Managing Interdomain Traffic in Latin America: A New Perspective Based on LISP

Marcelo Yannuzzi and Xavi Masip-Bruin, Technical University of Catalonia
Eduardo Grampín, University of the Republic
Roque Gagliano, Latin American and Caribbean Internet Addresses Registry
Alberto Castro and Martín Germán, Technical University of Catalonia and University of the Republic

ABSTRACT

The characteristics of Latin American network infrastructures have global consequences, particularly in the area of interdomain traffic engineering. As an example, Latin America shows the largest de-aggregation factor of IP prefixes among all regional Internet registries, being proportionally the largest contributor to the growth and dynamics of the global BGP routing table. In this article we analyze the peculiarities of LA interdomain routing architecture, and provide up-to-date data about the combined effects of the multihoming and TE practices in the region. We observe that the Internet Research Task Force initiative on the separation of the address space into locators and identifiers can not only alleviate the growth and dynamics of the global routing table, but can also offer appealing TE opportunities for LA. We outline one of the solutions under discussion at the IRTF, the Locator/Identifier Separation Protocol, and examine its potential in terms of interdomain traffic management in the context of LA. The key advantage of LISP is its nondisruptive nature, but the existing proposals for its control plane have some problems that may hinder its possible deployment. In light of this, we introduce a promising control plane for LISP that can solve these issues, and at the same time has the potential to bridge the gap between intradomain and interdomain traffic management.

INTRODUCTION

Latin America (LA) is characterized by a low degree of peering connections among local domains, where most of the regional and transit traffic is exchanged outside LA, in U.S. network access points (NAPs). Internet service providers (ISPs) in LA typically purchase expensive connectivity from multiple transit providers, so one of their most important traffic engineering (TE) objectives is to load as much as possible their transit links. They also seek to improve both the performance and reliability of their transit traffic while minimizing costs. Unfortunately, the lack of investment in new network infrastructures increases the complexity, since transit traffic needs to be engineered over a bundle of links with heterogeneous capacities that are usually connected to different transit providers.

Currently, the process of finding the best trade-off for this challenging optimization problem is essentially manual, and tuned on a trial and error basis. The tools available today for Latin American ISPs are coarse-grained, and basically consist of the combination of Border Gateway Protocol (BGP)-based techniques, such as the utilization of variable prefix lengths together with AS-path prepending. The idea behind these techniques is to de-aggregate Internet Protocol (IP) prefixes, thus increasing the granularity of the BGP advertisements, allowing the distribution of traffic to be better controlled. The problem of this practice is that it fuels both the growth and dynamics of the global BGP routing table. Indeed, the average de-aggregation of IP prefixes in LA is twice as large as the global average. Issues like multihoming, suboptimal address allocation, and CIDR block de-aggregation for TE purposes, among others, are impacting on the scalability of the global BGP routing table, which soon will reach 300,000 entries [1]. This is a major concern for the Internet community [2].

To address this issue, the Internet Research Task Force (IRTF) is considering the separation of the address space into end-system identifiers (EIDs) and routing locators (RLOCs). The basic idea is that an EID represents an end-host IP address, while RLOCs represent the IP addresses where end hosts are located. The scaling ben-

This work was partially funded by the Spanish Ministry of Foreign Affairs and Cooperation through AECI-PCI A/019977/08.

 0163-6804/09/$25.00 © 2009 IEEE

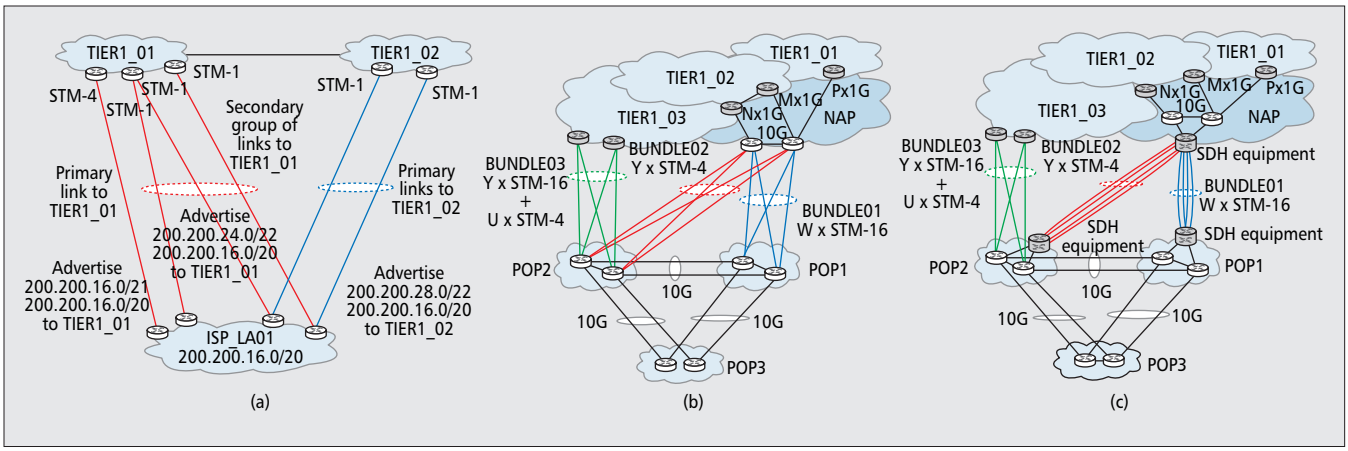


Figure 1. *a) Scenario #1: multihomed ISP with links of different capacity, load sharing, and backup routing policy; b) scenario #2: multihomed ISP with NAP presence; c) scenario #3: multihomed ISP with NAP presence and SDH multiplexers.*

efits arise when EID addresses are not routable through the Internet — only RLOC addresses are globally routable, allowing efficient aggregation of the RLOC address space.

