A Cooperative Approach for Coordinated Inter-domain QoSR Decisions

Alexandre Fonte*‡, Edmundo Monteiro*, Marcelo Yannuzzi†, Xavier Masip-Bruin†, Jordi Domingo-Pascual†
*University of Coimbra
Laboratory of Communications and Telematics
CISUC/DEI, Polo II, Pinhal de Marrocos, 3030-290, Coimbra, Portugal
Email: {afonte,edmundo}@dei.uc.pt
†Universitat Politecnica de Catalunya, Departament d'Arquitectura de Computadors
Avgda. Victor Balaguer, s/n-08800 Vilanova i la Geltru, Barcelona, Catalunya, Spain
Email: {yannuzzi,xmasip,jordid}@ac.upc.edu
‡Polytechnic Institute of Castelo Branco
Av. Pedro Alvares Cabral, n12, 6000-084, Castelo Branco, Portugal

***Abstract*— Currently, most of the automated Traffic Engineering (TE) techniques with BGP (Border Gateway Protocol) rely on selfish routing approaches. The limitation of a selfish routing scheme is that it is unable to anticipate the performance effect on downstream domains due to uncoordinated routing decisions. This paper is concerned with Inter-domain Quality of Service Routing (QoSR) issues. It presents and discusses an approach for coordinated Inter-domain QoSR decisions to protect domains against SLS (Service Level Specification) violations and undesired effects on downstream domains.**

I. Introduction

Nowadays the Internet is largely operated by commercial providers. Part of inter-domain traffic exchanges are governed by SLSs (Service Level Specification) negotiated between providers and customers. Customers claim for robust SLSs to satisfy their traffic QoS (Quality of Service) demands. An emergent approach to facilitate the provision of robust SLSs and to protect traffic aggregates against QoS degradations or even to optimize internal domain's metrics (e.g. cost) is to develop intelligent TE (Traffic Engineering) mechanisms coupled with BGP (Border Gateway Protocol) [1]. This mechanisms can be located at specialized Overlay Entities (OEs) in the managing of inter-domain traffic exchanges [2]. To achieve this goal and to influence the BGP route selection process, the OEs must be able to tweak the BGP attributes of routes on-the-fly.

The benefits of OE based approaches operating on short timescales, are evident and so they are also being developed as commercial products [3][4]. However, these approaches have been designed as selfish routing schemes and are only able to support outbound traffic control. While they allow a domain to select the most cost-effective routes (or providers) by tweaking their local-preferences BGP attributes, the routes are greedily selected. This means that the routing mechanisms are unable to anticipate the performance impact of their choices on downstream domains. Hence, network overload episodes or resources policy conflicts can occur. In addition routing instabilities can frequently be experienced, which can have global impact, and thus, large parts of the Internet routing infrastructure can be overloaded [6]. Furthermore, the available techniques to downstream domains that can provide support on inbound traffic control are limited to BGP in-band techniques (e.g. BGP multi-exit discriminator attributes tweaking). Despite this possibility, these mechanisms require external BGP updates and support from upstream domains, and above all, they operate on a large timescale of several minutes. Moreover, some studies have documented problems arising in the use of these mechanisms [7].

To address the above issues, this paper presents an out-of-band cooperative approach to operate on a very short timescale, able to support coordinated inter-domain QoSR routing decisions. The out-band approach is well-suited to provide predictability in inter-domain traffic exchanges to avoid performance degradations and routing instabilities.

Remaining sections are organized as follows. In Sect. II, a brief analysis of the related work is given. Then, in Sect. III, we introduce and describe our cooperative approach, including the basic concepts and a proposal of a step-by-step coordination mechanism. In Sect. IV, we present and discuss some implementation considerations. Finally, we conclude the paper in Sect. V and discuss some directions for future (already ongoing) work.

