS-disjoint BGP extensions. In this case no routing loops are possible and LSP blocking occurs only due to lack of resources. No BGP re-convergence is performed.

- *Mixed strategies*: Here the LSP requests are routed on the secondary paths during the BGP protocol re-convergence (using either the Head-end or the Local notification) and when the BGP protocol converges, the subsequent connection requests are routed on the new primary paths.

The actual failure notification can be performed in several ways, e.g. by the RSVP-TE Notify message or by extending the BGP Keep Alive messages. In our implementation we have employed the RSVP-TE Notify message, which is used to propagate a list of affected destinations in case of a link failure. The scope of propagation of the message depends on the applied notification procedure as described above.

6 Simulation results

The behavior of the extended BGP protocol was evaluated via two different simulation activities. First, a Quagga implementation of the BGP extension was used to validate the process of obtaining AS-disjoint paths and to evaluate the protocol overhead during BGP convergence. Then, the network performance under the outlined earlier application scenarios was

[4]Due to loop-detection mechanism within the RSVP-TE implementation.

evaluated via simulations with the event driven simulator tool OPNET [18]. We have evaluated the behavior of the modified protocol in two Pan-European topologies. For the Quagga implementation, each country is represented by one border node (see Fig. 4), whereas for the network performance evaluations we have used the COST 266 topology [19] (see Fig.**??**). Here the intra-domain topologies of the separate domains are randomly generated and have no more than 4 nodes acting as sourses/destinations. In total there are 46 source/destination nodes in 22 domains interconnected via 40 inter-domain links. The domain boundaries are assumed to be the geographical borders of the countries.

During the BGP path selection procedure the hop count is selected as a routing metric. No specific import/export policies are applied, i.e. it is assumed that all domains offer transit services to all their neighbors. The values for the Min_Route_Advertisement_Interval_Timers are set according to the specification in RFC 4271, i.e. 30 seconds for eBGP and 5 seconds for iBGP.

The two simulation activities are as follows. For the first one, we examine the entries in the Adj-Rib-In and Adj-Rib-Disj-In databases of all border nodes with respect to one specific destination and we observe the selection process and the backup path dissemination process. For the second activity we conduct several simulations in order to illustrate the potential benefits of using the proposed AS-disjoint BGP modifications. First, we illustrate the benefit for avoiding loss of traffic on established LSPs during BGP re-convergence. Then, we show the recovery success ratio of affected LSPs when two different restoration strategies are applied. Last, we focus on the effect of different failure notification strategies on the LSP blocking ratio for requests which occur during and after the link failure, i.e. we focus on the LSP rejection ratio. For these simulations the following settings are used. All connection requests have exponentially distributed duration with mean value of 600 seconds. The input load of the network is regulated by varying the mean value for the LSP interarrival time. Depending on the used amount of wavelengths per link we evaluate the behavior of the network at high, medium and low loads. Wavelength continuity constraint is assumed. For LSP signaling we use the RSVP-TE protocol. The wavelength assignment at the destination node is random among all free wavelengths along the path.

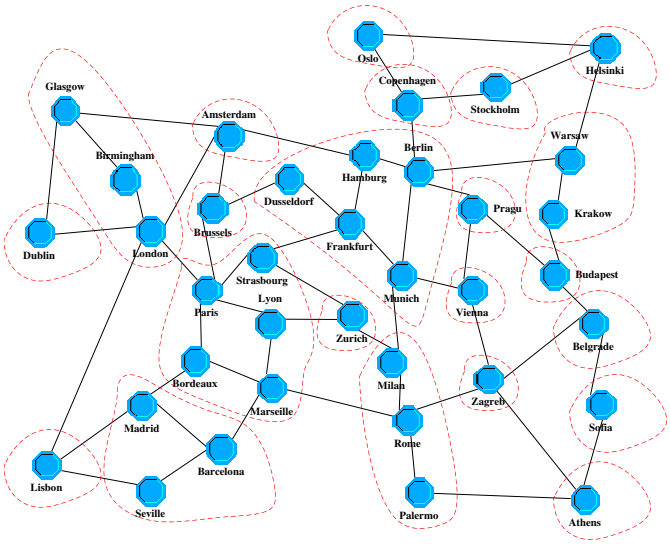


Figure 3: Cost 266 Pan-European topology.

