Implementing a BGP-Free ISP Core with LISP

Florin Coras*, Damien Saucez†, Loránd Jakab*, Albert Cabellos-Aparicio*, and Jordi Domingo-Pascual*
*Universitat Politècnica de Catalunya, Barcelona, Spain
†INRIA, Sophia Antipolis, France

***Abstract*—The sustained growth pace of the global routing table is exerting an economical strain on ISPs by requiring untimely router upgrades. Notably, it has been speculated that the growth rate of router FIBs is surpassing that of its supporting technology and that the deployment of IPv6 is only to make matters worse. In this paper, we propose LISP-MPS, an architecture based on LISP, that isolate the intra-domain routing of an Autonomous System (AS) from its inter-domain routing. The resulting separation implies the decrease of backbone routing table sizes and enables an AS to control the forwarding of traffic inside its network. For a seamless, cost effective, and incremental deployment, LISP-MPS leverages iBGP to implement the LISP mapping system functionality with minimal modification to a small subset of deployed equipment. Finally, an analysis of realistic topologies shows that, despite changing how packets transit a network, the architecture does not lose resilience to router failures. Moreover, we show that it can be a viable alternative to BGP/MPLS deployments due to its low implementation cost.**

I. Introduction

Recent developments have shown that the Internet's routing table is growing at a large pace [1]. Reasons for this growth are partly organic in nature, as the number of domains connected is constantly increasing, but also related to practices like multihoming and traffic engineering. The two generally defeat provider based address aggregation and are worsening the situation through support for prefix deaggregation. Additionally, the exhaustion of the IPv4 address space is now fostering the deployment of IPv6. Should the adoption move too slowly, the IPv4 address space might end up fragmented and this would lead to the end of hierarchical address aggregation. Should the converse hold, for a given time period the routers would have to store routing tables for both address families. Unfortunately, both paths seem to foretell an accelerated growth of the Internet's Default Free Zone (DFZ) routing table.

As exposed in [2] and [3], the growth of the DFZ routing table has detrimental effects on the operational costs of Internet Service Providers (ISPs) in the current operating environment. Driving costs up is the increased technical complexity required for the management of large tables. For instance, if scaling a router's Routing Information Base (RIB) is assured by the commonly accepted *Moore's Law*, the same can not be said about the Forwarding Information Base (FIB) table. The former is generally saved in cheap, mass produced, control plane memory whereas the latter is stored in much faster but also more expensive and difficult to scale line card memory. More technological limitations are discussed in [2]. Overall, they translate to increased router prices and, in the long run, due to accelerated table growth, to shorter router life spans.

The impossibility of the current Internet's architecture to scale with the routing table size was deemed in an Internet Advisory Board workshop as one of its most important problems [2]. After a detailed analysis, the participants have identified the overloading of IP address semantics with both *location* and *identity*, as the main source of this limitation. Map-and-encap [4] was suggested as a starting point mechanism for defining a solution however, since the workshop's conclusion, a plethora [5] of architectures have been proposed. Among them, Locator/Identifier Separation Protocol (LISP) [6] proposes a semantic decoupling of *identity* and *location* at network level. The two resulting namespaces are used to unambiguously address end-hosts and their Internet attachment points. Their binding is done by a mapping system and LISP routers use encapsulation to transparently exchange (i.e., tunnel) content over the Internet's core. LISP is an incrementally deployable solution that, besides counting with support from both academia and industry [7], also relies on a pilot network [8] dedicated to its ongoing development. Nevertheless, no significant DFZ routing table size reduction is expected until a considerable part of the Internet's edge ASes upgrade to LISP [9].

In this paper we propose to complement the traditional LISP's inter-domain use with a new deployment case restricted to the scope of an AS. Similar to the inter-domain location-identity dichotomy, in intra-domain context there's a distinction to be made between an IGP-external destination prefix and the location of the points whereby it could be reached. For instance, all IP routers in an AS's backbone are required to carry BGP routes although no BGP decision is taken within the domain. This needlessly exposes routers to external routes when information about egress points would suffice.

The goal of our work is to devise a mechanism that reduces the size of the routing tables in IGP backbone routers and enables advanced intra-domain traffic engineering. To this end, we propose the use of LISP's tunnelling ability to obtain a BGP-free core but also as a mechanism to control the points through which packets egress a domain. In our solution, border routers select local egress points for transiting packets and send them to the border by the mean of encapsulation.

Thinking in a swift deployment, we propose to reuse existing iBGP infrastructure as a mapping system and require just a mild upgrade to enable LISP functionality. The resulting mapping-system *pushes* bindings to tunnelling routers and therefore ensures no mapping misses and update propagation times no worse than those in current networks. Additionally, for traffic engineering and resilience purposes, a router and

router interface addressing scheme is proposed.