II. Related Work

In previous work concerning intelligent routing schemes willing to control inter-domain traffic transferences, two sets of TE mechanisms coupled with BGP have been proposed. The first set includes proposals of on-line and off-line techniques for BGP route optimization. Among these proposals only few papers deal with the design of algorithms for multi-objective (i.e. performance and cost) route optimization [8][9]. The second set of mechanisms were designed to meet another central TE issue that is the attainment of smooth traffic distributions on egress links. In [10] this problem is addressed by using an evolutionary TE algorithm. One important aspect common to the cited works is that the proposed mechanisms operate

on relatively large timescales of a few minutes to reduce the risk of requiring to tweak a large number of BGP routes and to avoid the corresponding BGP updates storms. This means that these techniques are not able to handle real-time metrics (e.g. in order of magnitude of the Round-Trip Time). In parallel to scientific work, commercial products also started to appear. These products are globally known as smart (or optimized) edge routing and they operate on short timescales [4][5]. In general commercial solutions try to select the most cost-effective routes or providers. In contrast with previous tools, their internal details are unknown. For instance, the effective performance improvements and the impact on BGP performance is not clear. Another important common aspect is that these techniques behave as selfish routing schemes.

Our work differs from the above described techniques in the following aspects 1) we are proposing a cooperative approach as part of (but not limited to) an overall solution to address inter-domain QoSR issues; 2) our approach is able to handle SLS violations, automatically reconfiguring the BGP routers parameters; 3) our approach takes into account the preferences of the downstream domains regarding to the admission of new traffic.

III. Cooperative Approach

As stated previously our inter-domain QoSR framework is based on the distributed OEs and BGP approach. A key idea is that the peering OEs belonging to remote domains, cooperate in a reflective (mirroring) manner. This cooperative mirroring scheme allows the OEs sharing QoS measurement data to dynamically manage its outgoing and incoming traffic among the available paths to the target domain. Although, this cooperative approach allows incremental developments, because it does not require cooperation with intermediate domains and most of the introduced complexity is located in the edge domains. The main shortcoming of this approach is its inability to anticipate the impact of a new traffic patterns in intermediate domains. To address this issue, we present in this section an extension of this cooperative approach able to perform coordinated routing decisions.

A. Basic Concepts

Our cooperative approach for coordinated routing decisions is conceptually simple. Figure 1 outlines the approach. Once an ingress domain's OE (e.g. the OE_1) detects a SLS violation (or an imminent SLS violation), it obtains the available alternative QoS paths to the egress domain constrained to QoS demands of the traffic aggregate affected. Consequent on, from the OEs members of each alternate path, it gathers their domain routing preferences regarding to the admission of this traffic aggregate. After an ingress OE obtains these values, it aggregates them using a given aggregation function. Then, the OE's path selection process selects among the alternate paths the social optimum path (i.e. the most favorable path to all member domains, including the ingress domain). Finally, the ingress OE sets up the selected path. This process is

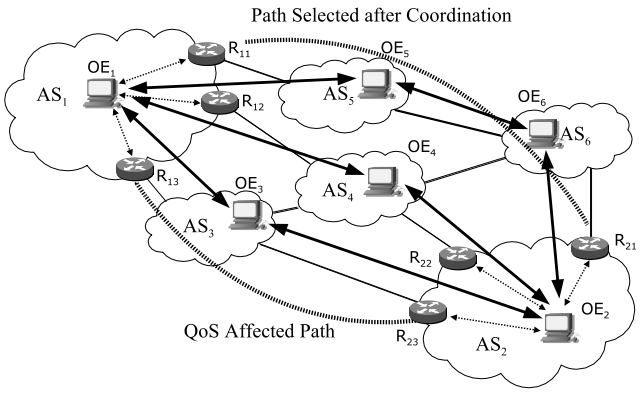

Fig. 1. Illustration of the Proposed Approach

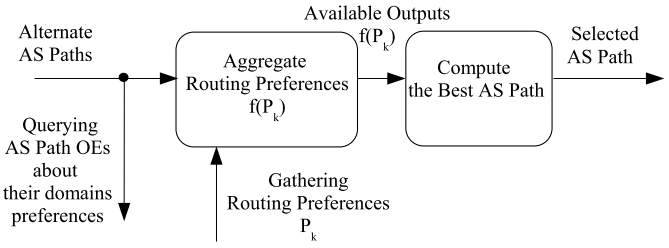

Fig. 2. Model of the Coordination Mechanism Problem

continuously repeated, step-by-step, for each traffic aggregate affected.