6.1 Extended BGP protocol behavior evaluation

In this section the simulation results conducted with the Quagga implementation of the BGP extension are presented. The emulated network topology can be observed on Fig. 4. The target destination is domain 17. Fig. 5(a) shows the entries in the Adj-Rib-In database[5] in all nodes regarding destination 17 during the BGP convergence (i.e., during obtaining the primary paths). The chosen paths are indicated in oval forms. Fig. 5(b) illustrates the entries in the database when the AS-disjoint dissemination is active (secondary advertisements are indicated with asterisks) and the chosen secondary paths (indicated with dashed lines). Fig. 6 illustrates the BGP overhead and convergence time when normal BGP is used and when the AS-disjoint option is activated for the emulated network for destination - domain 17. It can be seen, that when the AS-disjoint option is activated the overhead in the network has increased with about 65%, whereas the time to converge the protocol has increased with about 50%, compared to the case when no extensions are used in the network. This is due to the fact that both processes (for working path and for as-disjoint path) are running in parallel. The results show that obtaining two AS-disjoint paths per destination does not cause excessive overhead in the network.

[5]One database is used for both primary and secondary advertisements, where the incoming advertisements regarding secondary paths are indicated in asterisks

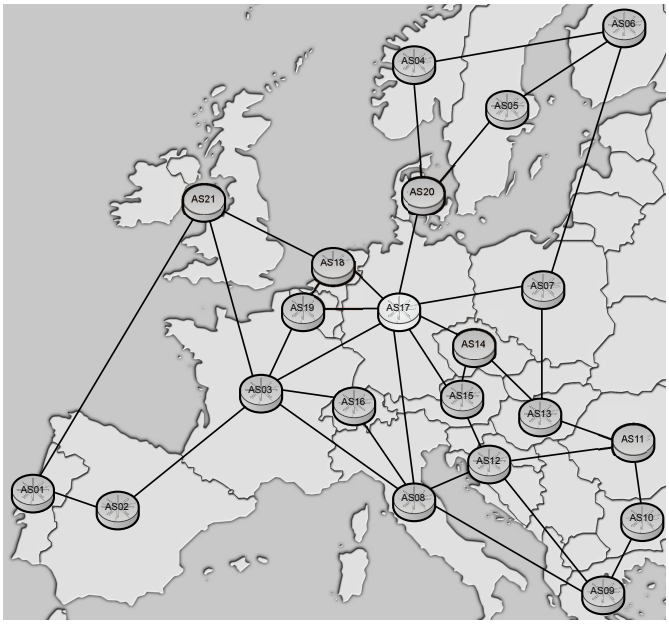


Figure 4: Tested Pan-European topology.

6.2 Traffic loss under BGP re-convergence

Here we focus is on the main goal of the BGP extension, namely saving traffic during BGP re-convergence. We investigate the performance under high network load (1 Erlang input load per node). Our performance metric is the BGP re-convergence time. Without having pre-established secondary paths during this period the source nodes cannot redirect the traffic from the primary LSPs on the secondary ones, thus the traffic on them will be lost. We evaluate the re-convergence time in case of failure of every inter-domain link in the network separately. The lost traffic under BGP re-convergence is proportional to the time to re-converge the BGP protocol, thus it is approximate value calculated as $N * T * C$, where N is the number of affected connections on the failed link, T is the BGP convergence time for that failure case and C is the capacity of the connections. Here we assume 10 Gbps connections.

On Fig. 7 it can be seen that for the failed links the convergence time varies between 5 seconds and 3 minutes. Considering the bit rate of the affected connections and the number of affected LSPs a link failure results in loss of approximately 0.5 to 115 Tb traffic due to path osculations and loss of visibility. Using our proposed mechanism for deriving AS-disjoint backup paths can significantly decrease the amount of lost traffic in the network since the source nodes do not have to wait for the BGP protocol to re-converge in order to obtain an alternative path.

