Computer Networks 57 (2013) 976–989

Contents lists available at SciVerse ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Efficient inter-domain traffic engineering with transit-edge hierarchical routing

Stefano Secci [a,*], Kunpen Liu [b], Bijan Jabbari [b]

[a] University Pierre and Marie Curie, LIP6, 4 place Jussieu, 75005 Paris, France
[b] George Mason University, Fairfax, VA 22030-4444, USA

ARTICLE INFO

Article history:
Received 4 July 2012
Received in revised form 28 October 2012
Accepted 22 November 2012
Available online 30 November 2012

Keywords:
Internet routing
Traffic engineering
Transit-edge routing
LISP
Game theory

ABSTRACT

The relentless growth of Internet, which has resulted in the increase of routing table sizes, requires consideration and new direction to address Internet scalability and resiliency. A possible direction is to move away from the flat legacy Internet routing to hierarchical routing, and introduce two-level hierarchical routing between edge networks and across transit networks. In this way, there is also an opportunity to separate the routing locator from the terminal identifier, to better manage IP mobility and mitigate important routing security issues. In this paper, we study the extended traffic engineering capabilities arising in a transit-edge hierarchical routing, focusing on those multi-homed edge networks (e.g., Cloud/content providers) that aim at increasing their Internet resiliency experience. We model the interaction between distant independent edge networks exchanging large traffic volumes using game theory, with the goal of seeking efficient edge-to-edge load-balancing 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 the available standard protocols such as BGP and LISP, we can achieve a much higher degree of resiliency and stability.[1]

1. Introduction

Internet traffic engineering is an important part of network design and engineering that deals with performance evaluation and optimization issues of operational IP networks. 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 [2]. 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 [3].

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 through manipulating attributes associated with the BGP decision process, partially fulfilling the needs of the Internet network actors (transit, content and Internet service providers) [4]. Nevertheless, BGP-based traffic engineering methods are usually applied in a try-and-hope fashion, given the impossibility to control inbound traffic with certainty, and given the

* Corresponding author. Tel.: +33 144273678.
E-mail address: stefano.secci@lip6.fr (S. Secci).

[1] A preliminary version of this paper has been presented at the 2011 IEEE Int. Conference on Communications (ICC 2011) [1].

1389-1286/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2012.11.012

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 42,000 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 430,000 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 context involving transit and edge networks appears a desirable and viable solution [6]. With a transit-edge hierarchical routing, the routing table size and its loading effect on the router can be drastically reduced, efficient mobility mechanisms can be deployed, the IP terminal's global 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 hierarchical routing context. In Section 2 we present the novel routing context. Section 3 shows how we model the routing interaction among independent edge networks with non-cooperative game theory. In Section 4 we define how efficient edge-to-edge load-balancing routing solutions can be built upon the routing equilibria. Section 5 reports the performance evaluation of our proposition for realistic settings. In Section 6 we discuss how the game modeling and the load-balancing solution can be generalized to multiple networks. Section 7 contains implementation aspects with the Locator–Identifier Separation Protocol (LISP). Section 8 concludes the paper.

2. Background

Currently, the Internet is composed of about 42,000 ASes. Analyzing recent transit routing tables from Routeviews [18], we find that roughly 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 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.[2] 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 destination networks, while with AS path prepending one can artificially increase the AS path to distract incoming traffic volumes from some providers [4,5]. 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 two 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 43–90% at different levels of transit-edge routing separation [11].

With transit-edge hierarchical routing, 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 [7]. One reason is that BGP multipath brings additional instabilities to the routing system. For edge ASes, forms of stable multipath routing would be useful as the edge-to-edge

[2] It is worth noting that the multi-homing degree is likely increasing in time and the figure refers to Aug. 2010; also note that the number of vintage points used by Routeviews is limited to about a dozen and this is a partial view.

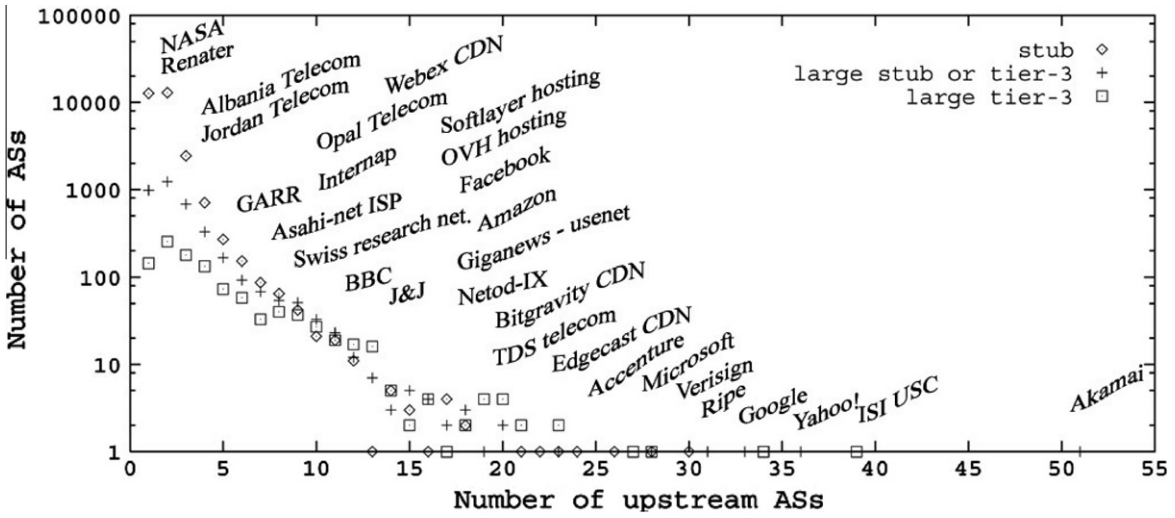

Fig. 1. Multi-homing distribution of destination ASes (as of 25 August 2010).

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 [6]. Namely, a viable direction is to address, in a scalable way, the hierarchical routing between the transit and the edge routing domains. Transit-edge hierarchical routing, 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 [8]) performing packet encapsulation and decapsulation at the transit-edge borders – can largely increase the level of path diversity in Internet routing by introducing gateway and locator middle-nodes.