We have supported our approach by using a step-by-step pathway mechanism for coordinated routing decisions. In the context of this paper, this mechanism is understood as a routing mechanism part of our overlay layer, which solves the problem of finding a social optimum path. There are some basic concepts concerning the overall design of our mechanism. Figure 2 outlines a useful model to give a help on the mechanism's understanding.

Domain Routing Preferences: A fundamental rationale is that the OEs members of each path within the set of alternative paths A, report their domain routing preferences. A domain routing preference is similar to a cost p_k. It reflects the effort (or preference) on the admission of a new traffic aggregate from a domain's perspective. With the introduction of this concept we supply a means to intermediate domains and the egress domain to control incoming traffic by creating preferred routes and influencing the path selection process (and therefore the performed traffic changes) at the ingress domains.

Coordination Metric: We introduce the concept of coordination metric to enable routing domains to express their routing preferences. The coordination metric values are computed by a function $M()$. Thus, a given preference is computed as $p_k = M(v_k)$, where v_k represents a given vector of internal

information of the intermediate domain k being mapped into the cost p_k. For example, a given implementation of the coordination mechanism could use a $M()$ function which combines a vector v_k of policy flags, traffic costs and IGP (Interior Gateway Protocol) QoS path costs. Further more, the coordination metric is an essential mechanism to handle the business constraints on the disclosure of internal information (e.g. internal configurations or policy details). Thus, it is required the use of functions $M()$, which the problem of finding its inverse $M()^{-1}$, is computationally unfeasible.

Preference Aggregation and Feasible Outputs: At an ingress domain, an OE considers that the gathering operation of the domains' routing preferences along a certain path is well-suceeded, when every preference from all path's domains members are gathered. The result of all well-suceeded gathering operations is a subset of alternate paths $R \subset A$. Consecutively, all collected preferences $P_m = (p_1, ..., p_k, ..., p_n)$, where p_k for $k = 1, ..., n-1$ are the intermediate domains routing preferences and p_n is the egress domain routing preference, are aggregated using an aggregation function $f()$. In case of additive routing preferences, the ζ_m value of a path, defined as $\zeta_m = f(P_m)$, is equal to the sum of the corresponding routing preference values along that path. For nonadditive routing preferences, the ζ_m value of a path can be the minimum (or maximum) routing preference along that path. In our model, the subset R (and the associated ζ_m values and end-to-end QoS metric measures) is the set of feasible outputs of the coordination mechanism.

Optimization Problem: The definition of the objectives on the traffic exchanged between the ingress domain and downstream domains has a direct impact on the path selection. Given a set of feasible outputs R of the coordination mechanism, an output R_m (i.e. a path) which its associated ζ_m and end-to-end QoS metric measures values are the best at same time is required. On the one hand, optimizing both objectives can lead to a conflict situation. However, optimizing these objectives as a single combined objective is not useful for our purposes. Recall that our goal is to select the best path constrained to QoS traffic demands, which at same time it takes into account the preferences of the downstream domains regarding to the admission of new traffic. A computational efficient optimization algorithm should be integrated into the coordination mechanism to solve this problem.

B. Interactions between Overlay Entities

Rather than assuming an identical arrangement of the routing domains independently of the existing inter-relationships (like BGP does), with our out-band approach it is possible to adopt a different strategy to increase the stability and scalability of our inter-domain QoSR model. We propose two-levels of interactions between the OEs. In the first level OEs perform coordinated path changes to reallocate the traffic aggregates affected by QoS degradations to an alternative path. In contrast, the second level the OEs perform coordinated link changes to reallocate these traffic aggregates to a different link

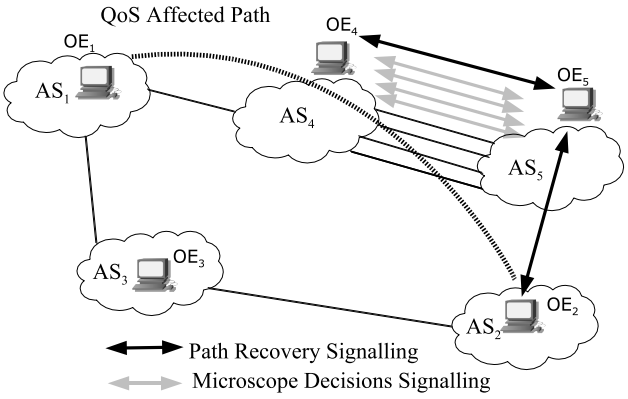

Fig. 3. Overlay Entities Recovering a QoS affected Path

shared by a pair of domains. Figure 3 illustrates an example of these interactions.