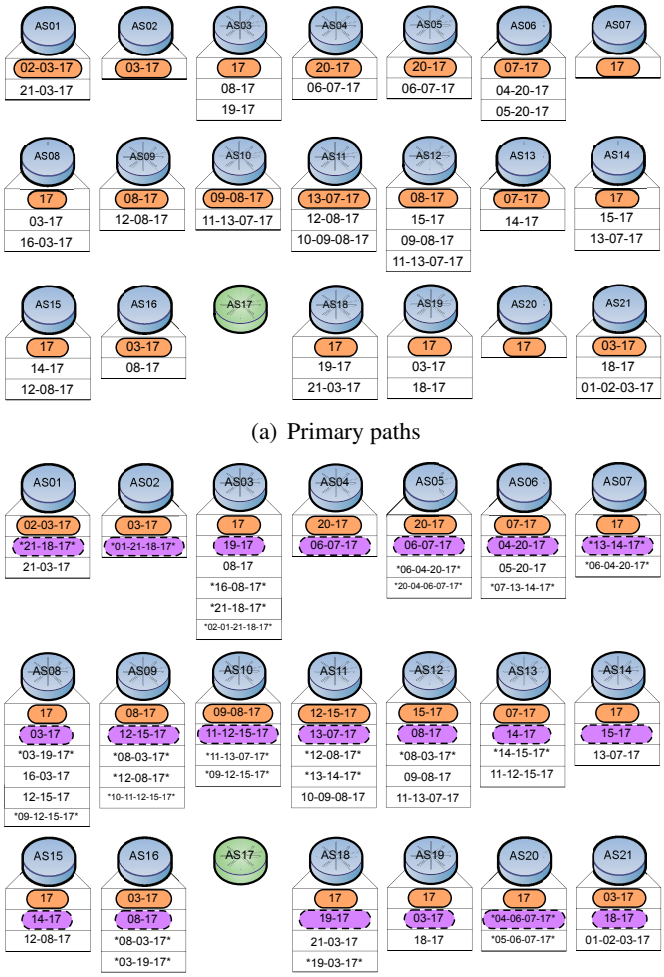


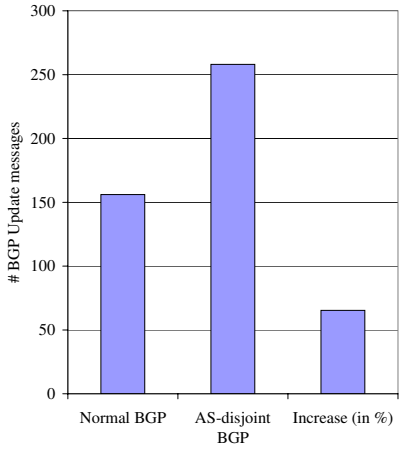
Figure 5: Adj-Rib-In and Adj-Rib-Disj-In entries and selected primary and secondary paths.

6.3 LSP restoration

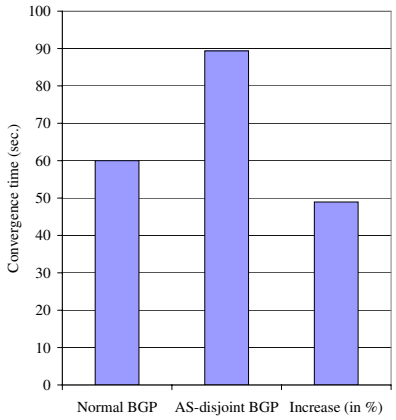
As we illustrated on Fig. 7, waiting for BGP to re-converge results in huge amount of lost traffic. Thus, using the proposed AS-disjoint BGP path dissemination we can re-establish the affected LSPs using the secondary paths. Since all routers in the network have two disjoint paths (if such exist and are policy compliant) two restoration approaches can be used - end-to-end (E2E) and local-to-egress (L2E)[6].

Fig. 8 illustrates the amount for saved traffic when applying restoration for different normalized input traffic loads at 50 wavelengths per link for two inter-domain link failures. In all cases the recovery process is in the order of 100 milliseconds, which is

[6]Local restoration is not an option unless there are parallel links between the border nodes, adjacent to the link failure. This is due to the fact that border routers have only visibility to reachable destinations and no visibility to other routers in neighboring domains.

