Resilient Traffic Engineering in a Transit-Edge Separated Internet Routing

Stefano Secci[a], Kunpeng Liu[b], Guruprasad K. Rao[b], Bijan Jabbari[b]
[a]LIP6, Pierre & Marie Curie University, France. [b]Communications and Networking Lab., George Mason University, VA.
E-mail: stefano.secci@lip6.fr, {kliu3,grao2,bjabbari}@gmu.edu

***Abstract*—The significant growth of Internet traffic and increase of routing tables require solutions to address Internet scalability and resiliency. A possible direction is to move away from the flat legacy Internet routing to hierarchical routing, separating edge networks from transit networks. In this paper, we study the extended traffic engineering capabilities arising in a transit-edge separated Internet routing, focusing on those multi-homed edge networks (e.g., Cloud/content providers) that aim at increasing their Internet resiliency experience. We model using game theory the interaction between distant independent edge networks exchanging large traffic volumes, with the goal of seeking efficient edge-to-edge load-balanced routing solutions. The proposed traffic engineering framework relies on a non-cooperative potential game, built upon locator and path ranking costs, that indicates efficient equilibrium solution for the edge-to-edge load-balancing coordination problem. Simulations on real instances show that in comparison to BGP and LISP we can achieve significantly higher resiliency and stability[1].**

I. Introduction

The main purpose of traffic engineering is to facilitate reliable network operation by providing methods that enhance network integrity and survivability, via routing and resource management, taking into account the occurrence of various network impairments, differentiated traffic scheduling and multi-class service provisioning [1]. The principal scope of implementation of Internet traffic engineering methods has been the intra-domain routing. Within the network of a single Internet carrier or service provider, the autonomous nature of the network has allowed the introduction of new capabilities, such as label-switching protocols, that natively allow for explicit routing and new services [2].

Within the inter-domain inter-carrier scope, instead, scalability, confidentiality and policy issues have limited reaching consensus for a systematic approach to inter-domain traffic engineering. With the current inter-domain routing protocol, the Border Gateway Protocol (BGP), levels of traffic engineering are possible manipulating attributes associated with the BGP decision process, partially fulfilling the needs of the Internet network actors (transit, content and Internet service providers) [3]. Nevertheless, BGP-based traffic engineering methods are usually applied in a try-and-hope fashion, given the impossibility to control with certainty inbound traffic, and given the uncertainty of traffic variations due to the decoupling between the communication layers.

In the current commercial Internet, we are witnessing the deployment of high access traffic bit rates (100 Gb/s interfaces) and the number of connected networks (about 36000 Autonomous Systems, ASes). Trials to perform traffic engineering for resiliency and multi-homing management via BGP are moreover amplifying the number of networks to be managed independently (about 400,000 lines in the BGP routing tables). It is well known that the scalability of the Internet, together with its acceptable performance, can be preserved by introducing hierarchical routing mechanisms. In particular, given the scale-free nature of the Internet graph with a few hub carrier networks, a two-level routing separation between transit and edge networks appears a desirable and viable solution [5]. With a transit-edge separated Internet routing, the routing table size and its loading effect on the router can be drastically reduced, efficient mobility mechanisms can be deployed, the user locator can be separated from the identifier, and the overall Internet path diversity and resiliency can be improved. In this paper, we study the novel traffic engineering capabilities emerging in a transit-edge routing separation context.

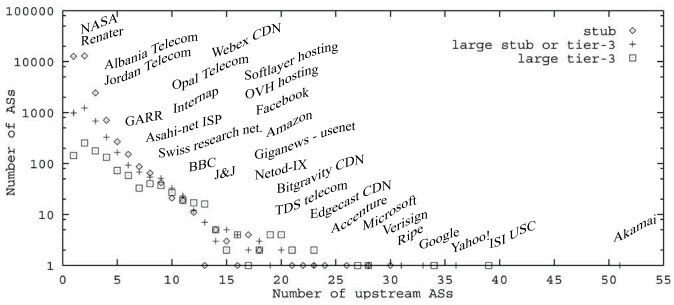

Fig. 1. Multi-homing distribution of destination ASes (as of 25 Aug., 2010).