We define an Internet traffic engineering framework to efficiently manage the additional edge-to-edge path diversity arising in a transit-edge hierarchical routing context. 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.

3. The routing game

We present how routing among distant edge domains in a transit-edge hierarchy can be modeled by non-cooperative game theory, starting with a simple game, then introducing the game properties and generalizing the model.

3.1. 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 [8]). 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 [5], which can be however discarded or ineffective in many cases (e.g., when the upstream AS uses adverse local preferences).

Table 1
A locator routing game.

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

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 networks I and II have two and three upstream AS providers, respectively. In transit-edge hierarchical routing, 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 [8], 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 locator cost on the destination network; this cost can express a network cost to use that link, in monetary terms, or in terms of performance level, reliability, load, similarly to what done with local preference in BGP, or with link weights in OSPF or ISIS link-state routing protocols.

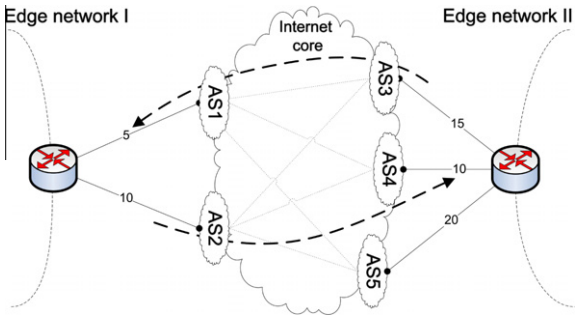

Fig. 2. Edge-to-edge routing interaction example.

In a naive context, the source chooses the locator following the announced destination's preferences (e.g., minimizing its routing locator 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 [15]. Table 1 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 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 3.1. *Without a coordinated routing mechanism, there is no traffic engineering incentive – e.g., locator priorities or weights – in following locator preferences in a transit-edge hierarchical routing context.*

All the profiles in Table 1 are (pure-strategy) Nash equilibria,[3] 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.[4] 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.

Table 2
Joint routing game.

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

3.2. 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 1, but also for the cost of its own decision as in Table 2 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 2, 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 2 indicates in bold the six Nash equilibria of the corresponding routing game.[5]

Among the six (pure-strategy) equilibria of Table 2, 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".

3.3. 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

[3] Let (S, f) be a game with two players, where S_i is the strategy set for player i, $S = S_1 \times S_2$ is the set of strategy profiles and $f = (f_1(x), f_2(x))$ is the payoff function for $x \in S$. Let x_i be a strategy profile of player i and x_{-i} be a strategy profile of all players except for player i. When each player $i \in 1,2$ chooses strategy x_i resulting in strategy profile $x = (x_1, x_2)$ then player i obtains payoff $f_i(x)$. Note that the payoff depends on the strategy profile chosen, i.e., on the strategy chosen by player i as well as the strategies chosen by all the other players. A strategy profile $\overline{x} \in S$ is a Nash equilibrium if no unilateral deviation in strategy by any single player is profitable for that player, that is $\forall i, x_i \in S_i : f_i(\overline{x}_i, \overline{x}_{-i}) \geqslant f_i(x_i, \overline{x}_{-i})$.

[4] Note that on this matter, the LISP specification highlights the traffic engineering capabilities beyond locator priorities, but does not propose any traffic engineering procedure because of being out of scope [8].

[5] 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.

(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.[6] 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 crude 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 [9]. We thus enrich the routing game with forward route costs to take benefit from the additional path diversity offered by transit-edge hierarchical routing. 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 3 (the exponent is explained hereafter), depicted in Fig. 3 – where directional costs are indicated close to the egress point over each link line (i.e., the cost close to a node along a link line indicates the cost to exit that node along the corresponding egress link); note that the common cost among all the routes passing through the same gateway in practice becomes the gateway cost to which is appended the remaining transit subpath cost (this gateway cost can be seen as an inverse BGP local preference) – with this time a single Nash equilibrium.[7]

Table 3
Bidirectional routing game with forward path costs.

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$^{(7)}$

3.3.1. Forward cost function design

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 allows 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.

Let Ω_{ij} 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 $\omega \in \Omega_{ij}$. 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 a path length measure, and the equivalent resistance (L_{eq}) can be computed. The path length measure can be the AS hop count as is done with one of the BGP rules, or an equivalent distance expressing performance and policy characteristics. With an equivalent resistor-like global length metric, the more paths there are between two edge network border routers, the lower the equivalent length. Lengthy paths bring more negligible contributions, and the more available paths the lower route cost we get. As the equivalent resistor, the equivalent length is be computed as: $\frac{1}{L_{eq}} = \sum_{\omega \in \Omega_{ij}} \frac{1}{L(\omega)}$. Therefore, the routing metric between border router i and border router j can be computed as:

$$c_{ij} = \lceil A \cdot L_{eq} \rceil = \left\lceil A \left(\sum_{\omega \in \Omega_{ij}} \frac{1}{L(\omega)} \right)^{-1} \right\rceil \quad (1)$$

where A is an arbitrary scaling constant.[8] Fig. 4 shows (1) for an example of fives paths, of length 2, 3, 4 and 5 for the first four paths, and of variable length for the fifth one. Certainly, other cost functions can be conceived, and

[6] The same would stand in a future Internet scenario with a BGP-free Internet core where other connection-oriented technologies would still take into account forms of AS-paths to identify tunnel/circuit/flow routing (as, e.g., under the distributed provider alliance control-plane proposed in [12], or under forms of cross-provider OpenFlow unified centralized forwarding plane [13]).

[7] For example, consider the profile (G_1L_3, G_3L_1): 22 is the cost for network I, i.e., the sum of the egress routing cost via gateway 1 (13), the transit forward path cost toward network II's locator 3 (4), and the ingress locator cost via locator 1 (5). Other cost components are computed following the same logic.

[8] For example, with $A = 1$ and using as path length metric the AS path hop count, for the above mentioned example, $c_{1,3} = 17$ can correspond to one path of 17 AS hops, or to two paths of 34 hops, etc.

