A Self-adaptive Interdomain Traffic Engineering Scheme

M. Yannuzzi, X. Masip-Bruin,
S. Sánchez-López, and J. Domingo-Pascual
Advanced Broadband Communications Center
Technical University of Catalonia
Avgda. Víctor Balaguer, s/n – 08800 Vilanova i la Geltrú Barcelona, Catalonia, Spain
Email: {yannuzzi, xmasip, sergio, jordid}@ac.upc.edu

A. Fonte[1,2], M. Curado[1] and E. Monteiro[1]
[1]Laboratory of Communications and Telematics
University of Coimbra
Pólo II, P. de Marrocos, 3030-290 Coimbra, Portugal
[2]Polytechnic Institute of Castelo Branco, Av. P. Álvares Cabral, nº12, 6000-084 Castelo Branco, Portugal
Email: {afonte, marilia, edmundo}@dei.uc.pt

***Abstract*—In this paper we propose, design, and test a self-adaptive Traffic Engineering (TE) scheme for multihomed stub Autonomous Systems (ASes). This scheme is based on a reduced set of self-adaptive and collaborative Smart Routing Managers (SRMs), which are exclusively located at those multihomed stub ASes participating in our scheme. These SRMs are able to improve the end-to-end traffic performance of delay-sensitive applications. In addition, the SRMs significantly contribute to overall network stability based on their capability to learn from and adapt to the mid and long-term network dynamics. In this sense, the TE actions performed by the SRMs not only depend on the current picture or state of the network. They also depend on the dynamic of these states, since the SRMs are able to evolve jointly with the network dynamics. This has two significant advantages. First, it substantially improves overall network stability. Second, this adaptive feature allows the traffic reallocation decision process to become transparent to the network dynamics.**

***Index Terms*—BGP, Interdomain, Multihoming, Self-Adaptive, Stability, Traffic Engineering.**

I. Introduction

One of the main issues with end-to-end Quality of Service (QoS) provisioning lies in the foundations of the current interdomain network paradigm. This paradigm is based on a highly scalable and completely distributed network architecture, which relies on the Border Gateway Protocol (BGP) [1] as the glue that keeps the Internet together. The central issue is that BGP has not inbuilt QoS capabilities. Although some researchers have proposed to replace BGP, in practice only "incremental" solutions are realistic and will have chance to become deployed. Then, such solutions should aim at complementing the deficiencies of BGP rather than replacing it, and hence, this is the approach we follow in this paper.

This work was partially funded by the Spanish Ministry of Science and Technology under contract FEDER-TIC2002-04531-C04-02, the Catalan Research Council under contract 2001-SGR00226, the European Commission through E-NEXT under contract FP6-506869, and the SATIN grants "New Interdomain QoS Routing Algorithms based on a Distributed Overlay Entities Architecture" and "Study of Coordination Mechanisms and Signaling Protocols for Interdomain Quality of Service Routing in a Distributed Overlay Entities Architecture".

Another essential issue is that stringent end-to-end QoS guarantees demands for interdomain resource reservation. This is due to the connection oriented nature of QoS. However, following such an approach at the interdomain level imposes, at least at present, several tough challenges in practical terms. As an alternative, it is possible to conceive dynamic end-to-end QoS without any kind of resource reservation, and to follow the IP connectionless paradigm, as long as only "soft" end-to-end QoS is guaranteed (i.e., QoS without reservations and hence without tight guarantees). Thus, this is the approach we follow in this work.

In this paper we propose, design and test an "incremental" Traffic Engineering (TE) scheme which aids BGP to improve the end-to-end performance of delay-sensitive applications. This scheme is addressed to multihomed stub Autonomous Systems (ASes) which are willing to cooperate with other (remote) multihomed stub ASes for economical or performance reasons. Medium and large enterprises with multihomed premises in different locations are good examples of this, and might benefit from this kind of approach. The TE scheme is supported by a reduced set of cooperative Smart Routing Managers (SRMs) exclusively located at those multihomed stub ASes participating in our scheme.

As we will see in Section VI our TE scheme aids BGP to significantly improve end-to-end traffic performance. Despite this fact, the main contribution in this paper is the design of the self-adaptive nature of the SRMs. Our tests clearly show that the best of the scheme is obtained when the SRMs are able to learn from the network dynamics and evolve with them in order to "socially" deal with the magnitude and frequency of the TE actions they perform. This is noticeable not only in terms of stability, but also in the achievements in terms of end-to-end performance such as averagely lower delays and higher throughput.

The rest of the paper is organized as follows. In Section II we present our motivations and goals. Section III

introduces our TE scheme, while Section IV presents the core of the design of the self-adaptive features of the SRMs. In Section V we present the self-adaptive TE algorithm. Then, we show our evaluation results in Section VI, and finally, Section VII concludes the paper and highlights some futures directions.

II. Motivations and Goals

Above all, multihomed stub ASes are those which are in need of new mechanisms that allow them to opportunistically manage and distribute their interdomain traffic in order to improve their performance. This particular fraction of ASes crowds together mostly medium and large enterprise customers, Content Service Providers (CSPs), and small Network Service Providers (NSPs). It is worth highlighting that nearly 80% of the total number of ASes that currently compose the Internet are stub, and the majority of this fraction is in fact multihomed. Therefore, the blast of multihomed stub ASes has gained huge interest in both research and commercial fields in the last few years. The proposals developed in this paper are addressed to this type of ASes.

Compelling recent studies like [2] demonstrate that the problem of tracking and controlling most of the traffic of multihomed stub ASes turns out to be impracticable. This is because the large variability of the topological characteristics of interdomain traffic, in addition to the limited aggregation of this traffic, indicates that the number of paths to be tracked and controlled is not only highly variable, but also really large. Despite these variability and lack of aggregation issues, several recent studies also show that a very small number of invariant paths are still responsible for a significant fraction of the existing traffic [2, 3]. By invariant we mean that these paths are stable, i.e., they are, typically, permanently present in the BGP tables and hence are not affected by the variability issues mentioned before. For instance, the measurements conducted in [2] reveal that only six invariant AS paths carried about 36% of the one-month total traffic of a real multihomed stub AS, which is indeed a significant fraction of the overall traffic. Thus, a realistic approach is that it is possible to track and manage an important portion of the existing traffic by simply controlling a reduced set of stable paths (typically 6-8 paths), and hence, this is the approach that we follow with our TE scheme.

A central issue however, is that multi-connectivity to the Internet does not necessarily guarantee improved end-to-end path diversity. This is due, first, to the fact that BGP only advertises the best path it knows, so BGP considerably prunes the total number of available paths between distant ASes, and second, to the topological characteristics of the Internet at the AS-level. Even though several studies have addressed these scarce path diversity issues [4], recent studies like [5] demonstrate that in practice multihoming in combination with route optimizers [6, 7] and TE tools are powerful techniques to improve end-to-end performance. Therefore, soft end-to-end QoS approaches are feasible. A sound explanation for this is that the Internet core is in effect an over-provisioned network.

Another important point is that cooperation among remote ASes is expected for a number of reasons. First, recent studies show that interdomain traffic is mainly exchanged among ASes that are not directly connected. Instead, they are typically 2, 3 and 4 AS hops away [3], and this also applies to the reduced set of paths to be tracked and controlled that we mentioned before. Second, cooperation between remote ASes is expected for both performance and economical reasons. Indeed, such scenario is perfectly suitable, for example, for medium and large enterprises with multi-connected offices in different cities or countries. In such a case, what each multi-connected office primarily needs is to improve the performance but for the relevant traffic exchanged with other enterprise's sites. Once again, the number of relevant paths to be tracked and controlled per-site will be typically small. For this reason, strong manufacturers such as Cisco have recently announced that they will go in this direction [6].