II. Background

Currently, the Internet is composed of about 35,000 ASes. Analyzing current transit routing tables from Routeviews [12], we find that 84% of the ASes are "stub ASes", i.e., they appear only as destination ASes, last in routing table's AS paths. Stub ASes typically represent large corporations, universities, or Cloud/content providers. Looking at the historical trend of AS stub number ratio, one can appreciate that it has been linearly increasing for the past few years. Moreover, those ASes appearing at most penultimate in AS paths are about 10%; these often are large stub ASes that have fragmented their operational network into many dependent ASes, or small service providers offering Internet services in small geographical regions (called tier-3 ASes in Internetworking jargon). Finally, those appearing at most in the third from last position are about 3% and are typically large tier-3s. Stub and tier-3 ASes thus represent the large majority, about 97%, and can be considered the *edge* of the Internet. Most of

[1]This work was funded by the ONR US project "Secure Protocols and Services for Resilient Internetworking." For additional details see the TAVRI project website http://cnl.gmu.edu/TAVRI.

them are "multi-homed", i.e., have more than one upstream provider connecting them to the rest of the Internet, and about 17% of them are connected to more than two providers. Fig. 1 shows the distribution of the number of upstream ASes per stub AS, large stub or tier-3 ASes (at most penultimate position in AS paths), and large tier-3s (at most third from last position), as visible from Routeviews routing tables. We indicate the name of the organization behind some edge AS; typically, those ASes with a large number of upstream ASes are Cloud/content providers (e.g., Amazon, Google) and content delivery networks (e.g., Akamai, Edgecast), while those with lower degrees are small ISPs (e.g., Asahi-net, Albania tlc), service providers (e.g., Verisign, Internap) or research networks (e.g., GARR, Renater).

Many reasons can be behind such high degrees of multi-homing. Namely, both traffic engineering and network reliability benefit from an augmented interconnectivity. Here, Internet traffic engineering consists of controlling the direction and the load of inbound and outbound traffic from and towards the upstream ASes. At present the legacy BGP protocol offers an attribute, the local preference, and a method, the AS path prepending, to perform traffic engineering via local filtering of BGP messages. The local preference can be assigned to incoming BGP messages to rank upstream networks, while with AS path prepending one can artificially increase the AS path to distract traffic volumes toward its other providers [3] [4]. Looking at routing tables, local preferences cannot precisely be inferred, while one can notice prepended AS paths; we find that about 17.5% of the edge AS networks are actively using the path prepending, with at least 2 upstream ASes. These edge AS networks have thus strict Internet traffic engineering requirements for their services. Nevertheless, while effective, the Internet traffic engineering resulting from BGP attribute tweaking remains deficient, time-consuming and highly computational intensive for routers. It also results in an excessive fragmentation of network prefixes that is exploding the BGP routing table size: about 30% of edge AS networks announce more than 100 network prefixes. Recent detailed analysis shows that the size of the routing table can be reduced by 43% to 90% at different levels of transit-edge routing separation [9].

With transit-edge separation, the edge-to-edge routing decision is enriched: not only the best path toward the destination edge network has to be chosen, but also the best locator and/or the best egress gateway for the source edge network. Furthermore, *Internet multipath routing*, a feature largely desirable for edge AS networks for load-balancing purposes, can be enhanced. It can be implemented either using the multipath mode of BGP, available for some routers (multipath on equivalent BGP routes with even load-balancing), or with load-balancing middle-boxes. However, recent studies show that inter-AS multipath routing is practically not used today [6]. One reason is that BGP multipath brings additional instabilities to the routing system. For edge ASs, forms of stable multipath routing would be useful as the edge-to-edge path length is expected to be longer than for the global average path length.

These major aspects are also highlighted in the recent Internetworking research guidelines by the Internet Architecture Board [5]. Namely, a viable direction is to address in a scalable way the separation between the transit and the edge routing domains. Transit-edge routing separation, besides allowing important performance enhancements - such as a significant reduction of the routing table size, seamless mobility management, Internet routing security preservation, e.g., with a Locator/Identifier Separation Protocol (LISP [7]) performing packet encapsulation and decapsulation at the transit-edge borders - can largely increase the level of path diversity in Internet routing introducing gateway and locator middle-nodes. In this paper, we address the traffic engineering requirements of those 17.5% edge AS networks actively performing Internet traffic engineering with BGP. We propose a rationally justified method to coordinate the multipath routing among distant edge networks (e.g., among a tier-3 provider and a content provider) for an efficient Internet-wide load-balancing.

III. The Routing Game

We present how routing among distant edge domains can be modeled with game theory, starting with a simple game and gradually generalizing the model.

A. An introductory scenario

Let us suppose that two edge networks exchange in a stable manner a relevant amount of traffic and that, with the aim to improve their routing, they announce to each other preferences on their routing locators (as possible, e.g., with locator priorities in LISP [7]). The preferences on the locators can be due to a variety of reasons (e.g., interconnection agreements, bandwidth, observed performance), similarly to what happens with the BGP's local preference. Differently from BGP local preferences that apply to outbound traffic, *locator preferences* apply to inbound traffic. Note that in BGP, a preference for inbound traffic can be globally expressed using AS path prepending [4], which can be however discarded or ineffective in many cases.