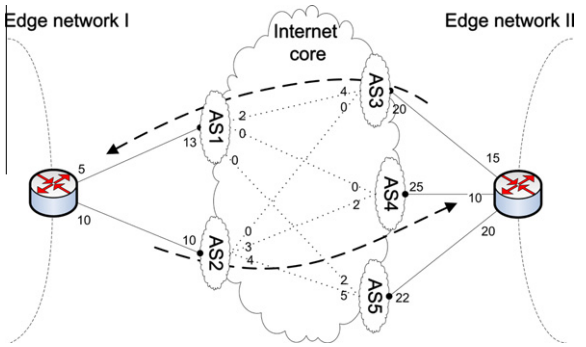

Fig. 3. Edge-to-edge routing interaction example with forward path costs.

functions of different players can be different to each other; it is worth noting, however, that in order to maintain the good game properties explained hereafter, the different player functions have to be independent of each-other.

3.3.2. On direct edge-to-edge interconnections

The resulting traffic engineering setting with forward path cost assumes edge networks are connected through transit networks. Some measurements on real data claim that, in terms of traffic volume, before 2010, the majority of Internet traffic was routed directly between edge networks, i.e., without using transit networks [10]. Our generic mathematical modeling and routing solution do encompass situations in which there are no transit networks, which simply corresponds to not including forward path cost components in the game setting, without any loss in terms of equilibrium properties and computation.

3.4. 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 (extending somehow the usage of BGP local preferences), impacted by the other network's routing decision (not taken into account in any form by the legacy BGP decision process) – we already discussed an example of dummy game in Table 1.

$G_s = (X, Y; f_s, g_s)$, 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 ($\mathbf{N}$ is the integer set). 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 3, e.g., consider the profile $(\tilde{x}, \tilde{y})$ with $\tilde{x} = G_2L_3$ and $\tilde{y} = G_4L_1$; we have:

$$f_s(\tilde{x}, \tilde{y}) = \phi_s(\tilde{x}) = c_{2,3} = 10$$

$$g_s(\tilde{x}, \tilde{y}) = \psi_s(\tilde{y}) = c_{4,1} = 25$$

$G_d = (X, Y; f_d, g_d)$, 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:

$$f_d(\tilde{x}, \tilde{y}) = \phi_d(\tilde{y}) = c(l'_1) = 5$$

$$g_d(\tilde{x}, \tilde{y}) = \psi_d(\tilde{x}) = c(l'_3) = 15$$

4. 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.

4.1. Pure-strategy equilibrium properties and computation

$G_s + G_d$ is a cardinal potential game [14], 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 3, e.g., represent the corresponding potential values.[9]

Generally, in non-cooperative games the Nash equilibrium existence is not guaranteed. As property of potential games [14], the P minimum corresponds to a (pure-strategy) Nash equilibrium and always exists. The inverse is not necessarily true, but the next theorem proves that it is true for G.

Theorem 4.1. *Every (pure-strategy) Nash equilibrium of G corresponds to a minimum of P.*

Proof. If (x^*, y^*) is an equilibrium, $P(x^*, y^*) \leqslant P(x, y^*)$, $\forall x \in X$. But, given a reference potential profile (x_0, y_0): $P(x^*, y^*) = \phi_s(x^*) - \phi_s(x_0)$ and $P(x, y^*) = \phi_s(x) - \phi_s(x_0)$, $\forall x \in X$. Thus $P(x^*, y^*) \leqslant P(x, y^*)$, $\forall x \in X$, is equivalent to $\phi_s(x^*) - \phi_s(x_0) \leqslant \phi_s(x) - \phi_s(x_0)$, $\forall x \in X$, that is $\phi_s(x^*) \leqslant \phi_s(x)$, $\forall x \in X$. Hence x^* is a minimum for ϕ_s. Idem for y^*. So $P(x^*, y^*) = 0$, that is a minimum of P. □

The exponents in the example of Table 3 indicate the potential value corresponding to the strategy profile.[10]

The Nash equilibrium is thus guided by G_s. The opportunity of using the minimization of the potential function to catch all the Nash equilibria represents a key advantage. It

[9] This decomposition is characterized for the general case in Appendix A.

[10] 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., $P(x_0, y_0) = 0 \ \ \forall (x_0, y_0) \in X \times Y | f(x_0, y_0) + g(x_0, y_0) = \min\{f(x, y) + g(x, y)\}$; for example, in Table 3 there are two such null-potential starting profiles, (G_2L_4, G_3L_1) and (G_2L_4, G_5L_1). Then, all the other potential values can be determined following unilateral moves and adding to the null potential the difference between the costs of the moving player: consider a move from (G_2L_4, G_3L_1) to $(G_2L_3, G_3L_1), 15 - 17 = -2$ is the potential difference hence the potential value of the profile is $0 - 2 = -2$; then, moving to $(G_2L_3, G_3L_2), 35 - 37 = -2$ hence the potential value is $-2 - 2 = -4$, and so on so forth.

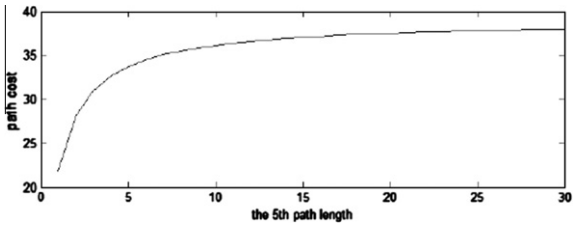

Fig. 4. Example of the path cost function (1) – $A = 50$.

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.

4.1.1. 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 players' costs. The *Pareto-frontier* contains the *Pareto-efficient* profiles, i.e., those not Pareto-inferior to any other. 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 3, 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., $c(l'_3) > c(l'_4)$.

4.2. Enforcing edge-to-edge load-balancing

In a transit-edge hierarchical routing 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. Moreover, with forward path ranking by the edges, load-balancing is particularly desirable to avoid possible routing oscillations; in fact, in the case multiple networks use the same path cost function and react synchronously to transit path performance degradation (assigning them higher costs), if single path routing is used the single path is likely to suffer from performance loss in turn because of traffic overload, leading to possible persistent routing oscillations. A systematic yet fine-selected load-balancing scheme can prevent from these events affecting the Internet routing stability.