The combination of all the aforementioned features made us focus on a cost-effective, incremental, and cooperative TE framework among a reduced set of strategically selected remote multihomed stub ASes (typically ~ 6-8 remote peers per-ASes). Our main goals are to design this framework so that:

(i) It should improve end-to-end traffic performance for those 6-8 very significant paths per-multihomed stub AS (using simply soft-based QoS)
(ii) It should accomplish with (i) but without impacting on the stability of the network
(iii) It should be easily deployable
(iv) It should be easy to use and configure
(v) It should avoid starvation of best-effort traffic

Goals (i), (ii), and (iv) supply strong motivations to add self-adaptive capabilities to our TE scheme. This is firstly because self-adaptive TE tools are suitable to handle the inherent tradeoff between (i) and (ii). Non-adaptive mechanisms typically need to be manually tuned in the mid and long-term, since they usually become either excessively conservative, underutilizing network resources, or lead to instabilities. Secondly, network managers of multihomed stub ASes are not prone to adopt complex TE mechanisms. They expect that their traffic could be opportunistically managed based on a set of plain decisions and configured policies and they wish that such decisions last in time. Thus, as we described in [8] a major advantage behind the self-adaptive SRMs is that they are able to provide transparency to the traffic reallocation decision process, i.e., the SRMs will hide the network dynamics from this latter.

III. An Incremental Traffic Engineering Scheme

The self-adaptive TE scheme that we propose in this paper is based on a distributed and incremental architecture in which a pair of SRMs within two non-peering multi-homed stub ASes are able to: (a) exchange a Service Level Specification (SLS) and agree upon a set of "soft" QoS parameters regarding the traffic among them; (b) examine the compliance with this SLS; (c) and accurately configure on-the-fly BGP to avoid link failures, or service degradation for a set of Classes of Service (CoSs). We assume that cooperation between these ASes is desired for monetary and/or performance reasons, and as indicated in Section II, for some ASes and manufacturers this is becoming a fact.

The essence in this approach is that the QoS perception between a pair of remote ASes in our scheme is basically the one that the SRMs have of each other. This scheme has several advantages. First, with only a very small number of SRMs, but located at a strategically selected remote multihomed sites is enough to control a significant portion of the traffic of an AS. As indicated in Section II typically only 6-8 peering SRMs will be needed per-AS. Second, the intermediate ASes do not need to participate of the control architecture, and hence neither SRMs nor any kind of modifications are required in transit ASes. Therefore, the complexity of dynamic QoS provisioning is pushed to the edge of the network. Third, our approach is that a SRM within a source AS dynamically manages BGP in order to control the allocation of its outbound traffic towards a remote AS in our scheme, depending on the network conditions and QoS constraints for each CoS. This allows tweaking BGP even in very short timescales given that no BGP messages will be ever spawned. Thus, instead of proposing a complex scheme to dynamically and accurately manage how traffic enters a target AS, we focus on a collaborative scheme which handles how the traffic exits from the source AS.

Fig. 1 depicts a possible scenario were our proposal could be applied. In this simple scheme, the SRM in AS1 has two SLS agreements, i.e., one with AS2, and another with AS3. However, no agreement exists between AS2 and AS3. Thus, AS3 is not in the TE scheme of AS2 (and reciprocally) since no significant traffic needs to be tracked and controlled between them. In addition, AS4 is in none of the TE schemes.

The self-adaptive SRMs introduced in this paper are especially designed to improve the performance of delay-sensitive applications, such as Voice of IP (VoIP) or video applications. We took this approach because these are major applications which are perfectly suitable for cooperative edge routing schemes. Therefore, we choose the One-Way Delay (OWD) [9] as the end-to-end QoS information that the SRMs will collect from the network. These SRMs are designed to take opportunistic TE actions, so they will take full advantage of multihoming allowing the reallocation of traffic for a given CoS each time a sufficiently better end-to-end path exists, and the network conditions are favorable. In other words, the self-adaptive features of the SRMs restrain this opportunistic and selfish behavior with the aim of avoiding network instabilities.

In order to gather the OWD information the SRMs are endowed with mechanisms to spawn probes targeting the reduced set of remote ASes in our scheme. To fix ideas about the burden behind this practice, typically, each SRM will only need to probe 6-8 remote SRMs, for 2-3 CoSs, and through 2-3 egress links of the AS. During our experiments we observed that this practice only generates a few Kbps per-AS, for all the CoSs and for all the remote SRMs involved. Following the recommendations in the literature, we used a Pseudo-Random Poisson Process to generate the probes [9]. In this process N_u random sampling times uniformly distributed are generated over consecutive intervals of duration T_u. Then, the probability distribution function for the sampling times is given by:

$$F(t) = Uniform\left\{N_u, t \in \left[ndT_u, (n+1)dT_u\right]\right\} \quad \forall\ t \geq 0\ \wedge\ \forall n \geq 0\ /\ \mathrm{n} \in Z \tag{1}$$

We set the size and frequency of the probes to correlate the measurements with the class of traffic (application) being controlled.

Gathering QoS measurements based on OWD implies that those measurements are performed by the SRM on the destination AS. However, the dynamic reallocation of traffic is performed in the source AS. Thus, the source SRMs require feedback from its peers. This is done using the SRM-SRM protocol, which is a highly improved version of a protocol that we designed in [10][1].

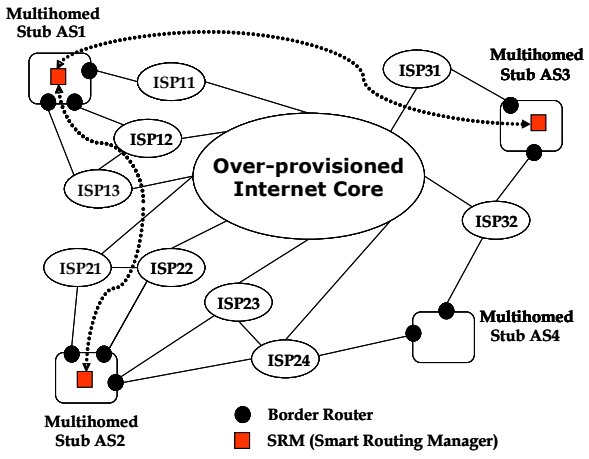

Fig. 1. Illustrative example of the proposed interdomain TE scheme

[1] We could not describe this here due to space limitations

IV. A SELF-ADAPTIVE COST METRIC FOR THE SRMS

In order to decide which is the best path to reach a remote AS in our scheme, each SRM collects QoS information and uses this information to compute the cost to reach that destination. This QoS information supplies two components to that cost. The first component consists of end-to-end QoS information, which is based on processing and filtering the OWD inferred from the probes. We call this a Smoothed OWD (SOWD)[2]. The second component consists of local QoS information, and is based on collecting the Available Bandwidth (ABW) from the egress links of the AS.