The clustering for the first level interactions between OEs is obtained by the identifications of domains, which have previous agreements to perform coordinated path change decisions. One the other hand, the clustering for the second level interactions between OEs is obtained by the identifications of pairs of dense domains, which have previous agreements to perform coordinated link change decisions. This last arrangement is motivated by recent tomography studies showing that transit domains are even more dense, and that these domains normally share multiple links [11].

When coupling both levels of interactions, instead of an OE immediately changing an affected traffic aggregate to an alternate path, it explicitly spawns a QoS degradation warning message on the current path asking the pairs of OEs with second level relations to seek for alternative links, which are able to improve the current offered QoS. In short, this feature has the advantage of keeping the current path in the case of a successful recovery operation pre-required, avoiding thus unnecessary path shifts and the corresponding BGP updates. This process is supported by the BGP path concept to be agnostic of any detail about domain interconnections.

C. Coordinated Link Changes

The replacement of current links carrying traffic aggregates affected by QoS degradations, in the coordinated link change case, the ingress domain's OE gathers the alternative routing options, and infers their corresponding link interfaces. The gathering and aggregation of link preferences is done by a similar process as in the path change case. Since, as depicted in Fig. 4, among the options, one which implies the use of a different egress point in the egress domain, could be selected as a solution. In the worst case scenario, this could result in the congestion of its downstream domains, and even on a subsequent cascade of coordinated link decisions processes before convergence. To avoid these effects and also to alleviate BGP, the egress domains OE should influence the decision

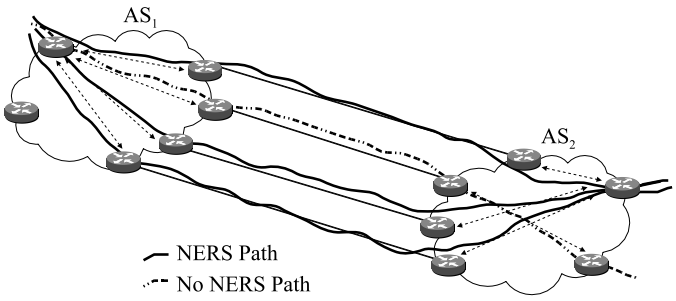

Fig. 4. Illustration of NERS paths

process of its peer to not select those links. This could be done by artificially increasing the coordination metric values. However, a more efficient but also restricted solution (similar to the use of a drain plug) is to apply an additional condition to the alternative links set. This could be done by filtering this set to obtain a Null Effect Route Subset (NERS) composed only by routes which keep the current egress point. Only the NERS routing preferences are reported.

D. Mechanism Proposal for Coordinated Routing Decisions

Figure 5 shows the algorithm of the mechanism for coordinated routing decisions. The proposed mechanism was designed as a step-by-step pathway toward the replacement of routes exhibiting QoS performance degradations. The modularity and independency of internal details about coordination metric computations, aggregation functions, traffic objectives, and optimization algorithms to select the social optimum path and finally signalling, which are part of our mechanism, were considered as fundamental requirements. Therefore, it has main focus on the flow of OEs functions execution and on interactions between OEs, and themselves and BGP routers. Once the ingress domain's OE detects an SLS violation event, according to our model, a new routing cycle starts. After being computed and established a social optimum path this cycle is finished. Because, it could be reasonable to admit that none social optimum path can be determinate. The mechanism assumes that OEs can perform a standard greedy route selection as last chance to improve the path's QoS.