The architecture we propose in this paper bears similarities with networks that jointly deploy MPLS and BGP [10], [11]. However, following the lead of [12] we advance our architecture as an IP-routing based alternative. In [12] Metz et al. express concerns that MPLS might possess a control-plane complexity factor and argue that IP mechanisms might be equally suited at performing MPLS functions. Furthermore, MPLS has a constrained footprint, and can't be natively forwarded between disjoint networks, whereas IP is ubiquitous and easily supports coordination of disjoint sub-domains. In homage to MPLS and because of the vague similarity between MPLS and our architecture, we named our proposal *LISP Multi Protocol Switching*, or LISP-MPS.

The remainder of this paper is organized as follow. First, we provide necessary background on LISP and current routing practices in Internet Service Providers (ISP) in Sec. II. Then, we present the details of our LISP-MPS architecture that relies on LISP encapsulation and an iBGP control plane in Sec. III. Further, we discuss the added value of LISP-MPS in Sec. IV and evaluate its benefits in Sec. V. We finally contrast LISP-MPS to the related work in Sec. VI and conclude this paper in Sec. VII.

II. Background

In this section, we present the necessary background for the understanding of our solution with a brief description of LISP in Sec. II-A and current practices in ISP networks in Sec. II-B.

All along the paper, we use the following taxonomy that splits a domain's routers in three categories: i) AS Border Routers (ASBRs), routers found at the border with other ASes, ii) Customer Border Routers (CBR), routers that connect local customer networks to the backbone and iii) Backbone Routers (BBR), all AS core routers not ASBRs or CBRs. We may refer to the first two simply as edge or Border Routers (BR).

A. LISP

LISP [6] is one [5] of the recently emerged architectural solutions to the Internet's scalability problem [2]. Its main goal is that of splitting the semantics of IP addresses with the aim of forming two namespaces that unambiguously identify core (locators) and edge (identifiers) network objects. To facilitate transition from the current Internet infrastructure, both of the resulting namespaces use the existing IP addressing scheme. As a result, the split does not affect routing within existing stub or transit networks. Nevertheless, as identifiers and locators bear relevance only within their respective namespaces, a form of conversion, from one to the other, has to be performed at border points between core and edge networks. LISP employs map-and-encap [4] as a means of storing both within a datagram. However, besides the need for data plane modifications (i.e., encapsulation), LISP also requires the introduction of a new control-plane *mapping function* able to provide bindings that link identifiers to locators.

Therefore, prior to forwarding a host generated packet (see Fig. 1), a LISP router maps Endpoint IDentifier (EID) (i.e., destination address), to a corresponding destination Routing LOCator (RLOC) by means of *mappings* obtained via a LISP specific distributed database, called the *mapping system* [13], [14], [15]. A mapping associates one EID prefix to a list of RLOCs. Each RLOC is assigned a *priority* and a *weight*. Among the list of RLOCs, the locator with the lowest priority value must be selected. If serveral of such locators are possible, the traffic load is balanced between them proporionaly to their weight. Once a mapping is obtained, the border router tunnels the packet from source edge to corresponding destination edge network by means of an encapsulation with a LISP-UDP-IP header. The outer IP header addresses are the RLOCs pertaining to the corresponding border routers. At the receiving router, the packet is decapsulated and forwarded to its intended destination. In LISP parlance, the source router, that performs the encapsulation, is called an Ingress Tunnel Router (ITR) whereas the one performing the decapsulation is named the Egress Tunnel Router (ETR). One that performs both functions is referred to as an xTR.

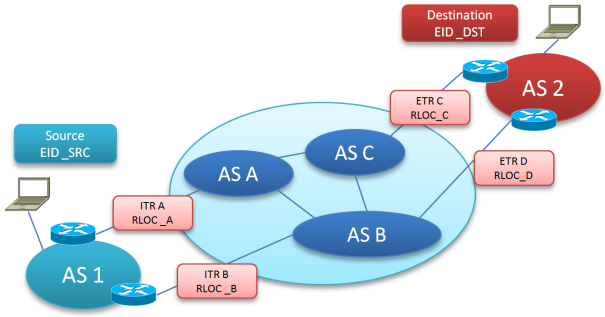

Fig. 1. Example LISP Architecture

Regarding encapsulation header size, LISP encapsulation is more inefficient than, for instance, MPLS encapsulation. However, the Maximum Transmission Unit (MTU) size in backbone networks is typically much larger than in access networks, and thus packet fragmentation is avoided.