Equation (2) presents the cost metric to be used by each SRM in our TE scheme. In terms of notation, M_{ij} represents the cost to reach a distant SRM through the i_{th} egress link of the source AS, for traffic of class j. The SOWD and the ABW components described before are represented as S_{ij} and ABW_i respectively. In addition, the non-negative parameters α_j and β_j are self-adaptive weights which endow the SRMs with self-adaptive capabilities. The bound $\overline{D_j}$ represents the maximum OWD tolerable to reach a remote AS for traffic of class j in our scheme. This constraint is specified in the SLS exchanged between the SRMs, using the SRM protocol. On the other hand, the bound B_i represents the minimum acceptable ABW in the i_{th} egress link of the AS. This is an important constraint which supplies a minimum bandwidth guarantee per-egress link avoiding best-effort traffic starvation. This will be described at the end of this section.

$$M_{ij} = \begin{cases} Floor\left[\alpha_j S_{ij} + \dfrac{\beta_j}{ABW_i}\right] & if \quad \left(S_{ij} \le \overline{D_j}\right) \wedge \left(ABW_i \ge B_i\right) \\ \infty & if \quad \left(S_{ij} > \overline{D_j}\right) \wedge \left(ABW_i < B_i\right) \end{cases} \quad (2)$$

If any constraint is violated the cost M_{ij} is set to infinite. This means that the i_{th} egress link of the AS should be removed from the list of available links for output traffic of class j as long as a violation to the SLS exists.

The main motivations for selecting this additive cost metric can be summarized as follows:

- This self-adaptive cost is not only able to reflect the present QoS dynamics, but also to evolve with them providing transparency to the traffic reallocation decision process (as it was indicated in Section II). Based on the computation of this cost the SRMs are able to tweak BGP in short timescales for the outbound traffic of the AS. In fact, network administrators of any multihomed stub AS using a SRM will only need to select and configure a fixed threshold, so that if this threshold is met then the reallocation of traffic is allowed. The self-adaptive capabilities mean that if the network conditions are adequate to reallocate the traffic, then the threshold will be more frequently met. However, if the conditions are inadequate, then larger variations in the QoS conditions are needed to meet the same threshold value.

- Equation (2) allows to flexibly design the cost unbalancing the sensitivity of the weights α_j and β_j. The motivation for this is to magnify the relevance of the OWD over the local ABW. However, the additive term of ABW in (2) preferably allows the allocation of traffic over links with more ABW when OWD conditions are similar. This allows an overall better traffic distribution for the outbound traffic from the AS. We will show in Section VI that this simple approach is suitable to comply with demanding SLSs even under stressful traffic network conditions.

In order to compute the cost in (2) the SRMs smooth the OWD samples gathered from the probes using two filters in cascade. The first filter corresponds to the median OWD through a sliding window of size W_j as shown in (3). The index n_{ij} in (3) simply represents the sequence number of the instantaneous samples of OWD.

$$Median\left(OWD_{ij}(k)\right),\ k \in [n_{ij} - W_j + 1, n_{ij}] \quad (3)$$

Our motivation for choosing the median is that it is widely accepted as the best estimator of the OWD that the user's applications are actually experiencing in the network. The sliding window is typically tuned in order to get a good trade-off between, the responsiveness of the filter, and a strong correlation between the measurements and the applications data being controlled.

The second filter is applied to the median OWD, using what we call a one-dimensional grid or simply the grid. This grid works like an A/D converter with a self-adaptive pace of conversion, in which the pace is constantly adapted depending on the QoS conditions on the network. If the conditions are smooth the pace is small, and in practice more traffic reallocations will be allowed. In such conditions a SRM is able to take full advantage of the multi-connectivity of its AS to the Internet. However, if the conditions may lead to instabilities the pace increases and the traffic reallocations allowed by the SRMs are diminished or even stopped until the network conditions become smooth once again.

In the rest of this section we will describe in detail the design of this grid and its relation with the weights α_j and β_j in (2). During this process Figs. 2 and 3 will help to understand the design of the grid and the cascade filters described before.

A. Designing the grid

The outcome of the second filter, i.e. the grid, is the term S_{ij} in (2). This grid provides some granularity to the samples of OWD in such a way that it can be exploited by an opportunistic TE algorithm. Our goal is to design this grid with self-adaptive capabilities, particu-

[2] Our aim is to avoid frequent changes of the QoS information, so the SRMs use a SOWD instead of instantaneous values of the OWD collected.

larly, depending on the mid and long-term QoS conditions in the network. To accomplish this goal, the interval $[0, \overline{D_j}\ ]$ is initially divided in N_{ij} subintervals, i.e.:

$$\left[m\frac{\overline{D_j}}{N_{ij}}, (m+1)\frac{\overline{D_j}}{N_{ij}} \right] \; m \in Z \,/\, 0 \le m \le (N_{ij} - 1) \tag{4}$$

defining an initial set of grids $\forall\ i, j$. In order to design this grid we define the following parameters using the first W_j instantaneous samples of OWD_{ij}.

$$\begin{cases} \overline{K_{ij}} = \max\left\{ OWD_{ij}(k) \right\} & \forall\ k = 1,\ldots,W_j \\ \underline{K_{ij}} = \min\left\{ OWD_{ij}(k) \right\} & \forall\ k = 1,\ldots,W_j \end{cases} \tag{5}$$

Then, the interval $\left[\underline{K_{ij}}, \overline{K_{ij}} \right]$ defines our first estimation of the range of variation of the instantaneous samples of OWD_{ij} (see Fig. 2). Our aim is to prevent frequent variations in the cost, so the main idea behind the grid is that moderate variations of the samples given in (3) generate the same numerical value of S_{ij}, and thus the same cost M_{ij}. Our first criterion while designing this grid is that the maximum variation $\left(\overline{K_{ij}} - \underline{K_{ij}} \right)$ fits into one subinterval of the grid. Moreover, we introduce an adjustable coefficient $\Delta_j \in R \,/\, \Delta_j \ge 1\ \forall\ j$, which assures at least a percentage of separation between the grid lines and the parameters defined in (5) given by $(\Delta_j - 1)10^2$. In this sense Δ_j basically reflects the degree of conservativeness while defining the initial grid. In addition, Δ_j will also play a fundamental role when adding self-adaptive capabilities to the grid. Fig. 2 shows the grid design approach. Accordingly, N_{ij} is bounded by:

$$\left.\begin{array}{l} \dfrac{m\overline{D_j}}{N_{ij}} \le \Delta_j^{-1}\underline{K_{ij}} \\ \dfrac{(m+1)\overline{D_j}}{N_{ij}} \ge \Delta_j \overline{K_{ij}} \end{array}\right\} \Rightarrow N_{ij} \le \frac{\overline{D_j}}{(\Delta_j \overline{K_{ij}} - \Delta_j^{-1}\underline{K_{ij}})} \tag{6}$$

$$N_{ij} \in Z \,/\, N_{ij} \ge 1$$

In order to provide a scalable design of the grid we impose the following restriction. For each CoS j, and $\forall\ i, k \,/\, i \neq k\ \ M_{ij}\ N_{ij} = N_{kj}$. This is indeed a reasonable decision since comparing costs M_{ij} and $M_{kj}\ \forall\ i \neq k$, only will make sense if the same grid is used for traffic of class j over every egress link from the source AS. Thus, we define: $N_j \equiv N_{ij}\ \forall\ i$.

Clearly, the advantages of this discrete arrangement may be lessened if the granularity is enough to cause that the opportunistic TE impels an SRM to frequently re-configure the border BGP routers of its AS. Thus a trade-off exists in terms of the granularity of the grid, and how proactively the traffic will be reallocated.

Following a conservative approach, our second criterion is to use the $\max_i [\Delta_j^{-1}\underline{K_{ij}}, \Delta_j \overline{K_{ij}}]$ when determining N_j. Thus:

$$N_j = Floor\left(\frac{\overline{D_j}}{\min\left\{ \overline{D_j}, \max_i \left(\Delta_j \overline{K_{ij}} - \Delta_j^{-1}\underline{K_{ij}} \right) \right\}} \right) \tag{7}$$

where (7) satisfies the restriction in (6) and provides a common grid $\forall$ egress link i for traffic of class j.