For the sake of simplicity, let us concentrate on cases with a single locator preference per provider (instead of per gateway router), as in the multi-homing example of Fig. 2 where the network I and II have two and three upstream AS providers, respectively. In transit-edge routing separation, the egress router of each edge network has the routing choice on the ingress provider for the destination network; e.g., as currently proposed in LISP [7], using a destination-to-locator mapping system, the source network can receive the available locators for a given destination together with some additional parameters such as the locator (cost) preference. Therefore, the locator routing choice of the source network impacts a routing cost on the destination network. In a naive context, the source chooses the locator following the announced destination's preferences (e.g., minimizing its routing cost); this would be strategically acceptable in the case of two edge networks belonging to the same AS authority (e.g., a Cloud provider or content delivery network), or to two strategically dependent ASes (belonging to the same company or dependent companies). We focus, instead, on a non-naive context in which the two edge networks are independent and normally act following their own preferences first. In such a context, we can model their strategic routing interaction with non-cooperative game theory [11]. Table I shows the locator routing game setting in strategic form corresponding to the scenario in Fig. 2, where the list of strategies available to network I corresponds to the three locator-providers of network II (and conversely). Each possible strategy profile indicates the cost for network I on the left and that for network II on the right, accounting for the cost

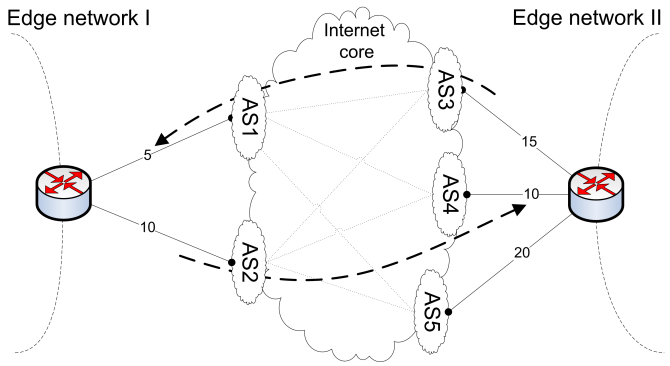

Fig. 2. Edge-to-edge routing interaction example

I \ II	AS1	AS2
AS3	**5,15**	**10,15**
AS4	**5,10**	**10,10**
AS5	**5,20**	**10,20**

TABLE I
A LOCATOR ROUTING GAME.

that each player's decision impacts on the other player, i.e., the locator cost. The profile (AS4,AS1), e.g., corresponds to the routing solution traced in Fig. 2.

Proposition III.1. *Without a coordinated routing mechanism, there is no traffic engineering incentive in following locator preferences in a transit-edge routing separation context.*

All the profiles in Table I are (pure-strategy) Nash equilibria, i.e., for each player there is no preference over the available strategies. Indeed, the game is a dummy game, which highlights that using the destination's locator preferences without a traffic engineering purpose would be a routing practice rationally not motivated. Therefore, it is of key interest to define coordination mechanisms to benefit from the novel traffic engineering capabilities beyond transit-edge locator-identifier separation. In fact, the introduction of locators for edge networks brings to a larger path diversity in Internet routing, which can undoubtedly increase the overall resiliency.

B. Coordinated joint routing

The two networks can agree in jointly routing their flows following implicit coordination equilibria of the corresponding joint routing game. This means accounting not only for the cost that the other player decision impacts on its own network as in Table I, but also for the cost of its own decision as in Table II where we simply assume (for the moment) that the locator preference applies also as a *gateway preference* for the egress direction, i.e., that the routing (cost) preference is considered valid for both the upstream and the downstream edge links - which makes sense when the two edge-to-edge flows are balanced (e.g., similar bit rates).

In Table II, the strategies have now the notation G_iL_j, where i and j indicate the gateway AS and the locator AS. In fact, now the decision is not simply on the destination's locator where to send the traffic, but also on its egress gateway; e.g., G_1L_4 is a strategy for network I that suggests to route the flow across AS1 toward AS4 on the way for the destination. Table II indicates in bold the six Nash equilibria of the corresponding

I \ II	G_3L_1	G_3L_2	G_4L_1	G_4L_2	G_5L_1	G_5L_2
G_1L_3	10,30	15,30	**10,25**	**15,25**	10,35	15,35
G_1L_4	10,25	15,25	***10,20***	**15,20**	10,30	15,30
G_1L_5	10,35	15,35	**10,30**	**15,30**	10,40	15,40
G_2L_3	15,30	20,30	15,25	20,25	15,35	20,35
G_2L_4	15,25	20,25	15,20	20,20	15,30	20,30
G_2L_5	15,35	20,35	15,30	20,30	15,40	20,40