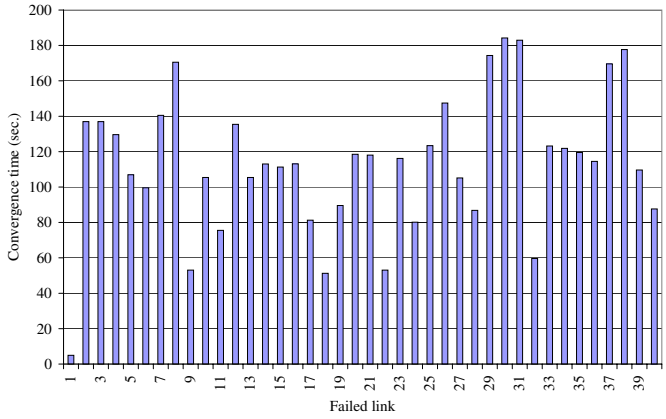


(a) BGP overhead

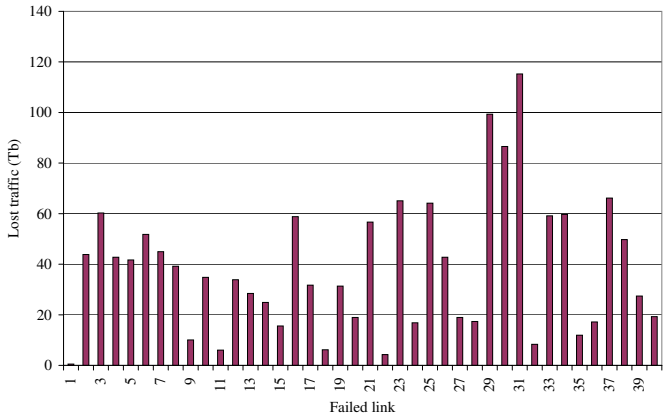


(b) BGP convergence time

Figure 6: BGP versus AS-disjoint BGP protocol evaluation.



(a) BGP convergence time after failure



(b) Lost traffic on affected connections

Figure 7: BGP convergence time and lost traffic on affected connections.

insignificantly small compared to the 3 minutes BGP re-convergence time. Furthermore, it can be observed that the efficiency of the recovery approach depends on the failed link. Applying our AS-disjoint path selection algorithm brings flexibility in the recovery process by facilitating differentiated failure handling, since all border routers have disjoint paths to all destinations. This brings advantage for operators which serve customers with diverse service requirements.

6.4 Failure notification analysis for future LSP requests

Here we evaluate the importance of proper failure notification method for reducing the blocking of LSP requests under BGP re-convergence. Our first case is a failure of the link with the most lost traffic from our previous experiment (i.e. link 31 which is Berlin - Warsaw). We consider the case of a medium and low load in the network. The results for the LSP blocking ratio using the different notification strategies are presented on Fig. 9.

As it can be seen, the LSP blocking ratio for the whole network is the highest for the *Local notification* strategy. This implies that the objective to preserve the failure information locally is not always the best choice. Applying the *Local notification* scheme may yield longer paths for the LSPs, which results in increased blocking probability. For the medium loaded network the remaining strategies perform almost equally good. This is due to the fact that under more loaded condition, the LSP blocking is dominated by the lack of resources. Thus, the difference between the schemes is difficult to observe. At the low loads though, the *Mixed strategy (Head-end)* performs the best. Under this scheme the LSP requests are routed from the Head-end on their secondary paths during the BGP re-convergence and on their new primary paths after the re-convergence. The achieved improvement compared to the *Local notification* strategy is about 50%.