Then, if we define $G_j = \dfrac{\overline{D_j}}{N_j}$ as the step of the grid for the j_{th} CoS, the S_{ij} to be used in (2) is defined as follows:

$$S_{ij} = \begin{cases} Median\left(OWD_{ij}\right) & N_j = 1 \\ Floor\left(\dfrac{Median\left(OWD_{ij}\right)}{G_j} \right) & N_j > 1 \end{cases} \tag{8}$$

Equation (8) is the most general expression for S_{ij}, and anticipates an interesting feature of our approach, that is, the SRMs can be configured in two different modalities: (i) an opportunistic SRM; (ii) a reactive SRM.

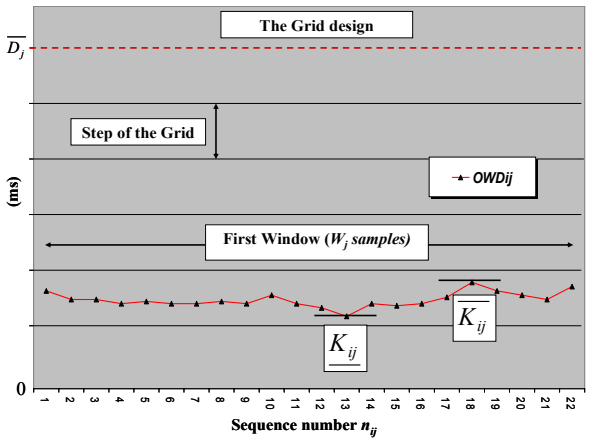

Fig. 2. The grid design

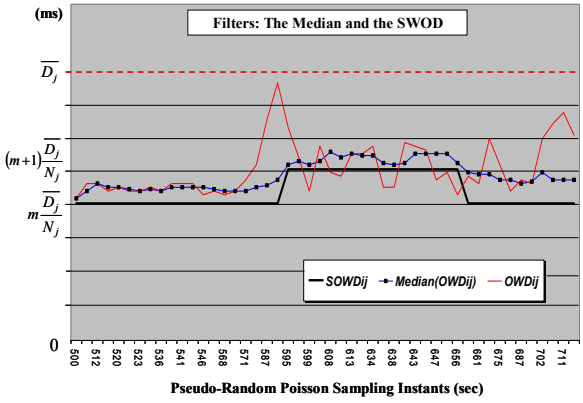

Fig. 3. The two filters: The Median and the SOWD S_{ij}

The difference in the reactive approach is that traffic is only reallocated when a violation to a SLS is detected. In other words, in this configuration the SRMs are not opportunistic any more.

Then, each network administrators of the multihomed stub ASes in our scheme could independently decide how to configure its SRM in order to comply with the SLS already agreed. An opportunistic SRM is a SRM using the "self-adaptive A/D converter", i.e. the grid, which corresponds to the case $N_j > 1$ in (8). Conversely, the reactive case (N_j =1) simply uses the first filter, since there is no need for the second filtering/smoothing process and compares the cost in (3) against $\overline{D_j}$.

Furthermore, our SRMs are designed to automatically switch from proactive to reactive when the network dynamics become aggressive (in the event of anomalies such as misconfigurations, link flaps, etc). Certainly, the SRMs are able to automatically resume their opportunistic behavior once the problem disappears. It should become clear that in a scenario like the one depicted in Fig. 1, each AS participating in our scheme may independently decide how its SRM will be configured (restricted to the compliance of the established SLS agreements).

Fig. 3 depicts the relation between the median and S_{ij}. As aforementioned, the development of a self-adaptive cost in terms of S_{ij}, demands that a SRM should be able to dynamically adapt this grid depending on the QoS conditions. Our approach is to avoid frequent recalculations of the grid during unstable network conditions, so we propose that each time a new grid is computed this is maintained for a superset of several windows S_{W_j} (several W_j windows). Then, we trigger the recalculation of the grid whenever:

$$\begin{cases} G_j^{(n)} < \max\limits_i \left(\overline{K_{ij}}^{(n)} - \underline{K_{ij}}^{(n)} \right) & \vee \\ \min\limits_i \left(\Delta_j \overline{K_{ij}}^{(n)} - \Delta_j^{-1} \underline{K_{ij}}^{(n)} \right) < \overline{K_j}^{(n)} - \underline{K_j}^{(n)} \end{cases} \quad (9)$$

where the supra-index (n) represents the current grid, and:

$$\begin{cases} \overline{K_{ij}}^{(n)} = \max \left\{ OWD_{ij}(k) \right\} & \forall k \in S_{W_j}^{(n)} \\ \underline{K_{ij}}^{(n)} = \min \left\{ OWD_{ij}(k) \right\} & \forall k \in S_{W_j}^{(n)} \end{cases} \quad (10)$$

with $\left[\underline{K_j}^{(n)}, \overline{K_j}^{(n)} \right] = \max\limits_i \left[\underline{K_{ij}}^{(n-1)}, \overline{K_{ij}}^{(n-1)} \right]$.

Then, a new grid (n+1) is obtained when the substitution of $\max\limits_i \left(\Delta_j \overline{K_{ij}}^{(n)} - \Delta_j^{-1} \underline{K_{ij}}^{(n)} \right)$ in (7) supplies a new $N_j^{(n+1)}$ / $N_j^{(n+1)} \neq N_j^{(n)}$. The first inequality in (9) reflects that the network conditions have become rather unsteady so the step of the grid $G_j^{(n)}$ needs to be increased, while the second inequality indicates that the conditions have become even steadier so the step of the grid could be diminished. It is possible that while an egress link *i* satisfies the first inequality in (9) for a given CoS *j*, another egress link *k* satisfies the second one. In such a case our decision is to follow a conservative approach so we choose to increase the step of the grid.

Finally, if the grid was not recomputed throughout $S_{W_j}^{(n)}$ instead of setting the current grid for a whole new superset, the SRMs began to search for either of the conditions in (9) through a sliding window of size S_{W_j}. This mechanism will speed up the reaction of our self-adaptive cost when network conditions have changed.

B. Linking the grid with the weights α_j and β_j

The natural next step in the design of our cost is to link the weights α_j and β_j in (2) with the self-adaptive features of the previous grid. Moreover, this needs to be done in such a way that the cost and the TE algorithm using this cost could bring transparency to the traffic reallocation decision process [8].

Given that the self-adaptive part of the grid is indeed its step, α_j and β_j should explicitly depend on the step G_j. Thus, the next propositions provide design criterions in order to choose a couple (α_j, β_j) that could supply the desired transparency to the traffic reallocation decision process, and fulfill our design motivations (recall our motivations behind the election of equation (2))

Proposition 1: When S_{ij} increases by one step of the grid, the weight α_j that will increase the cost M_{ij} at least by a "fixed" and positive integer value Q, and with maximum sensitivity, if given by:

$$\alpha_j = \frac{Q+1}{G_j} \quad (11)$$

Proof. Using (2) and assuming that M_{ij} increases by the fixed value Q, we have:

$$\left(\alpha_j \left(S_{ij} + G_j \right) + \frac{\beta_j}{ABW_i} \right) - a_2 = \left(\alpha_j S_{ij} + \frac{\beta_j}{ABW_i} \right) - a_1 + Q \quad (12)$$

where the parameters $a_h \in [0,1)$ $h = 1,2$ come from the Floor function, hence:

$$\alpha_j = \frac{a_2 - a_1 + Q}{G_j} \Rightarrow \frac{Q-1}{G_j} < \alpha_j < \frac{Q+1}{G_j} \quad (13)$$

Next, if we choose the upper bound this will supply the maximum sensitivity to M_{ij} given its linearity in α_j (except for the Floor truncation). Therefore, α_j explicitly depends on the resolution of the grid G_j, and hence it evolves with it.