A generic way to implement load-balancing is to arbitrarily assign at the source 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 [8]). 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.

4.3. The potential as an equilibrium refinement tool

In our framework, the important question is: what is the strategically acceptable load balancing distribution technique for edge-to-edge flows? Theoretically, an immediate answer to the question is to compute mixed strategy equilibria; however, for potential games they correspond to pure-strategy equilibria (see Appendix B).

In potential games, the potential value qualifies the profile propensity to reaching equilibrium and predicts the behavior of the potential game [14]: the lower it is, the finer the profile is. However, as of Theorem 4.1, all the equilibria of G have the same potential and therefore the potential value cannot 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[11] – 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 a transit-edge routing 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 3, 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.

[11] For the example of Table 3, multiple equilibria appear if $c_{1,4} = c_{2,3} = 10$, hence (G_1L_4, G_3L_2) as second equilibrium, or if $c_{5,1} = c_{3,2} = 20$, hence (G_1L_4, G_5L_1) as additional equilibrium; note that both the new equilibria are Pareto-superior to the incumbent one (G_2L_3, G_3L_2).

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 3, 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 3, this corresponds to discard (G_2L_3, G_3L_2) from the solution (even if it is the single pure-strategy equilibrium). 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.

4.4. Load-balancing distribution computation

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 $(x,y) \in \chi$; $b_{\tilde{x}}$ and $b_{\tilde{y}}$ the load-balancing ratio for strategy $\tilde{x} \in X$ and $\tilde{y} \in Y$, 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:

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

$$b_{\tilde{y}} = \frac{\sum_{(x,y)\in\chi}^{y=\tilde{y}} [1+\tau-P(x,y)]}{\sum_{(x,y)\in\chi} [1+\tau-P(x,y)]}, \quad \forall(x,\tilde{y}) \in \chi \tag{3}$$

The first is the load on first player's strategies that can be unilaterally computed by the first player, and dually the second can be unilaterally computed by the second player, implicitly and without the need of any signaling between the two players. We can in this way fairly assign higher weights to those unilateral strategies that cover many solution equilibria.[12]

The routing solution is summarized below:

Algorithm 1. Load-balancing distribution computing steps

1: compute the potential value vector of the game, its minimum and its potential threshold;
2: select all the profiles with a potential equal to or minor than the threshold;
3: apply the Pareto-restriction of the profile set; if empty, keep all the profiles;
4: compute the corresponding load-balancing distribution for the remaining profiles.

5. Performance evaluation

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 and locator costs 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 [18] routing tables 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 January 2009 to August 2010, so as to emulate successive game settings (providers, AS paths and path prepending often change, indeed). Datasets and MATLAB codes are available in [19].

In the following, we evaluate the performance of our solution (marked 'E2E-TE'). First, we characterize the equilibrium set (Fig. 5) and load-balancing (Fig. 6) dynamics given by our solution. Then, we compare it with the multipath BGP solution ('MP-BGP') and with the solution that one would obtain with normal LISP (marked with 'LISP' – i.e., the naive case presented in Section 3.1), with respect to the routing cost (Fig. 7), path diversity (Fig. 8) and routing stability (Fig. 8) and – hence resiliency – of the solution. For LISP, we used the same locator (priority) cost adopted by E2E-TE. As suggested by current IETF developments, LISP chooses the locator with higher priority (i.e., lower cost); moreover, when the locator cost is equal the edge upstreaming AS, LISP chooses the locator randomly, and when the locator costs are equal, a single one is chosen randomy. Finally, for the sake of a fair comparison between LISP and MP-BGP, we consider that if multiple equal-length path are available to the locator, LISP uses multipath (i.e., even load-balancing) among them. 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).

[12] For example, in Table 3, 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. Note that without the Pareto restriction we would obtain, instead, $b_{G_1L_4} = 3/25 = 12\%$ and $b_{G_2L_3} = 15/25 = 60\%$ and $b_{G_2L_4} = 7/25 = 28\%$ for network I, and $b_{G_3L_1} = 24\%$ and $b_{G_3L_2} = 52\%$ and $b_{G_5L_1} = 24\%$ for network II, hence a more fragmented distribution.

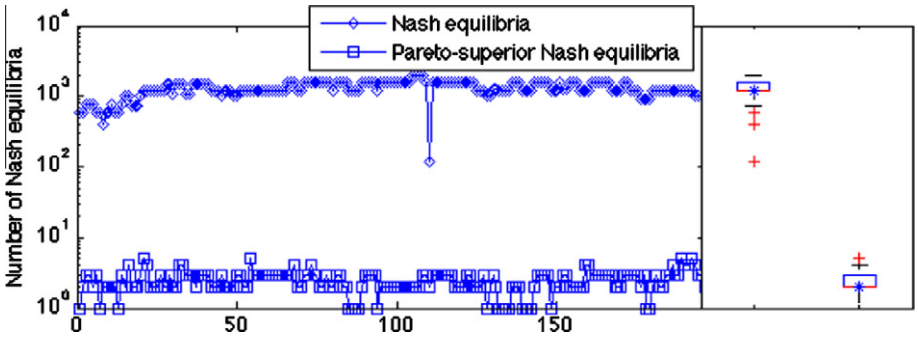

Fig. 5. Nash equilibria dynamics.

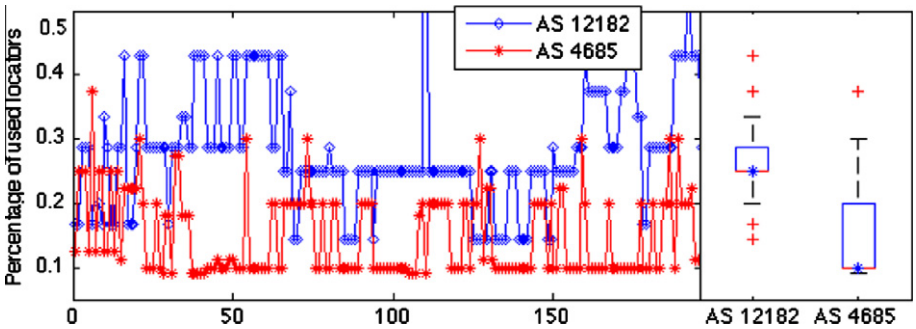

Fig. 6. Load balancing dynamics.