B. ISP Routing

In what follows we shortly review some of the mechanisms related to intra-domain routing and expose several of their limitations. The presentation is based on the assumption that the intra-domain and inter-domain packet routing for an AS are assured by an Interior Gateway Protocol (IGP) and the Border Gateway Protocol (BGP), respectively.

If not otherwise stated, we consider BBRs as BGP enabled. Further, we expect that iBGP is used for the intra-domain advertisement of BGP reachability information between BRs and to BBRs. Also, we suppose that Route Reflectors (RR) are are used for scaling the iBGP route redistribution.

One of the main drawbacks of such deployments is the need for BGP enabled BBRs. Normally, prior to forwarding a datagram, a router needs to determine a next-hop for the packet's destination address and subsequently an interface out on which this next-hop may be reached. Thus, all routers within a domain must be able to determine a next-hop for

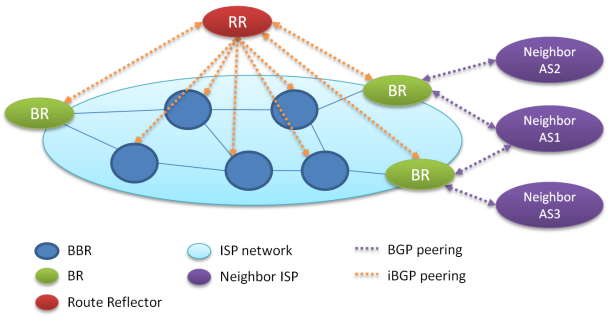

Fig. 2. ISP network example

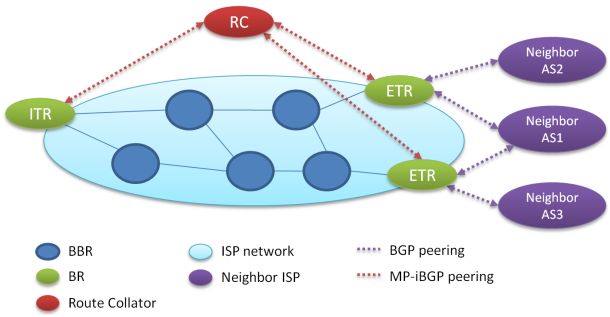

Fig. 3. Proposed LISP-MPS Architecture

any globally or intra-domain routable destination address a packet may hold. Typically, this results in the routers, besides participating in the IGP, being provided DFZ reachability information by means of iBGP (see Fig. 2). Consequently, they all need to store two different scope routing tables and deal with their associated protocol instabilities.

Additionally, due to iBGP's design, the two routing tables are coupled in the resolution chain of an outgoing interface for non-IGP destinations. In such a scenario, the next-hop of the iBGP learned route will typically be an address not adjacent to the resolver. Instead, it could either pertain to the router advertising the route in iBGP (a local BR) or to the foreign BGP peer from which the local BR learned the route. Therefore, a second resolution is needed, of the next-hop against the IGP table, for the discovery of an interface out on which the packet can be forwarded to the next-hop. Such resolution process can be intuitively interpreted as a double mapping. First, an address is mapped to a gateway, the BGP route's next-hop, which at its turn is mapped to an IGP route learned over a local interface. From the perspective of on path backbone routers the procedure is obviously redundant as they all perform identically the first mapping, if iBGP is converged.

To avoid storing BGP routing tables in BBRs, ISPs may use MPLS for tunnelling traffic between BRs. Additionally, this results in several traffic engineering benefits. First, the ability to speed up the the forwarding of traffic over a domain's backbone, optionally under QoS constraints. Second, due to MPLS's fast reroute capabilities good resilience to failures. Finally, in combination with Multipath BGP, MPLS tunnelling could be used for load balancing traffic between multiple egress points (BRs), instead of just one. However, as explained in [12], MPLS is quite complex to manage and requires support in all backbone routers. Furthermore, its deployment is typically limited to a domain so disjoint networks are hard to interconnect.

III. Proposed Architecture

The driving goals of our proposal, LISP-MPS, are to i) devise a solution for ISPs wishing to diminish the size of the routing tables in the routers part of their backbone networks and ii) enable more complex intra-domain traffic engineering policies. This section presents how these could be achieved with LISP. However, the proposed architecture boasts a much larger feature set which we will expand on in Section IV.

A. Overview

As explained in Section II-B, within an autonomous system network, backbone routers must store BGP routes, although itself this information can not influence the intra-domain routing of transiting packets. Furthermore, configuration of these intra-domain paths is not possible with a simple BGP-enabled core. As a result, network operators seeking a BGP-free core and intra-domain traffic engineering capabilities employ MPLS tunnelling over the network's backbone.