Corollary 1: Given that Q is a "fixed" and positive integer value, the variations of M_{ij} in terms of S_{ij} become independent of the state of the grid.

Proof. Using (11), the variation in M_{ij} when S_{ij} increases one step of the grid is:

$$\left(M_{ij}^{new} - M_{ij}^{old}\right) = (Q+1) + a_1 - a_2 \tag{14}$$

Then, the variations of M_{ij} are independent of G_j and hence are independent of the QoS dynamics. In other words, while the resolution of the grid evolves in time according to those QoS dynamics, we adapt the calculus of M_{ij} depending on the state of the grid so that the opportunistic TE algorithm remains transparent to them. The advantage of a self-adaptive α_j like the one proposed in (11) is that it allows multihomed stub ASes seeking for an opportunistic approach, to transparently opt for a degree of conservativeness which is independent of the QoS conditions in the network. Thus, this decision will last in time and it will be done by simply configuring a fixed threshold within the SRMs, so that network managers do not need to get into the details of the metric.

Proposition 2: The sensitivity of M_{ij} in the OWD dimension is higher than its sensitivity in the ABW dimension if we chose the following weight β_j:

$$\beta_j = \left(\frac{Q+1}{G_j}\right) \min_i \left\{B_i^{\,2}\right\} \tag{15}$$

Proof. Let $\widetilde{M}_{ij}$ represent the argument of the Floor function in (2), i.e., $M_{ij} = Floor\left(\widetilde{M}_{ij}\right)$. Then, our restriction in the sensitivity of M_{ij} can be expressed as follows:

$$\left|\frac{\partial \widetilde{M}_{ij}}{\partial S_{ij}}\right|_{ABW_i} \geq \left|\frac{\partial \widetilde{M}_{ij}}{\partial ABW_i}\right|_{S_{ij}} \quad \forall\, S_{ij} \leq \overline{D_j} \;\wedge\; \forall\, ABW_i \geq B_i \tag{16}$$

Then if:

$$\left|\frac{\partial \widetilde{M}_{ij}}{\partial S_{ij}}\right|_{ABW_i} = \max_i \left\{\left|\frac{\partial \widetilde{M}_{ij}}{\partial ABW_i}\right|_{S_{ij}}\right\} \Rightarrow\; \alpha_j = \frac{\beta_j}{\min_i [B_i^2]} \tag{17}$$

wherein substituting for α_j, we get (15). We emphasize that the motivation behind this last proposition is to amplify the relevance of OWD over the local ABW, but use the ABW as a tie-break when end-to-end OWD conditions are similar. An important corollary from (15) is that now β_j also becomes self-adaptive weight since it explicitly depends on the grid resolution G_j.

This concludes the design of the cost metric that will feed the TE algorithm running in the SRMs. Next, we briefly describe how we guarantee the survivability of the background traffic under highly stressful network conditions.

C. Avoiding Best Effort traffic starvation

In order to avoid starvation of the best effort traffic outgoing any AS in our TE scheme, we utilize the following design criterion to guarantee a minimum bandwidth for this traffic:

$$\sum_i \left(C_i - ABW_i\right) - \sum_i \sum_j B_{ij}^c = \sum_i B_i = << BE^{\,peak} \tag{18}$$

C_i link capacity of the i_{th} egress link of the AS.

B_{ij}^c bandwidth consumed for traffic of class j over the i_{th} egress link of the AS.

$BE^{\,peak}$ peak for the overall best effort traffic

A simple approach is to perform static provisioning over each egress link i so that the overall provision fulfills the restriction in (18). This is basically the approach followed in this paper. In Section VI we will show that with this simple reservation is enough to comply with demanding SLSs even under stressful traffic network conditions, and avoid the starvation of the background traffic.

V. An Opportunistic and Self-Adaptive Traffic Engineering Algorithm

The interdomain TE algorithm presented in this section feeds from the cost metric M_{ij} in an opportunistic way, so it triggers the reallocation of traffic even when no violations to the SLAs occur. This is precisely what the current route optimizers commercially available do. However, our SRMs have two important advantages when compared against these optimizers:

(i) They have self-adaptive capabilities to "socially" restrain their selfishness in case this is needed. In fact, they are able to automatically switch between proactive and reactive behaviors depending on the QoS conditions. In the extreme case, if the soft QoS SLSs cannot be fulfilled for a class of traffic $j \Rightarrow M_{ij} = \infty \;\forall\; i$ and packets belonging to this class will be forwarded based on the state of BGP advertisements. Thus, the SRMs will stop reallocating traffic for class j until the conditions become adequate once again.

(ii) Our SRMs don't work alone, and hence they try to improve the traffic performance but towards a reduced set of very important remote networks. Conversely, currently available tools try to control a large number of paths and are intrinsically selfish, which from our perspective is not only not necessary, but also prone to several improvements. In Section VI we will show that selfishness is precisely far from optimal.

Our opportunistic TE algorithm works as follows. Traffic of class j may be reallocated from an egress link i to link k if and only if their costs satisfy:

$$\left(\frac{M_{ij} - M_{kj}^{est}}{Q+1} \right) \geq R_j^{th} \quad (19)$$

Inequality (19) introduces the fixed threshold R_j^{th} that a network manager of a multihomed stub AS will need to choose and configure in its SRM. This will be done according to its "*degree of conservativeness*" for each CoS, i.e., the number of path shifts that the AS is willing to permit for each traffic class. The motivation behind this selection is that with this criterion the threshold basically counts the number of steps that S_{ij} needs to increase in the adaptive grid in order to reallocate traffic of class *j*. Furthermore, instead of directly comparing M_{ij} with M_{kj} we compare with the estimation given in (20), in which B_{ij}^c corresponds to the bandwidth consumed for traffic of class *j* over the current egress link *i*.

$$M_{kj}^{est} = Floor\left[\alpha_j S_{kj} + \frac{\beta_j}{\left(ABW_k - B_{ij}^c\right)} \right] \quad (20)$$

The next piece of pseudo-code succinctly summarizes the operation of our opportunistic TE algorithm:

Opportunistic and Self-Adaptive Interdomain TE Algorithm

```
if there exists a set of egress links
f which satisfy:
  (M_ij - R_j^th (Q+1)) ≥ M_fj^est
  Trigger traffic reallocation of
  class j from i to k / M_kj = min{M_fj},
  ∀ f ≠ i
   end.
```

Fig. 4. Piece of pseudo-code summarizing the operation of the opportunistic and self-adaptive TE algorithm

It is important to highlight that the only parameters that a network manager of a multihomed stub AS will need to configure are: (i) R_j^{th}; (ii) the set of constraints $\overline{D_j}$ and B_i. The rest of the parameters will be typically set to default values, but they can be manually tuned by network managers. For example, the values in *P* from Table I will be set to correlate the type of traffic being controlled. Moreover, the *Q* value is basically a "cosmetic" value which mainly controls the scale of the cost. In practice (19) shows that the TE decision is essentially independent of this value (except for the Floor truncation). Thus, setting these parameters using default values is not an issue, and they can always be reconfigured by network administrators.