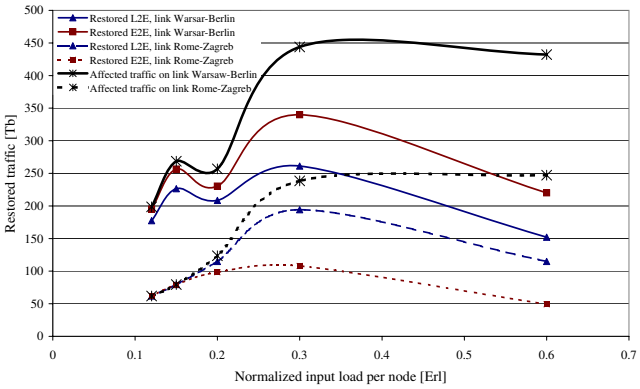


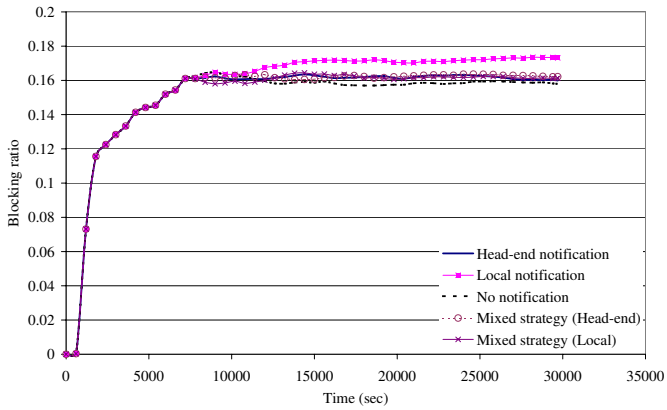
Figure 8: LSP restoration success ratio for two different link failures and two restoration strategies.

Fig. 10 illustrates the blocking ratio of two flows[7] in case of two different link failures for all tested notification mechanisms: flow England → Hungary with failed link Berlin - Warsaw and flow Poland → Greece with failed link Belgrade - Sofia. The mixed strategy is *Mixed strategy (Head-end)* and the network is under medium load condition. The goal is to see how different failures affect different individual flows and how the investigated notification schemes perform on a per flow basis.

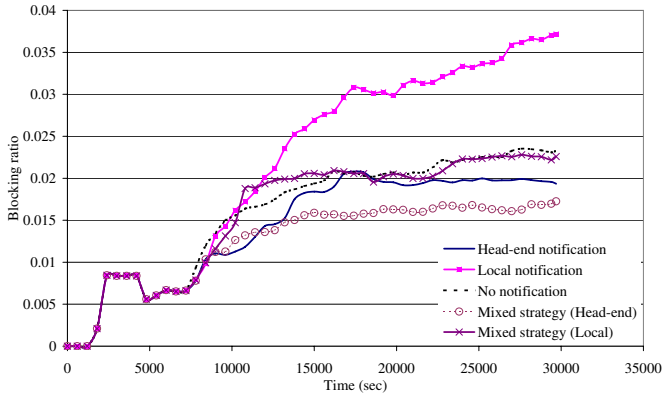
The first thing to be noticed is that different strategies affect the blocking ratio of flows differently. For the first flow (England → Hungary) informing the head-end of the connections (England in this case) brings significant LSP blocking improvement (around 50% better than the *Local notification*). For the second flow though, the *Local notification* yields the lowest blocking ratio (around 30% better than the *Head-end notification*). This calls for the development of schemes which handle affected flows in a differentiated manner.

The second interesting result is that for failed link Berlin - Warsaw the blocking of the observed flow under *Head-end notification* is lower after the failure than before the failure. Furthermore, the mixed strategy is performing worse than the *Head-end notification*. This is due to the fact that the obtained secondary path is better than both the old primary path and the new primary path obtained after protocol reconvergence, which yields lower blocking ratio. This implies that configuring the BGP routers of a certain domain taking only the bi-lateral agreements with the neighbors into account is not enough to obtain the best performance in a multi-domain environment. In the Future Optical Internet a global coordination is required in order to provide end-to-end QoS of connections crossing multiple domains.

[7]Here flow refers to set of connection requests between a fixed source/destination pair. Since the AS_PATH is the same for the same source/destination pair the LSP requests will follow the same set of ASes.



(a) Medium network load (17.25 Erl.)

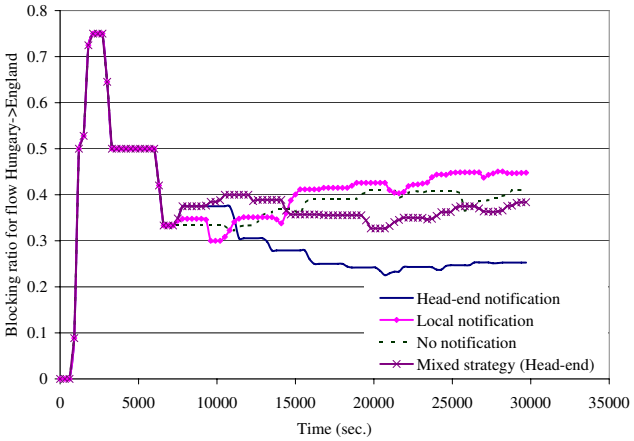


(b) Low network load (8.6 Erl.)