Following the lead of Metz [12] we propose the use of LISP as a more flexible alternative to MPLS. Thus, with LISP-MPS, for a packet transiting a domain, the egress BR is chosen at the ingress BR and stored in the datagram by means of LISP encapsulation. All further intra-domain routing of the packet will be done based on IGP information. This obviates the need for iBGP route redistribution to BBRs and therefore limits the scope of the DFZ routing information to the points of interaction with neighbouring domains, the ASBRs, and local customers, CBRs. Encapsulating BRs learn the mappings between external prefixes and the addresses of the BRs announcing their reachability by means of a mapping system (cf. Fig. 3). From traffic-engineering perspective, apart from the ability to precisely choose the traffic egress points at any ingress BR, LISP-MPS allows an operator flexibility in updating its running configuration in a timely fashion.

Henceforth, given a prefix, we shall refer to the IGP addresses of its iBGP originators, as Prefix Attachment Points (PAPs). By virtue of the previous definition, a PAP may be a synonym of the router itself or one of its interfaces. We further refer to the former as Router Name (RN) and to the latter as Router Interface Name (RIN). An addressing scheme for the two is suggested in Section III-D.

We detail in what follows the functioning of LISP-MPS's control and data plane.

B. Control Plane

Not to require the introduction of new network equipment, we exploit the iBGP implementation in edge routers and Route Reflectors (RRs) [16] for the distribution of mapping information. However, we do require an upgrade of the RRs,

or their pairing with an additional device, in order to support LISP functionality. To avoid confusion, we call the new route reflecting network element a Route Collator (RC).

So, similarly to an RR, an RC (see Fig. 3) is fed by BRs all their external BGP learned routes. As added constraints, all routes must have as next-hop attribute the RN of their advertising BR and must carry information about all the PAPs of the BR. This is achieved with the help of MP-BGP [17]. On the resulting RIB the RC runs the BGP decision process and selects for each external prefix the best route. If multiple border router advertisements tie in the selection process run by the RC, instead of breaking the tie by means of IGP metric, all the PAPs of the tied BRs should be saved as viable attachment points for the considered prefix. The resulting egress point diversity enables a fine grained tuning of the traffic engineering policies for a domain. To mark the preference of using a PAP out of a possible candidate set a *priority* is associated to each. Should more attachment points be equally preferred, load balancing is to be employed among them according to another associated value, the *weight*.

Once the decision process is finished the RC pushes through iBGP the selected routes to all BRs but without priority or weight information. If multiple BRs announcing disjoint AS paths were considered as valid egress points for a prefix, then the RC must generate a new BGP route containing an encompassing aggregate of the paths. This ensures that all BRs are provided with valid BGP paths they can forward to external BGP peers and avoids generating external BGP loops.

Apart from the iBGP updates, an RC must generate and push LISP messages to BRs such that they can build prefix-to-PAP map-caches. For prefixes with multiple PAPs the messages also convey priority and weight information. Considering that the LISP upgrade is the only disruptive change when moving from RR to RC functionality, a more cost-effective upgrade to LISP-MPS would be to pair a LISP capable device with an RR. Therefore, iBGP responsibilities would be fulfilled by the RR whereas LISP related ones by the new server with the help of iBGP feeds shared by the RR.

Note that BGP syntax could be enhanced to carry all LISP required information. However, we avoided this solution not to correlate iBGP and LISP updates and to avoid triggering the BGP decision process on LISP updates. Still, this alternative might be worth more consideration in the future.

C. Forwarding Changes

The simplification of the forwarding in BB routers is counter-balanced by a slight complication of the data-plane operations in border routers. On receiving a packet, a BR performs a longest prefix match of the destination address in the LISP map-cache. Besides the prefix encompassing the destination address, the router learns the PAP(s) of the BR(s) announcing reachability of the matched prefix and their associated priorities and weights. Having these, the BR selects one of the attachment points and then proceeds to LISP encapsulating the datagram. The resulting datagram is routed across the backbone network solely by the IGP. Once the packet reaches the destination edge router, it gets decapsulated and forwarded natively to the neighbouring AS.

D. Border Router Addressing

Aiming to improve intra-domain traffic engineering, we seek to provide the means to an RC to establish ITR-to-ETR paths distinct from those computed by IGP. As a result, we propose an intra-domain router and router interface naming scheme that makes use of IP prefix aggregation properties for enhanced router addressing. Each border router is allocated a local domain prefix whose reachability it must announce out all its interfaces. By convention, we consider the first address in the prefix to be the RN and attribute it to the router's loopback interface. The rest of the prefix is split in smaller blocks, each advertised out on and used to address one of the router's interfaces. Overall, a border router announces reachability for $N + 1$ prefixes, where N is the number of its IGP facing interfaces. The fact that an interface can be *selected* out of those pertaining to a router and the way the router addressing is performed are beneficial for traffic engineering and failure recovery. Both are discussed in more depth in Section IV.