For sake of simplicity, intentionally it was assumed implicitly that every participant OE in the coordinated routing decision process is honest when reports its routing preferences. However, typically domains are being managed by commercial organizations with their own optimization goals, which can lead them to a bad end, and thus to lie about the reported routing preferences. One possible way to minimize this problem is to extend the coordination mechanism to provide an incentive solution, including the corresponding rewards computation and distribution, that encourage each OE's domain to give truthful routing preference reports.

```
1. Begin {Start Running the mechanism}
2.   SLA Violation detected in Path to destination dst;
3.   Recovery=Path Recovery Process (Path);
4.   if(Recovery==unsuceeded)
5.      AlternatePaths=Alternate Paths Computation (dst);
6.      if!(AlternatePaths=={})
7.         NewPath=Preferred Path Computation (AlternatePaths);
8.         if(NewPath=={})
9.            Run Selfish Decision Process (AlternatePaths);
10.        Else Set-up (NewPath);
11.        Endif;
12.     Endif;
13.  Endif
14. End. {Stop Running the mechanism}

1. Path Recovery Process (Path);
2. Begin
3.  Send QoS Degradation Warning Message (Path);
4.  Upon receiving message (type,Path) ;
5.  If(type==WarningReply)&&(WarningReply->Trial Performed)
6.       QoSEvaluation=Evaluate QoS performance (Path);
7.       If (QoSEvaluation==valid)
         {Do nothing, the recovery operation was well-succeeded}
8.          Return  Recovery=suceeded;
9.       Else Return Recovery=unsuceeded;
10.      EndIf
11. EndIf
12. End

1. Alternate Paths Computation to (dst);
2. Begin
3.   PathOptions=Query Local BGP Routers (dst);
4.   QoSData=Query QoS History Database (PathOptions);
5.   FeasiblePaths=Compute feasible Paths (PathOptions, QoSData);
6.   Return FeasiblePaths;
7. End.

1. Preferred Path Computation (AlternatePaths)
2. Begin
3.   If!(AlternatePaths =={})
4.      {Start one transaction foreach Alternate Path;}
5.      Foreach Transaction
6.         Floods Request Preferences Messages to
                OEs members of current target Alternate Path;
7.         Upon receive all reply preferences messages;
8.         Incomingprefs=Retrieve all domains' preferences;
           {Output List will contain the set of feasible  Outputs}
9.         OutputList=Aggregate(Incomingprefs);
10.     EndTransaction;
11.     if !(OutputList=={})
        {Select from the OutputList a "social solution"}
12.     NewASPathdst= Compute New Path (OutputList);
13.     EndIf;
14.  EndIf;
15.  Return NewASPathdst;
16. End.
```

Fig. 5. Coordination Mechanism Algorithm

IV. Implementation Considerations

In this section, we describe the OE functions needed to support the deployment of the proposed cooperative approach.

QoS Measurements and SLS Violation Detection: A subjacent assumption is that the remote OEs exchange SLSs and agree upon certain QoS parameters. Since evaluating traffic volumes with accuracy is difficult and also due to the fact that IP is non-QoS aware by default, SLSs violations can occur. To support path performance evaluation and SLS violation detection, we adopt a strategy based on active end-to-end QoS measurements to derive the QoS parameters. The OEs incorporate efficient measurement methods following the recommendations of recent standardization efforts [12][13].

Gathering of alternative paths set: A fundamental requirement to deploy our out-band approach is that OEs must have administrative control over BGP speakers and thus full access to the Routing Information Base (RIB), namely to the Adj-RIBs-In and the Loc-RIB databases [1]. The gathering of alternative paths able to accommodate a traffic aggregate affected by a strong QoS degradation is an essential function. These paths and the corresponding next-hops are retrieved from the ingress domains Adj-RIBs-In, depending on the QoS measurements history.

Selected Path Set-up: The final step of the proposed mechanism is to set-up the selected path. This can easily be done by installing the route into the BGP Loc-RIB. Rather than this, our proposal is to enable OEs to modify the IP forwarding tables directly. This enables to create a soft state routing allowing OEs to rollback routing decisions and to avoid overload BGP during instability episodes. The new routes are inserted into the BGP Loc-RIB only when they are considered as stable. In the case of transit domains the advertisement to the upstream OEs that the path is at a soft state, can be demanded.