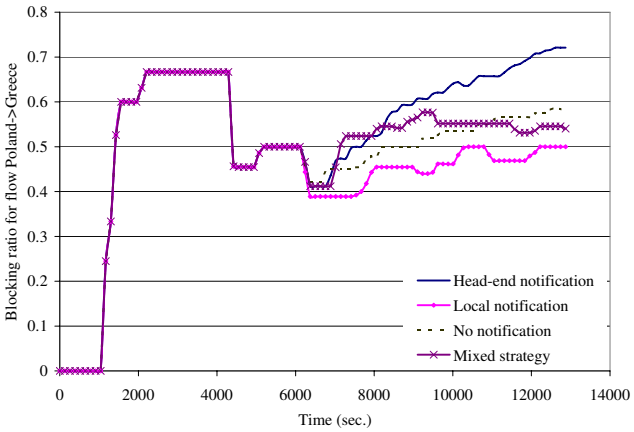
Figure 9: LSP blocking ratio for different notification strategies.

Fig. 11 presents the blocking ratios for different failure cases. The X-axis presents the failed link and its relative load calculated as the ratio between the affected LSPs at the time of failure and the total capacity of the link (30 wavelengths). The Y-axis presents the blocking ratio for one of the flows passing the particular failed link.

It is clearly visible that there is a relation between the load on the link and the efficiency of the applied notification strategy. The more loaded the failed link is the less effective the *Local notification* is. This is due to the fact that when a link is heavily loaded then re-directing all LSP requests on the same local backup path will saturate it faster, due to the presence of original traffic on that path. The *Head-end notification* on the other hand re-directs the affected flows from the head-ends of the connections and achieves in effect load balancing which decreases the blocking probability.



(a) Flow Hungary → England, failed link Berlin - Warsaw



(b) Flow Poland → Greece failed link Belgrade - Sofia

Figure 10: LSP blocking ratio for different notification strategies for two monitored flows.

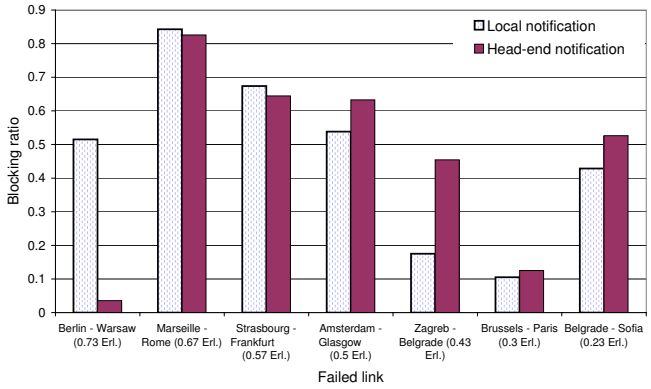


Figure 11: Blocking ratio of different flows for different failed links for two notification strategies.

7 Conclusions

In this paper we propose an extension of the BGP protocol for obtaining AS-disjoint paths in a multi-domain GMPLS network. We focus on the potential benefits of applying the proposed mechanism for improving network performance in case of inter-domain link failures. The conducted protocol evaluation analysis with a Quagga implementation reveal that the price for obtaining two disjoint paths per destination is not excessive. In fact, the generated overhead and the increase in the convergence time are about 50%. Simulation results, performed with an event driven simulator, illustrate that employing AS-disjoint paths for reestablishing affected LSP connections can potentially save huge amounts of traffic. With BGP reconvergence times within tens of minutes, this implies a lot of saved revenue.

Furthermore, we showed that deploying the correct failure notification strategy can considerably lower the blocking ratio of new LSP requests. Different failures affect LSP request flows in different way which calls for a differentiated approach for failure notification and failure recovery. Our results indicate that the more loaded the failed link is the less effective the *Local notification* strategy is. The presented results also imply that the position of the link failure, related to the head-end of the affected LSP request flow, as well as the actual position of the failed link within the multi-domain topology, should be taken into account when deciding the notification mechanism.