The number of additional entries to add in the FIB of each BBR is then given by:

$$\Omega = \sum_{r \in \mathbf{B}} |\mathbf{I}_r| + 1, \tag{1}$$

where $\mathbf{B}$ is the set of border routers and $\mathbf{I}_r$ is the set of IGP facing interfaces of a router r. Similarly, the number of entries necessary to add at a BR, for any $r \in \mathbf{B}$, is given by:

$$\Omega - (|\mathbf{I}_r| + 1)\,. \tag{2}$$

Note that Ω is independent of the global routing table (i.e., BGP) and only depends on the network topology (i.e., number of BRs and that of their IGP facing interfaces).

IV. Discussion

This section presents an analysis of the benefits and drawbacks of LISP-MPS. A comparison of the routing protocols ran by routers in domains with BGP, LISP-MPS and BGP/MPLS enabled backbones is shown in Table I.

1) Routing Table Reduction: One of the most important benefits of LISP-MPS is that it reduces the size of the routing tables on the backbone routers of an ISP. It does so by isolating the intra-domain routing from the inter-domain routing and by pushing all the inter-domain reachability state to the edges of the AS's topology. The result is that in our solution the BBR table sizes are bounded by the IGP size. Comparatively, in a BGP enabled backbone they grow proportionally to the number of prefixes in the DFZ. With BGP/MPLS they are also limited to the size of the IGP routing table but BBRs need to store an additional table for label switching.

2) Virtual Networks: LISP supports network virtualization with the help of an address-space extending field called *Instance-ID* (IID). Then, given multiple organizations with disconnected site-networks that use overlapping private address space, per organization site interconnection requires that per organization IIDs are used to tag each site. Obviously, all sites pertaining to an organization have the same IID and their extended address space is unique. We call such multi-site networks, where one organization controls all sites but not the network interconnecting them, *virtual networks*. To distinguish between all clients, routers at the transit-client border must install per virtual network forwarding tables and in-transit packets must carry the IID.

With LISP-MPS, at transit ingress ASBR, packets are matched (e.g., based on interface, VLAN tag) to a virtual network and thus to an established IID. The packets are subsequently encapsulated based on the map-cache associated to the virtual network and then forwarded. Finally, at the egress ASBR packets are decapsulated and forwarded to the client network that matches the conveyed IID. BRs continue to populate their map-caches by means of iBGP. Additionally, all ETRs pertaining to the transit provider tag the virtual networks route advertisements (i.e., mappings) with a *RouteTarget* [11] equal to the IID. As a result, the ITRs may build per *RouteTarget* map-caches.

In particular, this feature could be used by a service provider to offer virtual private network (VPN) services to its clients. An advantage over MPLS based VPNs is that this solution does not require the use of double encapsulation.

3) Multi Protocol Switching: LISP presently supports the encapsulation of a large set of protocols (e.g., IPv4, IPv6, or Ethernet Frames). Furthermore, by means of [18], can be extended to support virtually any protocol. As a result, LISP-MPS can be used to setup layer-2 VPNs or IPv6 networks independent of the underlying IGP routing protocol. Regarding IPv6 transit, besides requiring no backbone network upgrade, the solution avoids running two separate forwarding tables and thus worsening the FIB growth.

4) Flexible Routing Control: Access to IGP information should allow the collator to compute for all destination prefixes all the possible intra-domain paths. Depending on the network's complexity, an efficient distribution of traffic that minimizes metrics like link stress, bandwidth usage or latency could be implemented by configuration or with the help of an heuristic. The results may be imposed with the help of PAP priorities and weights. In this sense, traffic may be distributed among multiple PAPs with the same priority and for a specific PAP, traffic should ingress according to the weights associated to its interfaces.

5) Resilience to IGP Link and Router Failures: In the event of a BB router, or one of its interfaces failing, the IGP should generally deal with the re-routing of in-flight packets around the affected patch of network. Nevertheless, there are two IGP failure scenarios where a more complex network reaction is needed. First, if the disruption disconnects a border router's interface from the backbone network, its associated IGP route (as described in Section III-B) disappears. Consequently, all packets destined to the affected interface will match the associated aggregate prefix and will be delivered to the border router on one of its still active interfaces. The failure affects just intra-domain traffic engineering policies, if any were in place, but results in no packet loss. The second failure scenario is the result of a complete isolation from the backbone network or halting of an edge router. To avoid packet black-holing we propose the use of *re-encapsulators*. These devices attract with routes covering the whole PAP address space all packets whose egress points have failed. They then re-encapsulate this traffic towards alternative border routers. If no such router exists, the packets are dropped. ASBRs may act as re-encapsulators.