TABLE II
JOINT ROUTING GAME

routing game. For the sake of clarity, (G_1L_5, G_4L_2) is a Nash equilibria and the equal-cost (G_2L_3, G_3L_1) is not because, for the first, both the players have no incentive to change their strategies – for I, G_2L_x strategies have a cost of $20 > 15$, for II G_3L_x and G_5L_x have a cost higher than 30, and equal to for the remaining strategies – while for the latter both have incentives to change to a strategy with a lower unilateral cost.

Among the six (pure-strategy) equilibria of Table II, the one in italic (G_1L_4, G_4L_1) is the efficient one (more precisely, Pareto-superior to the others): it represents the distrustful strategic interaction "I'll route toward your preferred locator, only if you route toward my preferred locator."

C. Setting with forward route costs

An assumption made so far is that the locator preference cost is equal to that of the gateway, i.e., the same routing cost is considered for both the upstream and the downstream flows. A more realistic assumption is that these two costs are different to each other. In fact, since the transit-edge locator-identifier separation is incrementally deployable in the legacy Internet, the edge-border routers are BGP peers of the transit-border routers. Therefore, the edge-border router can receive as many AS-paths (towards each destination's locator) as its providers, which increases the available path diversity and allows evaluating each forward gateway-locator route independently.

The edge-border router does not receive the backward paths from the destination's locators towards its network, and forward and backward paths are generally different since Internet routing is asymmetric due to routing policies. Different ingress and egress costs should model ingress and egress edge links with asymmetric properties (different paths, and also different bandwidths, delays, interconnection policies, etc). In this way, the game slightly changes, with an ingress cost for the locator, and an egress cost for the forward route. The latter can also be seen as sum of a gateway cost, generally different from the locator cost, and a transit path performance-evaluation cost. Therefore each edge network accounts for the complete gateway-locator forward route cost, while assigning loose ingress costs for the backward flows (whose route is unknown to them). It is worth stressing that while exchanging the respective costs to build the routing game, because of the routing asymmetry, an edge network should not consider the other edge's forward route cost as part of its backward cost.

Different methods can be conceived to rank Internet routes. One can use rude yet efficient methods such as the AS hop count, or one can map in the cost monitored performance along

a route to assess its resiliency. Moreover, this may be done locally in the router or externally in a ranking middlebox server (made available also by other entities than the providers) as discussed in [8]. We thus enrich the routing game with forward route costs to take benefit from the additional path diversity offered by transit-edge routing separation. This consists of considering forward route costs $c_{i,j}$ from the source toward the destination passing by the source's gateway i and destination's locator j; in the example in Fig. 2, for network I, $i \in \{1,2\}$ and $j \in \{3,4,5\}$ passing via gateway 1 and 2 towards locators 3, 4 and 5, and conversely for network II. Considering, e.g., the setting:
$\{c_{1,3}=17, c_{1,4}=13, c_{1,5}=15, c_{2,3}=10, c_{2,4}=12, c_{2,5}=15\}$
$\{c_{3,1}=22, c_{3,2}=20, c_{4,1}=25, c_{4,2}=28, c_{5,1}=22, c_{5,2}=26\}$
we obtain the form in Table III (the exponent meaning is explained hereafter), with this time a single Nash equilibrium.

Since the main purpose of edge AS networks performing multi-homing is to increase their overall Internet resiliency experience, for the presented traffic engineering context one shall consider cost functions taking into consideration the level of path diversity for each transit route (from the gateway AS to the locator AS) along with other performance criteria (e.g., the AS hop count) of the available paths. This allow coping with the fact that the number and quality of available paths between two networks or gateway nodes can change in time.

The more paths are available, the more resilient the transit route is; in case of failure along one path, alternative paths shall be available to the gateway routers. As path performance criterion, we propose the simple yet efficient AS hop count, incorporating also possible path prepending (which represents a routing preference of downstream ASes).

Let $\Omega_{i,j}$ be the set of available AS-level paths between a gateway i and a locator j, and let $L(\omega)$ be the AS hop count of the path $\omega \in \Omega_{i,j}$. We believe it is appropriate to model the set of paths along a transit route as a system of resistors in parallel, where a resistance corresponds to an AS path length, and the equivalent resistance (L_{eq}) can be computed. Lenghty paths bring more negligible contributions, and the more available paths the lower route cost:

$$c_{i,j} = \lceil A \cdot L_{eq} \rceil \quad s.t. \quad {L_{eq}}^{-1} = \sum_{\omega \in \Omega_{i,j}} L(\omega)^{-1} \qquad (1)$$

where A is an arbitrary constant.