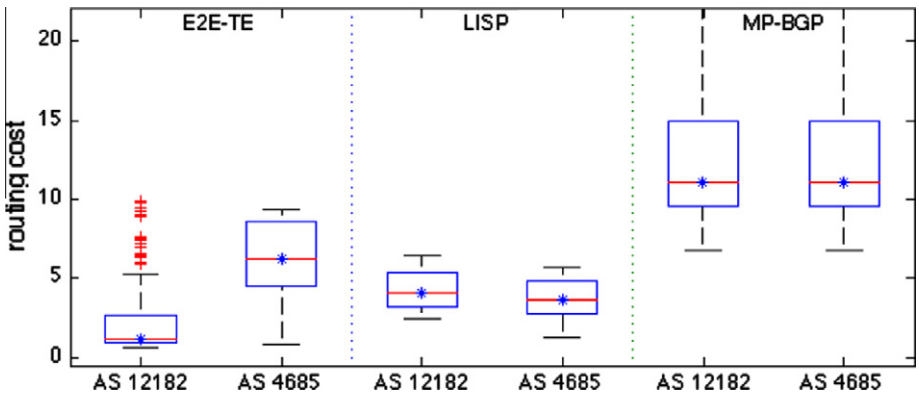

Fig. 7. Boxplot statistics of the solution's routing cost.

5.1. Equilibrium set and load-balancing dynamics

Fig. 5 shows the dynamics of the number of equilibria of the routing solution for all the iterations, and the boxplot statistics in the right side. We show also the Pareto-restriction of the equilibrium set. All in all we can see that we have around 1000 equilibria, of which around 'only' 3 of them are Pareto-superior to all the others. The Pareto-restriction is useful to get rid of those profiles that, even if considered as equilibria because of their low potential, show a strategically inefficient allocation. This equilibrium solution refinement finally produces a routing solution that has a very selective load-balancing toward a few locators. This aspect is described in Fig. 6, that shows the ratio of locators that are used by the load-balancing solution, knowing that the considered edge networks have between 6 and 9 locators, and between 8 and 12 locators, respectively. Therefore, the load-balancing solution brings to about a median of 1/4 of the locators used for one network and of 1/9 for the other.

5.2. Routing cost

Fig. 7 depicts the routing cost statistics, showing that while MP-BGP offers an inefficient solution with a cost about twice higher, there are no major differences between

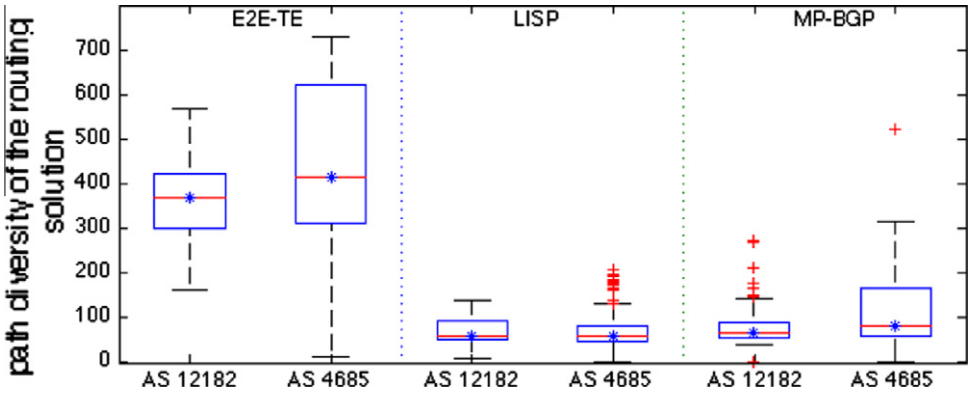

Fig. 8. Boxplot statistics of the solution's path diversity.

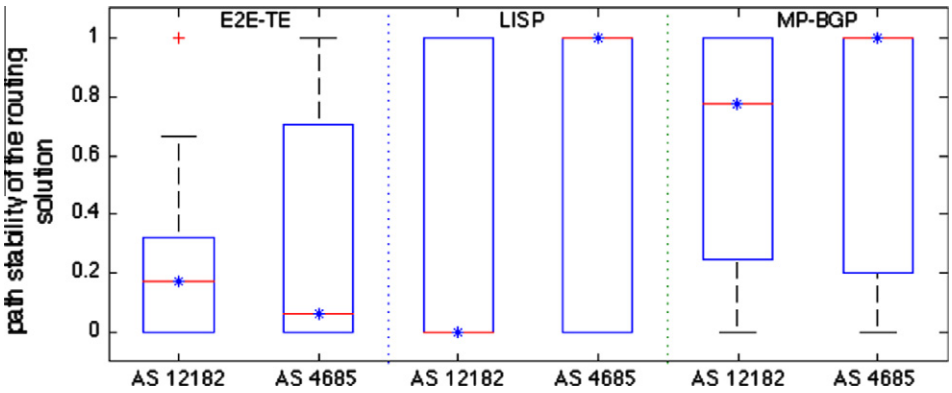

Fig. 9. Boxplot statistics of the solution's routing stability.

our method and the LISP solution based on locator cost minimization. This reflects that our approach does not merely follow the minimization of the routing cost, but rather accounts for the strategic pertinence of the routing profile.

5.3. Path diversity

Fig. 8 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.[13] This shows that resiliency route cost functions as intuitive and simple as (1) can allow reaching significant improvements with respect to legacy protocols.

5.4. Routing stability

Fig. 9 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.

6. Generalization to *n*/networks

We restricted our traffic engineering framework to a bilateral routing coordination between two edge net-

[13] 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.

works-players. In this section, we show how it can be easily generalized to more than two networks, and we propose additional traffic engineering enhancements.