TABLE 1.
SET OF PARAMETERS INVOLVED IN OUT TE SCHEME

J	set of CoSs { *j* }
$\boldsymbol{C_j}$	set of constraints for the CoSs: { $\overline{D_j}$ } $\forall j$
B	set of $\{B_i\}$ $\forall i$
S	set of remote SRMs $\{SRM_k\}$ $\forall k$
P	Pseudo-Random Poisson parameters { N_u, dT_u}
G	Parameters of the Grid: ▪ Sliding windows $\{W_j, SW_j\}$ $\forall j$ ▪ Δ_j (Conservativeness factor) ▪ Q (Fixes the scale of the cost M_{ij})
W	parameters for the sliding windows, $\{W_j, SW_j\}$ $\forall j$

VI. EVALUATION RESULTS AND COST OF IMPLEMENTATION

In this section, we investigate and test the behavior of the self-adaptive SRMs developed in the preceding sections.

A. Objectives and simulation setup

Our first objective is to assess how much they aid to improve end-to-end network performance. This is shown by evaluating: (i) the end-to-end OWD (or latency) for the different CoSs; and (ii) the traffic transfer efficiency to study the traffic performance of each CoS. This efficiency parameter is given by $Ef_{nsj} = C_{nj} / C_{sj}$, where C_{nj} is the throughput at a given destination *n*, and C_{sj} is the throughput originally sent by the source domain *s* for traffic of class *j*. Our second objective is to assess and contrast the time needed by the SRMs to react to a link failure, against the time it takes for BGP to converge to a new route. Finally, our third objective is to study how the self-adaptive SRMs contribute to overall network stability under variable QoS dynamics, and when different SLSs are used between remote multihomed stub domains. In this case, we use as performance indicator the total number of path shifts needed to meet the QoS constraints within the different SLSs.

It is important to highlight that our tests were conducted in a set-up where various sets of cooperative SRMs were simultaneously running and sharing the network resources. Each of these sets represents a scheme similar to the one depicted in Fig 1. In addition, each set is unaware of the existence of the other sets, so neither cooperation nor coordination exists among the distinct sets. This allows us to assess the impact on the traffic and the overall network stability caused by the interference between several groups of SRMs running in parallel. Therefore, when analyzing the performance of the SRMs it is important to keep in mind that we will be taking into account all the SRMs present in the network.

Our simulations were performed using the J-Sim [11] simulator with the BGP Infonet suite [12] in which we

have implemented the functionalities of the SRMs. A simplified picture of the network model used for our simulations is illustrated in Fig. 5. In this model, peering SRMs belonging to different ASes spanning across several AS hops are able to exchange a SLS and agree upon a set of soft QoS parameters concerning the traffic among them.

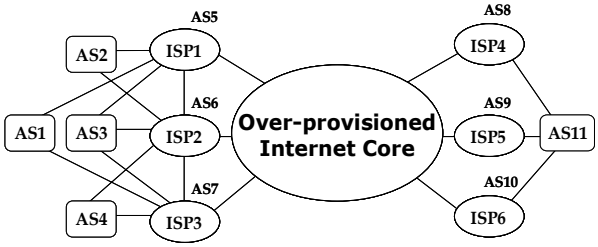

Fig. 5. Simulation network topology.

The simulated network aims at representing a multi-service part of the Internet composed by access ISPs able to provide some limited QoS services to some of their customers, and an over-provisioned Internet core (composed by Tier 1 and big Tier 2 ISPs). The motivation for this is that we believe this is the tendency that will be forthcoming in the near future. In this set-up, our tests were conducted using a traffic mix consisting of Voice over IP (VoIP) calls, video calls, prioritized data downloading, web browsing and email downloading.

In order to handle the different CoSs, we have chosen during the evaluations a classical approach which is to rely on Differentiated Services (DiffServ) capabilities at the edge of the network [13]. In other words, packets are classified and forwarded at the edge of the network using the standard Expedited Forwarding (EF), Assured Forwarding (AF), and Best-Effort (BE) traffic classes [14, 15]. In this case, what we mean with "edge of the network" is the set of multihomed stub AS customers and their corresponding ISPs (see Fig. 5). Thus, it is important to emphasize that DiffServ is not running in the rest of the network, so the Internet core is released from the burden of managing different CoSs (there is no need for packet differentiation since this is an over-provisioned network). In this model the complexity of QoS differentiation is pushed to the edge of the network using a standard framework. An important advantage of this proposal is that peering SRMs could agree upon a particular set of CoSs, and their corresponding ISPs do not need to know anything about this agreement. The only requirement is that this particular set of CoSs handled by the SRMs, needs to be mapped to standard EF and AF classes, so that their ISPs could interpret them and provide the desired treatment while forwarding the customer's packets.

In our experiments we run the same simulations separately (and incrementally) using four different scenarios:

(i) BGP, i.e., all traffic is Best-Effort hence neither DiffServ nor SRMs are present during these simulations.
(ii) BGP combined with DiffServ at the "edge of the network"[3], so no SRMs are present during these simulations.
(iii) BGP combined with DiffServ at the "edge of the network", as well as non-adaptive SRMs[4].
(iv) BGP combined with DiffServ at the "edge of the network", as well as self-adaptive SRMs.

During all the experiments, we used three different SLSs exchanged between remote multihomed stub domains, based on maximum OWD for voice, video and prioritized data traffic. These maximum OWDs tolerated per-service were heuristically chosen to represent reasonable but demanding values for the kinds of traffic sources considered. This is depicted in Table 2. Our aim is test the performance of the SRMs when realistic but tough SLSs are in use, so as to increase the probability of SLS violation events.

TABLE 2.
SLS DESCRIPTIONS

SLS ID	Voice (EF) [ms]	Video (AF21) [ms]	Prioritized (AF11) [ms]
1	45	55	75
2	55	65	85
3	65	75	95

B. Traffic models and default parameters

The source models are based on [16] and are as follows: (i) voice traffic is marked with the EF code-point. The EF source is an ON-OFF VoIP generator. In the ON state, the EF voice source generates traffic at a peak rate of 64kbps; (ii) video traffic is marked with the AF1 code-point. The AF1 source is a video traffic generator characterized by Pareto ON-OFF. In the ON state an AF1 source generates video traffic at a peak rate of 200 kbps; (iii) finally the prioritized data traffic is marked with the AF2 code-point. Both, AF2 and BE traffic are characterized by Poisson processes with sources generating traffic at 350Kbps, and 500Kbps, respectively. During the tests, new voice and video connections uniformly distributed arrived at the multihomed stub AS's border routers throughout the simulation runtime, while all data connections were active during almost all the simulation runtime.

In addition, the frequency and size of the probes spawned by the SRMs using the Pseudo-Random Poisson process were correlated with the CoS being controlled. Given that we handled three different CoSs (voice, video and prioritized data) we used three different types of probes matching with the corresponding

[3] The meaning of "edge of the network" is the same as above
[4] The non-adaptive SRMs are basically our SRMs but lacking of any sort of "social conscience", i.e., they are completely selfish as any of the current commercially available route optimizers.

traffic DiffServ Code Point (DSCP).

During the evaluations we configured the following default set of parameters i) $Q = 20$; ii) $\Delta_j = 1.05$; iii) and the sliding window parameters in $W_j = 8$ samples, and $S_{Wj} = 5\ W_j$. These default values were tuned after numerous simulation runs in order to get a good trade-off between speed of reaction (selfishness), and stability. Certainly, all these parameters may be configured (tuned) by the multihomed stub AS customers.