D. Mathematical notations

The routing game can be described as $G = (X, Y; f, g) = G_s + G_d$, sum of a selfish game and a dummy game, respectively; let f and g be the cost functions, and X and Y the strategy sets, of network I and network II, respectively. Each strategy $x \in X$ or $y \in Y$ indicates the source gateway and the destination locator. The strategy set cardinality is equal to the number of source gateways $\times$ the number of destination locators. G_s considers the forward path cost only, while G_d considers backward locator cost only, impacted by the other network's routing decision – we already discussed an example of dummy game in Table I.

$G_s = (X, Y; f_s, g_s)$, is a purely endogenous game, where $f_s,\ g_s : X \times Y \to \mathbf{N}$ are the cost functions for network I and network II, respectively. In particular, $f_s(x,y) = \phi_s(x)$, where $\phi_s : X \to \mathbf{N}$, and $g_s(x,y) = \psi_s(y)$, where $\psi_s : Y \to \mathbf{N}$. For the game in Table III, e.g., consider the profile $(\widetilde{x}, \widetilde{y})$ with

I \ II	G_3L_1	G_3L_2	G_4L_1	G_4L_2	G_5L_1	G_5L_2
G_1L_3	22,37 ^{5}	27,35 ^{3}	22,40^{8}	27,43^{11}	22,37 ^{5}	27,41^{9}
G_1L_4	*18,32* ^{1}	23,30^{-1}	18,35^{4}	23,38 ^{7}	*18,32* ^{1}	23,36^{5}
G_1L_5	20,42 ^{3}	25,40 ^{1}	20,45^{6}	25,48 ^{9}	20,42 ^{3}	25,46^{7}
G_2L_3	15,37^{-2}	**20,35**$^{-4}$	15,40^{1}	20,43 ^{4}	15,37^{-2}	20,41^{2}
G_2L_4	*17,32* ^{0}	22,30^{-2}	*17,35*^{3}	22,38 ^{6}	*17,32* ^{0}	22,36^{6}
G_2L_5	20,42 ^{3}	25,40 ^{1}	20,45^{6}	25,48 ^{9}	20,42 ^{3}	25,46^{7}

TABLE III
BIDIRECTIONAL ROUTING GAME WITH FORWARD PATH COSTS

$\widetilde{x} = G_2L_3$ and $\widetilde{y} = G_4L_1$; we have:
$f_s(\widetilde{x}, \widetilde{y}) = \phi_s(\widetilde{x}) = c_{2,3} = 10$; $g_s(\widetilde{x}, \widetilde{y}) = \psi_s(\widetilde{y}) = c_{4,1} = 25$

$G_d = (X, Y; f_d, g_d)$, is a game of pure externality, where $f_d,\ g_d : X \times Y \to \mathbf{N}$, $f_d(x,y) = \phi_d(y)$ and $\phi_d : Y \to \mathbf{N}$, $g_d(x,y) = \psi_d(x)$ and $\psi_d : X \to \mathbf{N}$. Let E be the edge link set, and let $c(l'_i)$ be the routing cost across the ingress link l'_i by provider/locator i, with $l_i, l'_i \in E$. For the above example: $f_d(\widetilde{x}, \widetilde{y}) = \phi_d(\widetilde{y}) = c(l'_1) = 5$; $g_d(\widetilde{x}, \widetilde{y}) = \psi_d(\widetilde{x}) = c(l'_3) = 15$

IV. LOAD-BALANCING EQUILIBRIUM SOLUTION

In this section we concentrate on the game equilibrium properties and on our proposition to compute a multipath routing solution for edge-to-edge load-balancing.

A. Pure-strategy Equilibrium Properties and Computation

$G_s + G_d$ is a cardinal potential game [10], i.e., the incentive to change players' strategy can be expressed with a single potential function (P) for all players, and the difference in individual costs by an individual strategy move has the same value as the potential difference. G_d can be seen as a potential game too, but with null potential. Hence, the potential $P : X \times Y \to \mathbf{N}$ depends on G_s only. The exponents in the profiles of Table III, e.g., represent the corresponding potential values.