6.1. Game extension to multiple players

From a strategic perspective, the extension to more than two networks implies that a subset of networks implicitly coordinate the bilateral routing of the respective flows. In order to have this extended coordination justified, the traffic among the participating networks have to be significant. The resulting load-balancing solution is thus computed including in the game only those networks with which an edge network exchange significant amount of traffic; the other edge networks with which the amounts are negligible can be discarded in the modeling.

It is worth noting that if all the Internet edge (stub) networks were to be included in the modeling, we would obtain a game with an infinite number of players. Besides being untreatable, this would also be ineffective since we can more pragmatically restrict the game modeling to the group of those edge networks with significant reciprocal traffic volume exchanges. Moreover, such a systematic approach would need to index all the networks, which would be impracticable given the rapid and decentralized evolution of the Internet ecosystem.

6.2. Game strategies, cluster size and complexity concerns

From a game setting perspective, supposing to have a set of N networks, in the n/player game each strategy of each player has to include routing indications for all the $n-1$ egress flows. Let X_i be the strategy set of the ith player, $i \in N$, and let P_i the number of providers/locators of network i. Then, $|X_i| = \Pi_{(i,j)\in N\times N}^{i\neq j}(P_i \cdot P_j)$. For example, in a case of 3 networks with 2, 3 and 4 providers each, respectively, we obtain a set of 48 strategies for player I, 72 for player II and 96 for player III, with a bi-dimensional potential array of 331 776 elements. A strategy $x \in X_1$ may, e.g., be $x = G_2L_3, G_1L_9$ indicating to route the flow from network I to network II via the gateway 2 and the locator 3 and the flow from network II to network III via the gateway 1 and the locator 9.

Nevertheless, for larger instances with a high number of networks, one may obtain untreatable instances. Let us suppose a large case of m networks, each one with k providers/locators; we obtain sets of k^{2m-1} strategies elements. For large settings (e.g., $k > 5$ and $m > 50$) there may be thus need to define a more scalable and less precise modeling. Very large instances would be, however, likely uncommon; in any case, a possible technical solution would be to implement multi-cluster settings with per-cluster edge link reservation levels and routing costs (somehow similarly to multi-level topologies for link-state Interior Gateway Protocols).

6.3. Notation

Therefore, the extended edge-to-edge routing game is a straightforward extension of the 2-player game:

- G_s and G_d maintain exactly the same structure,
- the number of strategies increases due to the higher number of flows, gateway and locators.

The generalized game is $G = (X_1, \ldots, X_n; f^1, \ldots, f^n)$, where $f^i = f_s^i + f_d^i$ is the cost function of the i^{th} network in the cluster such that $f^i : \prod X_{j_{j\in N}} \to N, f_c^i : X_i \to N$, and $f_d^i : \prod X_{j_{j\in N}}^{i\neq j} \to N$. The game therefore remains a potential game with at least one pure-strategy Nash equilibrium. Note that the cost functions now contain many cost components, one for each flow (whose ingress gateway and egress locator are indicated by the strategy X_i).

For example, if three flows (toward as many destination locators) routed across the same egress edge link, the egress unitary routing cost in f_s^i for that edge link is triplicated. It is worth mentioning that, alternatively to the multiplication of the same link cost by the number of routed flows, in practice there is the opportunity to implement congestion control mechanisms; this can be done by adding congestion cost components to f_s and f_d as function of the used link bandwidth, if flow bandwidths are known by all the networks in the cluster.

7. Remarks on Implementation

The current Locator–Identifier Separation Protocol (LISP) specification [8] is an IETF proposition implementing a form of transit-edge hierarchical routing with routing locator metrics. At present, it is implemented in some new routers (e.g., in some Cisco routers) and is under testing in the http://www.lisp4.net testbed. In LISP, the translation from destination identifier to routing locators (RLOC) is performed by a distributed database system called mapping system. The mapping systems provides all the locators announced for an identifier or an identifier space, and can optionally set for each locator a priority cost (lower is preferred) and a load-balancing integer weight (from 0 to 100) announced by the corresponding network gateways. In its current form, the LISP weight is used only when there are equal LISP priorities. As already argued, the naive usage of such weights and priorities is not strategically justified when the communicating networks are independent and have significant equivalent traffic exchanges. Our framework can be thus seen as a Traffic Engineering LISP (LISP-TE) framework.

From a practical standpoint, we are interested in using integer percentage values out of b_x and b_y (or b_{x_i} for the n/networks case) ratios for backward compatibility with the LISP's integer weights. The LISP priority field might be used as a coordination channel, and might possible be extended to allow the coding of both backward locator cost and forward path costs. It is worth mentioning that the LISP priorities and weights are to be announced globally, while in the bilateral interaction case a private bilateral signaling is needed. For the bilateral case, another coordination channel may be managed independently of, but coupled with, the global LISP mapping system. For the case of a cluster of edge networks, the load-balancing solution obtained can either be similarly restricted to the routing among cluster members only, or can be applied to *any*

Table 4
A generic 2-player symmetric game

I\II	L	R
T	(c,c)	(a,d)
B	(d,a)	(b,b)

Table 5
Decomposition of a 2-player symmetric game

I\II	L	R
T	$(0,0)$	$(d-c,d-c)$
B	$(d-c,d-c)$	$(d-c+b-a,d-c+b-a)$
I\II	L	R
T	(c,c)	$(a-d+c,c)$
B	$(c,a-d+c)$	$(a-d+c,a-d+c)$

Table 6
Decomposition of a 2-player symmetric game

I\II	L	R
T	$(0,0)$	$(-1,-1)$
B	$(-1,-1)$	$(-2,-2)$
I\II	L	R
T	$(2,2)$	$(5,2)$
B	$(2,5)$	$(5,5)$

source edge network if the sum of all the traffic contributions from all other edge networks is negligible with respect to the intra-cluster volume. Therefore, this last setting would be directly implementable under the current LISP proposition.

From an operational standpoint, an appropriate execution policy can be:

- when the LISP priorities are different, extract from them the locator costs and the forward path costs;
- compute the coordinated load-balancing solution;
- set the LISP weights accordingly;
- set the LISP priorities equal to each other.