To be noted that both types of failures are detected by LISP through probing after a time threshold and subsequently the PAP used in the encapsulation is changed with a valid one.

6) Resilience to eBGP Adjaceny Failure: In this case, reachability of the prefixes advertised only through the affected adjacency will be, independent of LISP-MPS, lost. Still, the prefixes with multiple potential egress points will have their best path recomputed once the failure is advertised. Therefore, once the new routes are distributed to the BRs, the transit paths of affected prefixes switch to valid egress points. However, all in-flight packets are dropped if they reach the affected border router before it updates its forwarding table with new egress points for destinations it lost connectivity to. Alternatively, re-encapsulators could be used to avoid all packet loss for prefixes with multiple egress routers.

7) Deployment: Because of a limited number of upgrades, the proposal presents a low overall deployment cost. The architecture's data plane requires just the upgrading of a domain's BRs. Furthermore, the mapping system reuses the iBGP protocol and only requires the upgrading of RRs to LISP functionality. Alternatively, RRs could be coupled with devices that perform LISP mapping-system specific functions.

If the scalability of the RC is a concern due to associated operational complexity, solutions like [19] could be implemented for distributing the collator.

TABLE I
COMPARISON OF SOLUTIONS (DIFFERENCES IN GRAY)

Router	BGP Backbone	LISP-MPS	BGP/MPLS
ASBR	IGP + eBGP + iBGP		
		LISP	MPLS
CBR	IGP + iBGP		
		LISP	MPLS
BBR	IGP		
	iBGP		MPLS
RR(RC)	iBGP	MP-iBGP	MP-iBGP

V. EVALUATION

LISP-MPS offers operators flexibility in controlling their transit traffic over the different egress points. In this section,

we evaluate two aspects of LISP-MPS. On the one hand, we estimate the gain in term of path diversity that an operator can expect if it deploys LISP-MPS. On the other hand, we determine the cost of using the technology to leverage the diversity by estimating the overhead in the routing table caused by the injection of interface related prefixes.

A. Path diversity incentives

BGP is such that only one route can be used to reach a destination. However, it is frequent that an AS receives several routes for each prefix, and this diversity is lost because of the BGP decision process. To quantify the potential path diversity that an operator can use by using LISP-MPS, we studied the diversity of BGP routes. For that purpose, we analyzed the BGP feeds of the four routers belonging to the University of Oregon available at Routeviews [20]. For each router, we took the Routing Information Based snapshot at midnight on March 15^{th}, 2012. Fig. 4 shows distribution of route diversity for three different filtering rules. More precisely, the figure shows the cumulative distribution of the number of prefixes (among the 424,833 prefixes) grouped by the number of routes that remain to reach them after being filtered. The curve label `no filter` gives the number of routes received for each prefix. As we can see, 95.5% of the prefixes have at least 2 routes. In other words, in general prefixes have path diversity. However, some routes should not be used because they are too long and would impact the performance. The curve labeled `shortest AS path` takes the length of the path into account and filters the RIB to only keep the routes that minimize the path length. In this situation, the proportion of prefixes with at least two routes is still 5.6% which means that the traffic for more than half of the prefixes could be load balanced between paths of same length. Finally, the curve labeled `same AS path` determines all the routes that have the same AS path as the route that would have been chosen by BGP's decision process. We observe that we have still 50.6% of the prefixes with at least two routes. In this particular case, the routes can be used in parallel, without disrupting BGP, as the AS path is preserved.

As a summary, an operator can see benefits in using LISP-MPS as it enables the use of several routes in parallel. This increases its traffic engineering capabilities and potentially reduces the traffic's transit cost [21].

B. Routing overhead

In the previous section, we saw that operators could gain in terms of diversity when using LISP-MPS. In this section we put this gain in perspective by estimating the routing overhead caused by the router addressing scheme required for enhanced traffic control. The proposed scheme consist in advertising all IGP facing interface of border routers in the IGP as well as an aggregate to protect against failure.