C. Evaluation of overall performance optimizations

Figs. 6 and 7 contrast the end-to-end traffic performance for the four different scenarios, and the three different SLSs, under stressful traffic conditions, i.e., with high load in some links at the edge of the network. Our aim was to generate some bottlenecks in order to stress the competition between the self-adaptive SRMs for less-loaded paths, and observe how they handle to comply with rather challenging SLS. This is typically the case of a number of edge Asian networks were the average link occupancy is really high.

In terms of latency, Fig. 6 shows that the first scenario exhibits the worst performance. In this case only BGP is running (all traffic is BE), and clearly BGP is unable to dynamically avoid the most congested paths. The way traffic will flow in this scenario completely depends on the state of the local routing policies, and so, it depends on how routes are being advertised by BGP. Thus, the only way to improve end-to-end traffic performance in this case is by means of manual intervention. The network administrators need to start tuning their BGP border router configurations on a trial-error basis. Fig. 6 clearly shows that under stressful traffic conditions, on average, BGP is incapable of complying with the SLSs in Table 2. The average latency is extremely high ($\cong$ 900ms) due to the contribution of the most congested links to the additive OWD. Moreover, Fig. 7 shows that the total number of losses is unacceptably high in this case (for instance, $\cong$ 20% for voice traffic).

In the second scenario, i.e. BGP combined with DiffServ at the edge of the network, the latency is improved by more than 10 times. However, the traffic throughput efficiency in Fig. 7 visibly shows that this is accomplished by strongly penalizing BE traffic, which is absolutely undesirable. Additionally, the SLS IDs 1 and 2 for video, and the SLS ID 1 for prioritized data, are still averagely unfulfilled. Therefore, this scenario does not constitute a definitive solution to improve end-to-end performance.

The third scenario adds the non-adaptive SRM sets to our previous scenario. Figs. 6 and 7 clearly show that the inclusion of the non-adaptive SRMs supplies significant improvements, both, in terms of latency and in terms of efficiency. The SRMs are able to bypass the most congested links without needing to penalize BE traffic (due to the simple starvation avoidance mechanism presented in Section IV).

Even though these improvements, the non-adaptive SRMs exhibit two important weaknesses. Firstly, on average they are incapable of complying with some particular combinations of SLSs, like SLS ID 2 for video traffic. The explanation for this is that the selfishness of the SRMs, in addition to the interference caused by the various sets of SRMs operating in the network at the same time, generates a large number of path shifts, especially under stressful traffic conditions. This leads to some oscillatory behaviors of some SRMs for the CoSs mapped to AF, which results in the non-compliance of the SLS during these unsteady states. Accordingly, the non-adaptive SRMs might not be able to averagely comply with certain combination of SLSs.

The second and most important weakness of the non-adaptive SRMs clearly resides in the number of path shifts needed to fulfill all the SLSs. This disadvantage can be seen at a glance in Fig. 9 (we will analyze Fig. 9 in Section E).

In our fourth scenario, i.e. BGP combined with DiffServ and the self-adaptive SRMs, we observe the higher performance benefits under stressful traffic conditions. The self-adaptive SRMs are able to route around congested links, outperforming the accomplishments of the non-adaptive SRMs. In particular, for the cases of video and prioritized data traffics in Fig. 6, the self-adaptive SRMs clearly improve the end-to-end average latency by $\cong$ 10%. It is important to notice that in this scenario all the SLSs are fulfilled for all the traffic classes involved. Thus, the capability of the SRMs to adjust the cost metric M_{ij} based on their learning from the mid and long-term network conditions, clearly improves the route optimization decisions that the SRMs are able to take. Another important fact is that the contribution of the self-adaptive SRMs to reduce the latency of delay-sensitive applications does not lessen the traffic transfer efficiency as shown in Fig. 7. In fact, both kinds of SRMs provide almost the same traffic transfer efficiency. From Figs. 7 and 9 it is evident that, even though the adaptive SRMs are prone to present in some cases slightly more losses than the non-adaptive ones, the improvements in terms of overall network stability are overwhelming as shown in Fig. 9.

Finally, from Fig. 7 it is clear that the efficiency of the background traffic (BE traffic) is highly improved by the starvation avoidance mechanism used by the SRMs when compared against the first two scenarios.

D. Response time of the SRMs to link failures

Fast recovery from failures is a requirement that cannot be fulfilled by BGP. Thus, in this section we try to assess how the SRMs are able to manage and react to a remote link failure. Fig. 8 contrasts the results obtained for the first two scenarios against the results obtained in our fourth scenario, i.e. when the self-adaptive SRMs are in use. For the first two scenarios we have aggres-

sively configured small keepalive and hold timer values (30 and 90 seconds respectively) so as to increase the speed of reaction of BGP [1]. From Fig. 8 we can observe that in case of links failures occurring a few hops away from a certain AS, the SRMs are able to react, on average, between 3-4 times faster than the time that BGP needs to converge to new route. In addition, since the SRMs gather end-to-end measurements their responsiveness becomes independent of the place where the remote link failure occurs. As a result, the self-adaptive SRMs are also able to reduce performance degradations (e.g. packet loss and latency) arising from inaccurate routing states during BGP transient fail-over periods.

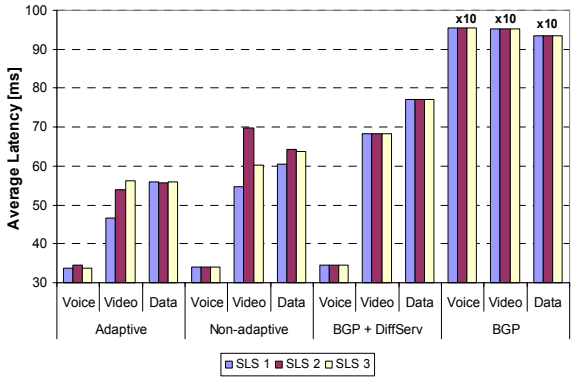

Fig. 6. Overall average latency for all scenarios and SLSs[5].

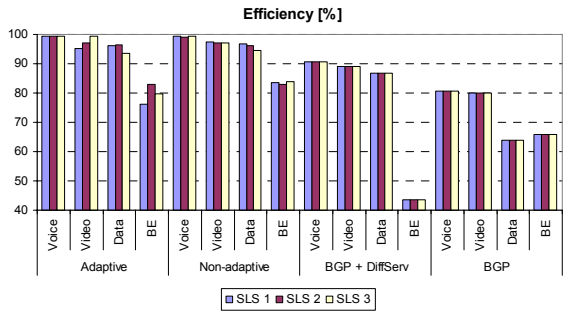

Fig. 7. Overall traffic transfer efficiency for all scenarios and SLSs.

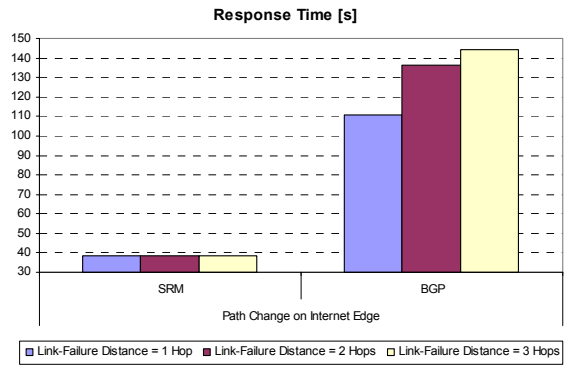

Fig. 8. Contrasting the response time to a link failure.