Generally, in non-cooperative games the Nash equilibrium existence is not guaranteed. As property of potential games [10], the P minimum corresponds to a (pure-strategy) Nash equilibrium and always exists. The inverse is not necessarily true, but it is easy to prove that it is due to the endogenous nature of G_s. The exponents in the example of Table III indicate the potential value corresponding to the strategy profile[2]. The Nash equilibrium is thus guided by G_s. The opportunity of using the minimization of the potential function to catch all the peering Nash equilibria represents a key advantage. It decreases the time complexity, which would have been very high for instances with many providers and locators. When there are multiple equilibria (possible with equal forward path and/or locator costs), G_d can help in selecting an efficient equilibrium in the Pareto-sense.

Pareto efficiency: Recall that the Nash equilibrium can be inefficient and far from the social optimum: the paid price is the price of anarchy due to the non-cooperative modeling of edge networks' independency. A strategy profile p is *Pareto-superior* to another profile p' if a player's cost can be decreased from p to p' without increasing the other

[2]to explicate P in calculus an arbitrary starting potential has to be chosen; we set to 0 the potential of social welfare profiles, i.e., $P(x_0, y_0) = 0 \ \ \forall (x_0, y_0) \in X \times Y | f(x_0, y_0) + g(x_0, y_0) = \min\{f(x,y) + g(x,y)\}$.

players' costs. In our routing game, locator costs affect the Pareto-efficiency (because of the pure externality of G_d); In particular, given many Nash equilibria, their Pareto-superiority strictly depends on G_d. For example, in Table III, the strategy profiles in italic are Pareto-superior to the Nash equilibrium, but are not equilibria since at least one player is interested in deviating to reduce its cost. Moreover, those underlined are the Pareto-efficient profiles of the game, and also correspond to the social optimum (which is not true in general). Hence the game has the form of a Prisoner-Dilemma game, where the players see the convenience to adopt a Nash equilibrium solution despite other non-equilibrium profiles are more efficient for both of them. Moreover, it is a good exercise to check that, if we decrease $c_{1,4}$ to 10, we obtain a second equilibrium in (G_1L_4,G_3L_2) which is Pareto-superior to the other equilibrium (G_2L_3, G_3L_2). This is due to the external effect of G_d, i.e., $c(l'_3) > c(l'_4)$.

B. Enforcing edge-to-edge load-balancing

In a transit-edge routing separation framework, it is technically possible and desirable to implement *edge-to-edge load balancing* schemes. The presence of multiple locators for the same destination radically increases the Internet path diversity available to the source network. Indeed, an egress router can dispose of a much larger path diversity than under the legacy flat-routed Internet (namely, using the multipath mode of BGP) - more precisely, a path diversity approximately proportional with the number of available locators.

A generic way to implement load-balancing is to arbitrary assign a percentage weight to each route-strategy, indicating the distribution of egress traffic toward the destination along that route. Alternatively, a percentage weight can be assigned to the locators by the destination network as its desired distribution for the upstream network(s). Both ways are technically possible and somehow equivalent; the latter is in fact more scalable (and is in fact the way to enforce inbound load-balancing currently included in the LISP specification [7]). We are thus interested in defining a method to arbitrary set such traffic distribution weights that is strategically acceptable.

The selection of n multiple equilibria could result in an even load-balancing distribution (at most $1/n$ load on each locator). Although acceptable, it is desirable to rank the equilibria following some rational criteria better considering the game dynamics so as to better meet routing stability requirements.

C. The potential as an equilibrium refinement tool

In potential games, the potential value qualifies the profile propensity to reaching equilibrium and predicts the behavior of the potential game [10]: the lower it is, the finer the profile is. However, all the equilibria of G have the same potential and therefore the potential value can not help in ranking the available pure-strategy equilibria. Moreover, remember that the occurrence of multiple equilibria in G is not guaranteed - it happens only with equal egress and/or ingress costs - and may be a rare event for small instances; in these cases, load-balancing could not be implementable.

Since load balancing is a key feature in an transit-edge separation context to improve Internet resiliency, it is desirable to increase the number of strategy profiles in the routing solution. The potential value can in fact help in extending the equilibrium set including also those profiles that are not pure-strategy equilibria, but that have good chances of becoming so in future settings. For example, in Table III, the profiles having a potential equal to -2 have a good chance to become an equilibrium after slight changes of one or a few cost components; such profiles can be considered as better strategy profiles than other profiles with a higher potential.

With the aim of increasing the path diversity of the routing solution, we can thus elevate those profiles that are not Nash equilibria, but that have a very low potential, to the equilibrium status and include them in the routing solution. This corresponds to selecting as routing equilibrium all the strategy profiles that have a potential equal or below a pre-computed threshold (i.e., not only those with the minimum potential). Since the maximum and the minimum potential values change with the game configuration, the threshold can be set accounting for the statistical potential distribution. An acceptable threshold corresponds to the first quartile of the potential distribution. For example, in Table III, the first quartile potential is equal to 1; therefore, the routing solution includes seven strategy profiles with a potential of 0 and less. The threshold computation can, however, be adapted to the problem instances; for very large instances, more conservative threshold levels than the first quartile could be used.