We have estimated Ω, see eq. (1), for 7 different topologies. Among them, there is the topology provided by Internet2 [22], the topology of Géant [23] and the last 5 are taken from Rocketfuel [24]. For Géant and Internet2, all the details

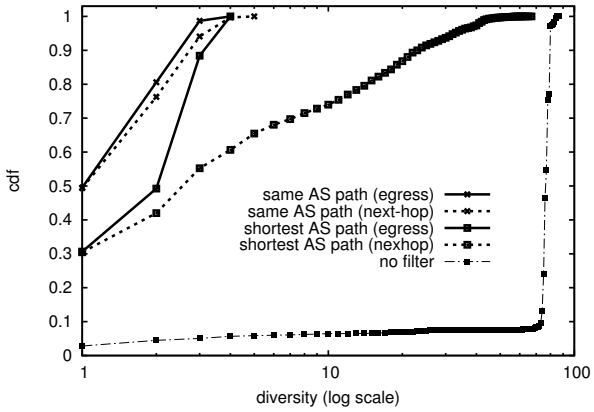

Fig. 4. Distribution of the BGP path diversity (i.e., the number of routes for a prefix) under different filtering rules

are provided so we can determine exactly the BRs and the BBRs. Alas, this is not possible for the Rocketfuel topologies. Therefore, we assigned the role of BR to two routers in each city. For this, we assumed that every city is a Point-of-Presence (PoP) and that a PoP must be protected against the failure of one router, hence two BR per city. For the considered topologies, we observe the value of Ω to be 8, 21, 128, 151, 166, 200, 294, and 513. In addition, we found that the number of IGP facing interfaces at the BRs is 4 ± 0.48. What is interesting in these results is that even in the case of large networks, the number of additional routing entries remains small in comparison to those necessary to operate BGP. The network with the largest Ω (of 513) is the one reported by Rocketfuel for Sprint. It has no less than 1944 links, 315 routers, and for which we accounted no less than 83 BRs.

VI. Related Work

The section reviews the ideas of some similarly aimed works carried in the field.

FIB Aggregation is an opportunistic technique that offers per router FIB size reductions by algorithmically removing specific forwarding (child) entries which share the same next hop with their trie ancestors. The procedure ensures forwarding correctness however, depending on the employed algorithms, it may introduce previously non-routable address space in the FIB. There are several proposals [25], [26], [5], [27] that recommend the use of these techniques for reducing routing table sizes. Notably, [27] presents a systematic analysis of costs and benefits for FIB aggregation and it concludes that it is a viable short-term solution.

Several works propose, like us, the use of *tunnelling* for relieving the pressure exerted by the size of forwarding tables on routers. Virtual Aggregation [28] tries to diminish the routing tables of routers within an AS by having the legacy routers forward their traffic to several *aggregation point routers* (APRs) instead of the best egress points. The forwarding on this second section (from the APR to the ASBR) is done by using MPLS tunnels in order to avoid routing loops. As a result, the number of FIB entries in legacy routers is

limited to the number of APRs. A downsides of this solution is that it introduces additional path-strech within the AS. Many Loc/ID split proposals [5] make use of encapsulation to decouple core from edge routing. Depending on how their deployment is to be done, they could reduce the size of the DFZ routing table.

These solutions manage to decrease the intra-domain routing tables, either through aggregation or by exclusion of edge-networks (EID) address space. Even so, there is still a direct relation between the size of the RLOC space and the size of the routing tables in a domain's backbone network. Our solution however, isolates intra-domain from inter-domain routing and directly relates the backbone routing table size to the number of BRs.

RCF 3107 [10] suggests the distribution of BGP routes with MPLS label mappings piggybacked onto them. Should border routers be using this mechanism together with a intra-domain label distribution protocol, then there is no need for BB routers to run iBGP if they support MPLS. At the edge of the domain a packet would get encapsulated with the label mapped on its matched route and subsequently MPLS forwarded over the backbone to its intra-domain next hop. We make use of BGP label mapped routes in our proposal however, instead of using MPLS, we use LISP encapsulation. This saves the need to support MPLS in the network's backbone and the deployment of a label distribution protocol.

VII. Conclusions

In this paper, we have devised and analyzed LISP-MPS, a LISP based solution that increases the lifespan of ISP backbone network routers by reducing the size of their routing tables. Once identified the source of growth as the inefficient intra-domain DFZ route redistribution, we have offered LISP encapsulation as a simple and efficient solution. Details regarding a domain constrained architecture were presented and an incremental deployment of a mapping-system that reuses existing iBGP infrastructure was proposed with the possibility to implement high-level routing policies thanks to a centralized entity named Route Collator. Finally, we show with BGP traces obtained from RouterViews and topologies from Rocketfuel that the traffic engineering opportunities of an AS are drastically increased whilst using LISP-MPS. Furthermore, we show that the offered feature set reduces the operational costs but maintains strong resiliency capabilities. More work is needed to understand how to implement the Route Collator high level routing policies in a distributed way.

Acknowledgements

We want to express our gratitude to Noel Chiappa for providing both inspiration and valuable feedback and to Pierre François for his insightful comments. This work has been partially supported by the Spanish Ministry of Education under scholarship number AP2009-3790, Catalan Government under project 2009SGR-1140, research project TEC2011-29700-C02 and a Cisco grant.