Signalling: Communication between OEs is asynchronous and over a limited network bandwidth. A signalling protocol is required. This must include requests, replies and acknowledgments for gathering of domain routing preferences, QoS data sharing or other actions. Messages are sent to individual OEs or to groups of OEs. In addition, recent recommendations in the proposal of an IP signalling protocol with QoS signalling, should take into account in the deployments [14]. When the coordination mechanism is being provided with an incentive solution, this signalling might be regarded as an useful means for rewards distribution to downstream domains.

V. Conclusion and Future Work

In this paper, we have proposed an out-band cooperative approach for coordinated inter-domain QoS routing decisions. We believe this kind of approach is the way to allow predictable inter-domain traffic exchanges between domains and to support robust SLSs in the Internet environment. As discussed in the paper, two main set of open issues are still part of our research agenda. First, it would be desirable to design of a novel coordination metric, and efficient traffic optimization algorithms to find social optimums. Additionally, to encourage OEs to deliver truthful routing preferences an incentive solution is need. Secondly, to ensure scalability, it is essential to build hierarchical OEs organizations and to define their relations in order to design a signalling protocol for OEs data sharing and actions' requests and acknowledges. Currently, we are implementing the described cooperative approach in a simulation environment based on J-Sim simulator and Infonet BGP suite [15][16]. We have planned extensive simulations contrasting its behaviour with current in-band BGP mechanisms. The results collected will enable the evaluation of the strengths and limitations of the contributions and will lead to refinements.

Acknowledgment

This work was partially funded by the European Commission through Network of Excellence E-NEXT (contract FP6-506869) under SATIN Grant *Study of Coordination Mechanisms and Signaling Protocols for Inter-domain Quality of Service Routing in a Distributed Overlay Entities Architecture.*

References

[1] Y. Rekhter, T. Li , A Border Gateway Protocol 4 (BGP-4), IETF, RFC 1771, March 1995
[2] M. Yannuzzi, A. Fonte, X. Masip, E. Monteiro, S. Sanchez, M. Curado, J. Domingo, A proposal for inter-domain QoS routing based on distributed overlay entities and QBGP, In Proc. of WQoSR2004, LNCS 3266, October 2004
[3] R. Dai, D. Stahl, and A. Whinston. The economics of smart routing and QoS. In Proc. of the Fifth Inter. Workshop on Networked Group Comm. (NGC'03), 2003.
[4] Internap Network Services, Internap Flow Control Platform, http://www.internap.com/
[5] Cisco Systems, Optimized Edge Routing, http://www.cisco.com/
[6] T. Griffin, What is the Sound of One Route Flapping?, IPAM talk, 2002
[7] T. Griffin, and G. Wilfong. Analysis of the MED Oscillation Problem in BGP. In Proc. of the 10th IEEE International Conf. on Network Protocols (ICNP02), 2002
[8] A. Akella, B. Maggs, S. Seshan, A. Shaikh, and R. Sitaraman. A measurement-based analysis of multihoming. In Proc. of ACM SIGCOMM 2003, August 2003
[9] D. Goldenberg, L. Qiu, H. Xie, Y. Yang, and Y. Zhang. Optimizing cost and performance for multihoming. In Proc. of ACM SIGCOMM 2004, August 2004
[10] S. Uhlig, O. Bonaventure, Designing BGP-based outbound traffic engineering techniques for stub ASes, ACM SIGCOMM CCR, October 2004
[11] The Internet Mapping Project, http://research.lumeta.com/ches/map/
[12] S. Shalunov, B. Teitelbaum, One-way active measurement protocol (OWAMP) requirements, IETF, RFC 3763, April 2004
[13] G. Almes, S. Kalidindi, M. Zekauskas, A one-way delay metric for IPPM, IETF, RFC 2679, September 1999
[14] R. Hancock, G. Karagiannis, J. Loughney, S. Bosh, Next Steps in Signalling, IETF, Internet-draft, November 2004
[15] J-Sim Homepage, http://www.j-sim.org
[16] Infonet Suite Homepage, http://www.info.ucl.ac.be/ bqu/jsim/