A further implicit step that is rationally acceptable is to restrict the equilibrium set only to those that are not Pareto-inferior to any other selected equilibrium; in Table III, this corresponds to discard (G_2L_3, G_3L_2) from the solution. Finally, we propose to use the potential value of the remaining equilibria as the index to set the load-balancing distribution, so that lower potential values bring to a higher load ratio.

Let $\chi \in X \times Y$ be the set of the equilibria kept as solution; τ the potential threshold; $P(x, y)$ the potential value of $(x, y) \in \chi$; $b_{\widetilde{x}}$ and $b_{\widetilde{y}}$ the load-balancing ratio for strategy $\widetilde{x} \in X$ and $\widetilde{y} \in Y$, for network I and network II, respectively. We propose to set the load-balancing ratios as the proportional weight, with respect to the distance from the potential threshold, of the unilateral strategy over all the available strategy profiles:

$$b_{\widetilde{x}} = \frac{\sum_{(x,y)\in\chi}^{x=\widetilde{x}} [1 + \tau - P(x, y)]}{\sum_{(x,y)\in\chi} [1 + \tau - P(x, y)]}, \quad \forall(\widetilde{x}, y) \in \chi \qquad (2)$$

and the dual for $b_{\widetilde{y}}$. We can in this way fairly assign higher weights to those unilateral strategies that cover many solution equilibria. For example, in Table III, we obtain the load-balancing solution $b_{G_2L_3} = 8/16 = 50\%$ and $b_{G_2L_4} = 8/16 = 50\%$ for network I, and $b_{G_3L_1} = 37.5\%$ and $b_{G_3L_2} = 25\%$ and $b_{G_5L_1} = 37.5\%$ for network II.

V. Simulation results

We simulated the edge-to-edge interconnection of two sample ASes, AS 12182 (Internap) and AS 4685 (Asahi-Net ISP), that have had between 6 and 12 AS providers in the last few years. We chose these two ASes because both of them actively use AS path prepending at different levels with most of their providers, i.e., both perform actively Internet traffic engineering and would benefit from our framework. Forward path costs and locator cost need to be on similar scales because of the Pareto-superior condition; hence we set $A = 50$ in (1) to have similar maximum costs in worst case scenarios (with very lengthy AS paths). We used Routeviews [12] routing tables

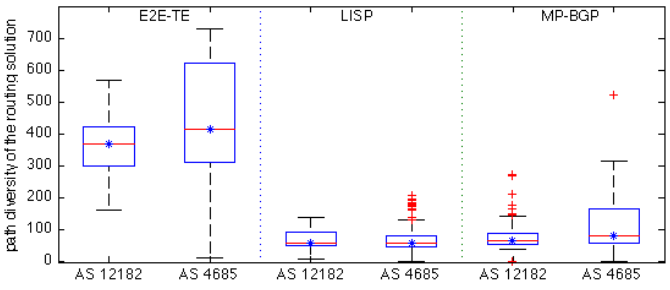

Fig. 3. Boxplot statistics of the solution's path diversity.

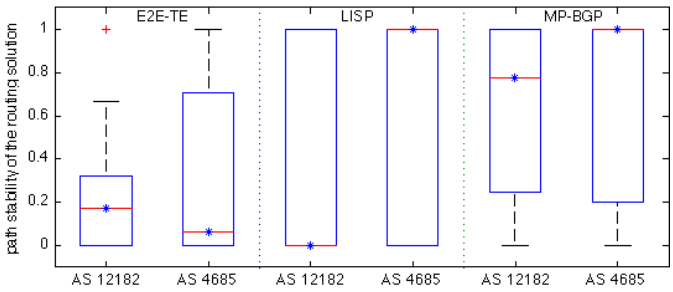

Fig. 4. Boxplot statistics of the solution's routing stability.

to qualify the AS graph, path prepending, and path diversity between gateways and locators (i.e., Ω). We set the locator cost to the detected path prepending amount to emulate a realistic configuration behavior. We used 197 successive 3-day spaced routing tables from Jan. 2009 to Aug. 2010, so as to emulate successive game settings (providers, AS paths and path prepending often change, indeed). Datasets and MATLAB codes are available in [13].