References

[1] G. Huston, "BGP Report Sep 2011." [Online]. Available: http://bgp.potaroo.net/
[2] D. Meyer, L. Zhang, and K. Fall, "Report from the IAB Workshop on Routing and Addressing," RFC 4984 (Informational), Internet Engineering Task Force, Sep. 2007.
[3] T. Narten, "On the Scalability of Internet Routing," draft-narten-radir-problem-statement-05, Internet Engineering Task Force, Feb. 2010, work in progress.
[4] R. Hinden, "New Scheme for Internet Routing and Addressing (ENCAPS) for IPNG," RFC 1955 (Informational), Internet Engineering Task Force, Jun. 1996.
[5] T. Li, "Recommendation for a Routing Architecture," RFC 6115 (Informational), Internet Engineering Task Force, Feb. 2011.
[6] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "Locator/ID Separation Protocol (LISP)," draft-ietf-lisp-15, Internet Engineering Task Force, Jul. 2011, work in progress.
[7] "IETF Locator/ID Separation Protocol WG." [Online]. Available: https://datatracker.ietf.org/wg/lisp/charter/
[8] "LISP Testbed." [Online]. Available: http://www.lisp4.net/
[9] L. Jakab, A. Cabellos-Aparicio, F. Coras, J. Domingo-Pascual, and D. Lewis, "LISP Network Element Deployment Considerations," draft-ietf-lisp-deployment-01.txt, Internet Engineering Task Force, Jul. 2011, work in progress.
[10] Y. Rekhter and E. Rosen, "Carrying Label Information in BGP-4," RFC 3107 (Proposed Standard), Internet Engineering Task Force, May 2001.
[11] E. Rosen and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)," RFC 4364 (Proposed Standard), Internet Engineering Task Force, Feb. 2006, updated by RFCs 4577, 4684, 5462.
[12] C. Metz, C. Barth, and C. Filsfils, "Beyond MPLS Less Is More," *Internet Computing, IEEE*, vol. 11, no. 5, pp. 72–76, 2007.
[13] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "LISP Alternative Topology (LISP+ALT)," draft-ietf-lisp-alt-01, Internet Engineering Task Force, May 2009, work in progress.
[14] V. Fuller, D. Farinacci, and D. Lewis, "LISP Delegated Database Tree (LISP-DDT)," draft-ietf-lisp-ddt-00, Internet Engineering Task Force, Nov. 2011, work in progress.
[15] L. Jakab, A. Cabellos-Aparicio, F. Coras, D. Saucez, and O. Bonaventure, "LISP-TREE: A DNS Hierarchy to Support the LISP Mapping System," *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 8, pp. 1332 –1343, october 2010.
[16] T. Bates, E. Chen, and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)," RFC 4456 (Draft Standard), Internet Engineering Task Force, Apr. 2006.
[17] T. Bates, R. Chandra, D. Katz, and Y. Rekhter, "Multiprotocol Extensions for BGP-4," RFC 4760 (Draft Standard), Internet Engineering Task Force, Jan. 2007.
[18] D. Farinacci, D. Meyer, and J. Snijders, "LISP Canonical Address Format (LCAF)," draft-farinacci-lisp-lcaf-07, Internet Engineering Task Force, Mar. 2012, work in progress.
[19] I. Oprescu, M. Meulle, S. Uhlig, C. Pelsser, O. Maennel, and P. Owezarski, "oBGP: an Overlay for a Scalable iBGP Control Plane," *NETWORKING 2011*, pp. 420–431, 2011.
[20] University of Oregon, "Routeviews project." [Online]. Available: http://www.routeviews.org
[21] A. Dhamdhere and C. Dovrolis, "ISP and Egress Path Selection for Multihomed Networks," in *INFOCOM*, april 2006, pp. 1 –12.
[22] "Internet2." [Online]. Available: http://www.internet2.edu/
[23] "Geant."
[24] "Rocketfuel."
[25] R. Draves, C. King, S. Venkatachary, and B. Zill, "Constructing optimal IP routing tables," in *INFOCOM*, 1999, pp. 88–97.
[26] B. Cain, "Auto aggregation method for ip prefix/length pairs." [Online]. Available: http://www.freepatentsonline.com/6401130.html
[27] X. Zhao, Y. Liu, L. Wang, and B. Zhang, "On the Aggregatability of Router Forwarding Tables," *INFOCOM*, pp. 1–9, Mar. 2010.
[28] H. Ballani, P. Francis, and J. Wang, "Making Routers Last Longer with ViAggre," in *NSDI*, 2009.