Applying the proposed BGP enhancement facilitates the operation of a highly dynamic and automatic multi-domain network, by providing flexibility for differentiated failure handling. It is an important step towards making BGP a viable solution for the Future Optical Internet under the GMPLS umbrella.

Acknowledgements: The work described in this paper was carried out with the support of the BONE-project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme.

References:

[1] E. Mannie et al. "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," RFC 3945, October 2004.

[2] Y. Rekhter and S. Hares. "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, January 2006.

[3] A. Manolova, S. Ruepp, R. Romeral. "Enhancing network performance under single link failure with AS-disjoint BGP extension," In Proc. 4th WSEAS International Conference on CIRCUITS, SYSTEMS, SIGNAL and TELECOMMUNICATIONS (CISST), January 2010.

[4] J.W. Suurballe. "Disjoint Paths in a Network," Networks, 4:125-145, June 1974.

[5] A. Farrel, J.-P. Vasseur, and J. Ash. "A Path Computation Element (PCE)- Based Architecture," RFC 4655 (Informational), August 2006.

[6] J.P. Lang, Y. Rekhter, and D. Papadimitriou. "RSVP-TE Extensions in sup- port of End-to-End Generalized Multi-Protocol Label Switching (GMPLS)- based Recovery," RFC 4872, May 2007.

[7] A. DÆAchille, M. Listanti, U. Monaco, F. Ricciato, D. Ali, and V. Sharma. "Diverse Inter-Region Path Setup/Establishment," draft-dachille-diverse-inter-region-path-setup-01.txt, October 2004.

[8] N. Kushman, S. Kandula, D. Katabi and B. Maggs. "R-BGP: Staying Connected in a Connected World," 4th USENIX Symposium on Networked Systems Design & Implementation, April 2007.

[9] M. Bhatia, J.M. Halpern and P. Jakma. "Advertising Multiple Next_Hop Routes in BGP," Internet Draft, August 2006, work in progress. http://tools.ietf.org/html/draft-bhatia-bgp-multiple-next-hops-01.txt.

[10] D. Walton, A. Retana, E. Chen and J. Scudder. "Advertisement of Multiple Paths in BGP," Internet Draft, July 2008, work in progress. http://tools.ietf.org/html/draft-walton-bgp-add-paths-06.txt.

[11] X. Li, Sarah Ruepp, Lars Dittmann, and Anna V. Manolova. "Survivability-Enhancing Routing Scheme for Multi-Domain Networks," GLOBECOM 2008: 2232-2236.

[12] A. Manolova, S. Ruepp, J. Buron, and L. Dittmann. "On the Efficiency of BGP-TE Extensions for GMPLS Multi-Domain Routing," 13th ONDM conference, February 2009.

[13] M. Tuba. "Computer Network Routing Based on Imprecise Routing Tables," WSEAS TRANSACTIONS on COMMUNICATIONS, Issue 4, Volume 8, April 2009.

[14] J. P. Vasseur, M. Pickavet, and P. Demeester. "Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS" Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[15] T. Xiansi, Ch. Jingwen, M. Yajie, O. Liang, Y. Zongkai. "An Analytical Model and A Fast Mechanism for Fault Restoration with QoS Constraints in MPLS Networks Based on n:m Protection," WSEAS TRANSACTIONS on COMPUTERS Volume 7, 2008.

[16] H. Hwang, S. Ahn, Y. Yoo, C. Kim. "Configuration of Shared Backup Cycles for Local Restoration in ATM Mesh Networks," WSES / IEEE SSIP '01, MIV '01, SIM '01, RODLICS '01 International Conferences, September, 2001.

[17] P. Pan, G. Swallow, A. Atlas. "Fast Reroute Extensions to RSVP-TE for LSP Tunnels," RFC 4090, May 2005.

[18] OPNET Modeler, http://www.opnet.com

[19] R. Inkret et al. "Advanced Infrastructure for Photonik Networks," Extended final report of COST Action 266, available at http://www.ufe.cz/dpt240/cost266/.