We compare our framework ('E2E-TE') to the multipath BGP solution ('MP-BGP') and to the normal LISP solution (as the naive case of Sect. III-A), with respect to the path diversity (Fig. 3) and route stability (Fig. 4), hence the solution resiliency. The routing cost results are not plotted due to space limitations, however, they do not show major differences between the three methods. We use boxplots to display statistical properties (each box, between the min. and the max., displays the first quartile, the median with a '*', third quartile).

Fig. 3 shows how many diverse AS-paths are available along the selected gateway-to-locator transit routes, for both routing directions from AS 12182 and AS 4685 (opportunely weighted accordingly to the load-balancing distribution, if any), and for the three solution methods, respectively. While the analysis of routing cost does not show relevant differences, one can appreciate how important improvements can be reached in terms of Internet reliability: we pass from a median of about 50 paths with both MP-BGP and LISP to a median around 400 with our approach[3]. This shows that resiliency route cost functions as intuitive and simple as (1) can allow reaching significant improvements with respect to legacy protocols.

Fig. 4 shows what percentage of traffic has been moved at each new solution. The higher it is, the less stable the previous solution can be considered (an instability of 1 indicates that 100% of the traffic volume has be rerouted across different paths). MP-BGP shows a quite high instability, which is in fact not a surprise, with a median above 70%. LISP shows a very high variance and opposite behaviors for the two networks, this probably relates to the fact that AS 12182 reconfigures much more often the path prepending than AS 4685 for traffic engineering purposes. All in all, our method clearly offers a more resilient solution in terms of Internet routing stability with (a median of) less than 10% of the traffic rerouted at each new reconfiguration.

VI. Conclusions

The legacy flat-routing approach to Internet routing, under which the source network decides the AS path directly to the destination network, is showing all its deficiencies in terms of scalability and resiliency. Placing intermediate gateways and locators separating edge networks from transit carrier networks can allow important performance improvement.

In this paper, we study the novel traffic engineering capabilities arising in a transit-edge routing separation context. We model the routing interaction between independent edge networks with non-cooperative game theory. We define a strategically rational approach to coordinate the routing of equivalent traffic flows following routing equilibria, which results in a fine-selected edge-to-edge load-balancing.

We experimentally show that our solution outperforms the current practice, offering far more resilient solutions also with respect to the basic routing mode of the LISP protocol currently under standardization. Our approach brings to solutions with a much higher resiliency in terms of achievable transit path diversity and routing stability. In particular, our simulation for an illustrating case shows 4-times more stable multipath routing solutions with 5-times larger path diversity[4]. Our work represents an important step toward the definition of novel Internet traffic engineering methods for edge networks, where content and computing services (the "Clouds") are located.

References

[1] D. Awduche et al., "Overview and Principles of Internet Traffic Engineering," RFC 3272, 2002.
[2] D. Awduche, B. Jabbari, "Internet traffic engineering using multi-protocol label switching (MPLS)," *Computer Networks*, 2002.
[3] B. Quoitin et al., "Interdomain traffic engineering with BGP," *IEEE Communications Magazine*, Vol. 41, No. 5, pp: 122-128, 2003.
[4] R. Gao et al., "Interdomain ingress traffic engineering through optimized AS-path prepending," in *Proc. of Networking 2005*.
[5] D. Mayer, L. Zhang, K. Fall, "Report from the IAB Workshop on Routing and Addressing," RFC 4984, Sept. 2007
[6] E. Elena, J.-L. Rougier, S. Secci, "Characterisation of AS-level Path Deviations and Multipath in Internet Routing", in *Proc. of NGI 2010*.
[7] D. Farinacci, V. Fuller, D. Mayer, D. Lewis, "Locator/ID Separation Protocol (LISP)," draft-ietf-lisp-08, Aug. 2010.
[8] D. Saucez et al., "Interdomain Traffic Engineering in a Locator/Identifier Separation Context," in *Proc. of INM 2008*.
[9] Y. Wang, J. Bi, J. Wu, "Empirical analysis of core-edge separation by decomposing Internet topology graph," in *Proc. of GLOBECOM 2010*.
[10] D. Monderer, L.S. Shapley, "Potential Games", *Games and Economic Behavior*, Vol. 14, No. 1, May 1996, Pp: 124-143.
[11] R.B. Myerson, *Game Theory: Analysis of Conflict*, Harvard Univ. Press.
[12] Routeviews website: www.routeviews.org
[13] Details, datasets and codes website: http://cnl.gmu.edu/TAVRI

[3] It is worth mentioning that these can be considered too high numbers for real cases; we indeed counted all the loop-free available paths collected exploring Routeviews tables; in reality, this number is expected to be lower due to policy filtering and limited visibility.

[4] More details not included due to space limits are given in [13].