[5] Both Figs. 7 and 8 were obtained using the normalized threshold $R_j^{th} = 1$, which is the most demanding configurable value of R_j^{th}.

E. Evaluation of overall network stability

The main goal of the SRMs is to provide steady traffic patterns were all the SLSs are fulfilled. However, such a goal may lead to network instabilities, especially, when a large number of SRMs are present on the network and frequent selfish routing optimizations are made. Consequently, the impact of these optimizations on network needs to be reduced as much as possible. During our experiments we observed that, although both types of SRMs are able to bypass congested segments of network and avoid SLS violations, the non-adaptive SRMs exhibit some oscillatory routing behaviors. On the other hand, we do not observe these anomalies when self-adaptive SRMs are in use.

The contribution of the self-adaptive SRMs to the overall network stability is evaluated by the total number of path shifts occurred in all the multihomed stub ASes using our TE scheme. This is assessed for different values of the degree of conservativeness R_j^{th}, as illustrated in Fig. 9. This figure shows that the number of path shifts introduced by the self-adaptive SRMs is clearly much smaller than those introduced by the non-adaptive SRMs. In some cases the number of path shifts is reduced by nearly one order of magnitude. Moreover, the configuration of different thresholds R_j^{th} in the SRMs helps to dramatically reduce the number of path shifts supplying an additional improvement to overall network stability.

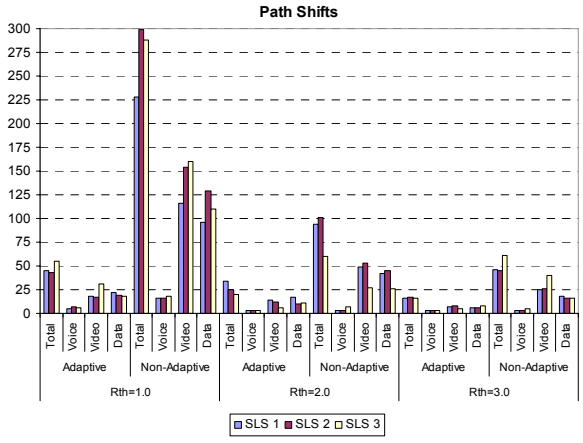

Fig. 9. Overall number of path shifts for scenarios 3 and 4, for all the SLSs, and for different normalized thresholds.

F. Summary and cost of implementation

To sum up, our fourth scenario with the self-adaptive SRMs is the only one where all the SLSs are fulfilled for all the traffic classes involved. From our perspective, the main conclusions that can be drawn from the results we found are:

- The usefulness of the self-adaptive SRMs to straightforwardly improve end-to-end performance is evident.
- Selfishness is clearly far from optimality. A better performance is obtained when the SRMs are endowed with some sort of "social conscience" and learn from the network dynamics.
- For the set-ups and experiments that we conducted it was possible to simultaneously comply with several challenging SLSs without any kind of manual intervention when our TE scheme is in use.

The cost to implement our TE scheme in an operational AS can be summarized as follows:

- A commodity PC running Linux is enough to support and implement the SRM capabilities for the current Internet
- Each SRM will typically interact with only 6-8 remote SRMs. This is enough to track and control a significant portion of the traffic of the AS
- 2-3 CoSs will provide enough granularity to differentiate traffic in the present Internet
- The total control traffic generated by the Pseudo-Random Poisson process is typically in the order of a few Kbps.
- The SRMs need full access to the BGP border routers within the AS
- Multihomed stub ASes using our scheme are small-sized ASes which don't use hot potato routing
- The impact of the optimizations carried out by the SRMs on iBGP is negligible for these small-sized ASes

It is worth mentioning that we are going to implement and test this scheme in a real environment in the frame of an European research project [17].

VII. Conclusions and Future Work

In this paper we have proposed, designed, and tested a cooperative self-adaptive TE scheme for multihomed stub ASes. The TE actions are controlled and performed by a set of SRMs which are able to improve the performance of delay-sensitive applications. In addition, these SRMs significantly contribute to overall network stability based on their capability to learn from and adapt to the network dynamics. We believe this is an appealing proposal for a number of reasons. First, neither SRMs nor changes are needed in any transit domain. Second, this incremental scheme is easily deployable. And third, it demands nearly no additional resources from the ASes participating in it. The scheme focuses on tracking and controlling only a reduced set of very important paths per-multihomed site.

However, the overall scalability of the approach requires further investigation, in the sense of the impact that these TE schemes could have on the downstream AS's routing policies. This is particularly relevant under the assumption of wide deployment of our TE schemes. In addition, both the local and global stability require additional investigation. For this reason, we have plans to develop a stability model for our TE scheme.

We have also plans to use the SRMs with reverse engineering purposes. Finally, we are currently working towards testing our self-adaptive SRMs in a real testbed.

References

[1] Y. Rekhter, T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, IETF, March 1995.

[2] S. Uhlig, V. Magnin, O. Bonaventure, C. Rapier and L. Deri, "Implications of the Topological Properties of Internet Traffic on Traffic Engineering," Proceedings of the 19th ACM Symposium on Applied Computing, Special Track on Computer Networks, Nicosia, Cyprus, March 2004.

[3] B. Quoitin, S. Tandel, S. Uhlig, O. Bonaventure, "Interdomain Traffic Engineering with Redistribution Communities," Computer Communications, 27(4), 2004.

[4] A. Akella, B. Maggs, S. Seshan, A. Shaikh, R. Sitaraman, "A Measurement-Based Analysis of Multihoming," in Proceedings of ACM SIGCOMM 2003, Karlsruhe, Germany.

[5] A. Akella, J. Pang, B. Maggs, S. Seshan and A. Shaikh, "A Comparison of Overlay Routing and Multihoming Route Control," in Proceedings of ACM SIGCOMM04, Portland, USA, August 2004.

[6] Cisco Optimized Edge Routing, http://www.cisco.com/

[7] Internap Network Services Corporation, http://www.internap.com/

[8] M. Yannuzzi, X. Masip-Bruin, E. Monteiro, "Towards Self-Adaptive Interdomain Edge Routing," accepted for publication in the IEEE Infocom Student Workshop, Miami, USA, March 2005.

[9] G. Almes, S. Kalidindi, M. Zekauskas, "A one-way delay metric for IPPM", Internet Engineering Task Force, Request for Comments 2679, September 1999.

[10] M. Yannuzzi, A. Fonte, X. Masip, E. Monteiro, S. Sánchez, M. Curado, J. Domingo, "A proposal for interdomain QoS routing based on distributed overlay entities and QBGP," in Proceedings of the First International Workshop on QoSR (WoQoSR), co-located with QoFIS'04, LNCS 3266, Barcelona, Spain, October 2004.

[11] J-Sim Homepage, http://www.j-sim.org.

[12] Infonet Suite Homepage, http://www.info.ucl.ac.be/~bqu/jsim/

[13] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services," RFC 2475, December 1998.

[14] V. Jacobson, K. Nichols and K. Poduri, "An Expedited Forwarding PHB," RFC 2598, IETF, June 1999.

[15] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group," RFC 2597, IETF, June 1999.

[16] H. Alshaer and E. Horlait, "Expedited Forwarding Delay Budget Through a Novel Call Admission Control," 3rd European Conference on Universal Multiservice Network (ECUMN' 2004), Porto, Portugal.

[17] EuQoS, "End-to-end Quality of Service support over heterogeneous networks," EU Research Project, http://www.euqos.org/