When a network needs to announce new cost settings to reflect changes in traffic characteristics, Internet paths and their performance, or topology properties, it simply resets accordingly the LISP priorities so that the upstream networks detect they are different and thus the change to the coordination setting; then, all the participating networks implicitly converge to a new coordinated load-balancing solution.

8. Conclusions

The Internet infrastructure has been rapidly evolving for the last few years. 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 (from a routing perspective) can jointly solve both the routing scalability and connection resiliency issues.

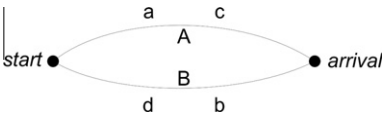

Fig. 10. Representation of a 2-player symmetric game.

In this paper, we study the novel traffic engineering capabilities arising in a transit-edge hierarchical routing context. We model the routing interaction between independent edge networks with non-cooperative game theory, and propose a strategically rational approach to coordinate the reciprocal routing of equivalent traffic volumes following routing equilibria, resulting in fine-selected edge-to-edge load-balancing. We mathematically demonstrate and experimentally show that our solution outperforms the current practice, and offers far more resilient solutions also with respect to the basic routing mode of the LISP protocol currently under standardization. Solutions brought by our approach show 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.

After describing how the model can be generalized to multiple network-player settings, we discuss that, from an implementation standpoint, our approach can be seen as a traffic engineering extension of the LISP architecture, presenting a LISP-based implementation policy. Our work represents an important step toward the definition of novel Internet traffic engineering methods for edge networks, where the content and the services (the "Clouds") are located.

Appendix A. Prisoner dilemma and potential games

We provide in this appendix a brief "tutorial" on how to decompose a prisoner's dilemma game as sum of two interesting types of games (extracted from [16]). Consider the generic symmetric game in Table 4, where $a,b,c,d \in \mathbb{R}$. We have a prisoner dilemma *cost* game if $a > b > c > d$, with (B,R) as Nash equilibrium, inefficient since both would prefer (T,L), which is however a dominated strategy profile. Indeed, this is the rationality dilemma offered by such games.

The game can be decomposed as sum of the two games shown in Table 5. For the first game, the cost components for the two players are equal for every profile. For the second game, the cost components of a player do not depend on its choice, but they depend on the other player's choice. The second game can be called "dummy game" since for a player there is no possible discrimination in choosing one strategy instead of the other. It can also be called "game of pure externality" meaning that its action has an effect *only* on the other player. This type of decomposition allows to clearly see the externality effect in the prisoner dilemma game.

It is easy to remark that the generic game in Table 6 is a potential game when $d-c=a-b$ and $c-a=d-b$. With

the setting: $a = 4, b = 3, c = 2, d = 1$ we obtain the game decomposition in Table 6. The choice of B allows to decrease the cost of I by 1, independently of the choice of II. At the same time, this choice increases by 3 the cost of II in the second game. It is worth noting that, in the first game, the costs are equal for the two players and that the choice of B has a positive externality effect for II: it decreases by 1 also its cost. Clearly inefficiency stems from the fact that externalities prevail upon selfish improvements.

With a broader perspective, one can note that such a decomposition is a general property of the so-called potential games [14]. For a game in strategic form $G = (X, Y; f, g)$, where X and Y are the strategy sets for the two players, and f and g are real functions, G admits a potential if it exists a function $P : X \times Y \to \mathbb{R}$ such that $\forall x', x'', x \in X, \forall y', y'', y \in Y$:

$$\begin{aligned} P(x', y) - P(x'', y) &= f(x', y) - f(x'', y) \\ P(x, y') - P(x, y'') &= g(x, y') - g(x, y'') \end{aligned} \tag{4}$$

P is called *potential function*. The analogy with physics relates, e.g., to the ability to substitute a "vector field" (the two payoff functions) with a single scalar valued function, or to the condition of being an irrotational field. Minima of the potential function are Nash equilibria for the game, which guarantees that finite potential games have equilibria in *pure* strategies [14].

Potential games emerge from congestion problems [17]. Indeed, we can represent the game of Table 4 with Fig. 10. Both players have to go from *start* to *arrival* taking either path A or path B (strategy A corresponds to T for I and to L for II, B corresponds to B for I and to R for II). The lower-case letters on each path in Fig. 10 indicate the transit cost for the players in case they walk alone (on the left) or together (on the right). If they travel together on the same path, the path is more congested than if they travelled alone along different paths, i.e., the cost is higher for both.

Appendix B. On mixed strategy equilibria

In a non-cooperative routing game, a strategically acceptable way to seek an *arbitrary load-balancing distribution* (e.g., 24%, 47% and 29% for three locators) might theoretically be reached implementing "mixed strategy" equilibria that could appear in addition to pure-strategy equilibria (the "type" discussed so far).

It is worth doing a small digression on this aspect. In game theory, with mixed strategies the player no longer chooses a single strategy, but a probability distribution on its (unilateral) available strategies. Somehow the player can rely on a random process that implements his decision following the probability distribution. In non-cooperative games, players adopt independent random processes, and the probability distribution of a strategy profile (e.g., an equilibrium) is given by discrete multiplication of the probabilities each player assigned to its corresponding strategy. Note that an equilibrium in pure strategies can be seen as a particular (degenerated) equilibrium in mixed strategies where each player strategy, hence the strategy profile, has a probability equal to 1. For example, in the game of Table 3, the equilibrium strategy G_2L_3 is played by network I with probability $p = 1$ and the other five strategies with probability $1 - p = 0$, and the same for network II and the equilibrium strategy G_3L_2 played with probability $q = 1$, so that the equilibrium profile (G_2L_3, G_3L_2) is played with probability $p \cdot q = 1$.

It has been proven that the mixed extension of a finite cardinal potential game, such as G, is also a cardinal potential game [14]. Therefore, we are interested in knowing if there can be additional mixed-strategy equilibria for G.

Corollary 8.1. *All the equilibria of the game G are pure-strategy equilibria, i.e., no additional equilibria are added with mixed strategies.*

In game theory parlance, this is quite straightforward once noted that the Nash equilibrium (a) of G can be found by iterated reduction of strongly dominated strategies. For example, in Table 3 the equilibrium can be obtained by first excluding, for network I, all G_1 strategies and G_2L_4 and G_2L_5 strategies since whatever network II chooses the network I cost is always minor, and by then conversely excluding G_4 and G_5 and G_3L_1 strategies for network II. The reduced game is the game degenerated to the single Nash equilibrium, if it is unique, and thus no mixed strategy is conceivable. If multiple equilibria exist for the general setting, the reduced game is composed of as much strategies and strategy profiles as needed to encompass the equilibria, and no additional mixed-strategy equilibria arise. Mixed strategies are therefore not implementable as a load-balancing distribution in a transit-edge routing game modeling.

Acknowledgment

The authors would like to thank Guruprasad K. Rao for his support in building the simulation datasets.

References

[1] S. Secci, K. Liu, G.K. Rao, B. Jabbari, Resilient traffic engineering in a transit-edge separated internet routing, in; Proceedings of 2011 IEEE International Conference on Communications (ICC 2011), Kyoto, Japan, 5–9 June 2011.
[2] D. Awduche et al., Overview and Principles of Internet Traffic Engineering, RFC 3272, 2002.
[3] D. Awduche, B. Jabbari, Internet traffic engineering using multi-protocol label switching (MPLS), Computer Networks (2002).
[4] B. Quoitin et al., Interdomain traffic engineering with BGP, IEEE Communications Magazine 41 (5) (2003) 122–128.
[5] R. Gao et al., Interdomain ingress traffic engineering through optimized AS-path prepending, in: Proceedings of Networking, 2005.
[6] D. Mayer, L. Zhang, K. Fall, Report from the IAB Workshop on Routing and Addressing, RFC 4984, September 2007.
[7] E. Elena, J.-L. Rougier, S. Secci, Characterisation of AS-level path deviations and multipath in internet routing, in: Proceedings of NGI, 2010.
[8] D. Farinacci, V. Fuller, D. Mayer, D. Lewis, Locator/ID separation protocol (LISP), RFC 6830, January 2013.
[9] D. Saucez et al., Interdomain traffic engineering in a locator/identifier separation context, in: Proceedings of INM, 2008.
[10] C. Labovitz et al., Internet inter-domain traffic, in: Proceedings of SIGCOMM, 2010.
[11] Y. Wang, J. Bi, J. Wu, Empirical analysis of core-edge separation by decomposing Internet topology graph, in: Proceedings of GLOBECOM, 2010.

[12] R. Douville, J.L. Le Roux, J.L. Rougier, S. Secci, A service plane over the PCE architecture for automatic multidomain connection-oriented services, IEEE Communications Magazine 46 (6) (2008).
[13] S. Das et al., Packet and circuit network convergence with OpenFlow, in: Proceedings of OFC, 2010.
[14] D. Monderer, L.S. Shapley, Potential games, Games and Economic Behavior 14 (1) (1996) 124–143.
[15] R.B. Myerson, Game Theory: Analysis of Conflict, Harvard Univ. Press.
[16] F. Patrone, Giochi con potenziale (Potential Games), Lettera Matematica (Pristem) 69 (2009) 17–19.
[17] R.W. Rosenthal, A class of games possessing pure-strategy Nash equilibria', International Journal of Game Theory 2 (1973) 65–67.
[18] Routeviews website <http://www.routeviews.org>.
[19] Details, datasets and codes website <http://cnl.gmu.edu/TAVRI>.

Stefano Secci received the M.Sc. degree in communications engineering from Politecnico di Milano, Milan, Italy, in 2005, and Ph.D. degrees in computer science and networks from Politecnico di Milano and Telecom ParisTech, Paris, France, in 2009. He is an Associate Professor at the LIP6, Université Pierre et Marie Curie (UPMC-Paris VI), Paris, France, since Oct. 2010. Before, he worked as Post–Doctoral Fellow at NTNU, Norway, and George Mason University, USA. Before the Ph.D., he worked as a Network Engineer with Fastweb Italia, Milan, Italy, and as a Research Associate with Ecole Polytechnique de Montréal, Canada, and with Politecnico di Milano. He actively participated in the ANR ACTRICE project,the FP7 EuroNF and ETICS projects, the CELTIC TIGER2 project, and the ONR TAVRI project. His works space from optical networking to IP routing optimization and traffic engineering. His current research interests are about future Internet network resiliency, mobility, and policy. Dr. Secci is the recipient of the NGI 2009 Best Paper Award. He is author of several papers published in leading conference proceedings and journals. He has been member of many Technical Program Committees including IEEE ICC, WCNC, GLOBECOM, and NGI, TPC co-chair of IEEE CloudNet 2012 and NoF 2011, referee for the Italian Ministry of Research and University, the Romanian Council for Scientific Research, and associate editor for IEEE Communications Surveys & Tutorials and Journal of Network and Systems Management. He is Secretary of the Internet Technical Committee (ITC), joint TC between the IEEE Communication Society and the Internet Society (ISOC), since 2011.

Kunpeng Liu received his Bach. degree in electrical engineering from Tsinghua University, Beijing, China, in 2005, and his M.Sc. degree from George Mason University in 2009, where he is now Ph.D. student at the Communications and Networking Lab. of GMU. His research interests are about future Internet routing and switching architectures.

Bijan Jabbari is a professor of electrical and computer engineering at George Mason University, Fairfax, Virginia, USA, and an affiliated faculty member with Telecom ParisTech (ENST), Paris, France. Dr. Jabbari served as the Editor for Wireless Multiple Access for the IEEE Transactions on Communications, was the International Division Editor for Wireless Communications of the Journal of Communications and Networks, and was on the editorial board of Proceedings of the IEEE. He is the past chairman of the IEEE Communications Society technical committee on Communications Switching and Routing. He is a Fellow of IEEE and IEE. He is a recipient of the IEEE Millennium Medal in 2000 and the Washington DC Metropolitan Area Engineer of the Year Award, in 2003. He helped industry adoption of MPLS by service providers and large corporations in their networks. He continues research on multi-access communications and high performance networking. He received his Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.