Recent studies show that one of the solutions under discussion at the IRTF, the Locator/Identifier Separation Protocol (LISP) [3, 4], offers some key advantages. For instance, Quoitin *et al.* [5] show that the size of the global routing table can be reduced by roughly two orders of magnitude with LISP. That work also shows that LISP provides improved interdomain TE capabilities using a nondisruptive approach.

Despite these strengths, the proposals for the LISP control plane present some major challenges that are exposed and addressed in this article. These challenges lie in the fact that since EIDs are not globally routable through the Internet, a mapping system is necessary between EIDs and RLOCs.

In this article we provide up-to-date data about the characteristics and central issues of interdomain traffic management in LA. We outline the strengths of LISP, highlighting a set of important TE opportunities for LA, and we also introduce a novel control plane for LISP that deals with the challenges mentioned above. We discuss the importance of architecting the control plane so as to concurrently provide a highly efficient coupling to the DNS system, the path computation element (PCE) — if present — and an EID-to-RLOC mapping engine that borrows concepts from intelligent route control (IRC) techniques [6]. We conclude with directions for future research, especially in a promising field for LA we refer to here as intelligent route mapping (IRM).

Characterization of Interdomain Traffic Management in LA

In this section we describe some relevant characteristics of LA connectivity infrastructure, and examine their influence in aspects like interdomain routing and IP prefix de-aggregation policies.

LA Interdomain Routing Scenarios

Latin American Internet traffic has been growing at annual rates higher than 70 percent, and it will continue to grow steadily; therefore, capacity upgrade is of crucial importance to regional ISPs. In developed countries it is possible to plan capacity growth according to the observed usage and estimated traffic demands. Unfortunately, this is not the case for LA, where ISPs basically get the bandwidth they can buy, constrained to the available capacity along transit circuits. As a result, interdomain connectivity in LA is frequently composed of bundles of circuits of variable bit rates that are aggregated and used as primary as well as backup links. We proceed to illustrate some of these peculiarities by sketching three reference scenarios.

The first scenario, depicted in Fig. 1a, shows a small multihomed ISP with several links connected to two transit providers. In order to cope with the traffic growth, ISP LA01 has upgraded one of its primary links from an STM-1 to an STM-4. This is a common situation in LA, where economical and infrastructure constraints prevent ISPs from upgrading all its links at the same time.

The policy of provider ISP LA01 for managing its inbound traffic through links with different capacities is to use a combination of two BGP-based TE mechanisms, specifically, the de-aggregation of its prefix 200.200.16.0/20 and the utilization of AS-path prepending. More precisely, ISP LA01 splits the prefix 200.200.16.0/20 into three more specific prefixes: 200.200.16.0/21, 200.200.24.0/22, and 200.200.28.0/22. Prefixes 200.200.16.0/21 and 200.200.24.0/22 are advertised to TIER1 01 through the STM-4 and the secondary group of links, respectively, whereas prefix 200.200.28.0/22 is advertised to TIER1 02. In order to provide backup paths, ISP LA01 advertises the less specific prefix 200.200.16.0/20 to both TIER1 01 and TIER1 02. The effect of ISP LA01's inbound TE policy is that instead of simply advertising the prefix 200.200.16.0/20 to each transit provider, a total of six prefixes are advertised upstream. This de-aggregation of prefixes increases the size of the global BGP routing table. In addition, ISP LA01 may further

Region	IXPs	# of prefixes	De-aggregation factor (DF)
Africa	21	5K	3.46
Asia & Pacific	73	66K	2.81
Europe & Mid. East	123	67K	1.74
LA & Caribbean	24	26K	4.38
North America	88	124K	1.87
		Global BGP table	**Global average**
		288K	2.12

Table 1. *Statistics by region (data of April 2009, extracted from [1] and APNIC [7]).*

split the prefixes and will often tune its advertisements (on a trial and error basis), so as to adapt to variations in the inbound traffic. This, in turn, increases the dynamics of the global routing table.

In LA outbound traffic typically represents a small fraction of inbound traffic. However, the policy for managing outbound traffic requires special care, since the utilization of links with different capacities might degrade the performance of end users' applications, especially in frequent cases where the egress links are chosen using a round-robin scheme.

The second scenario considered here represents another frequent situation in LA, which comes up when the intercontinental connectivity from LA to the NAPs is part of the ISP network, meaning that the ISP owns termination equipment at both ends. This scenario is depicted in Fig. 1b, where the ISP backbone is composed of local (inexpensive) 10G bundles, and intercontinental (expensive) STM-x bundles to different transit providers. Note that traffic among POP1, POP2, and POP3 should be kept local, and clearly the choice of uplink/downlink for interdomain traffic might considerably affect the ISP's economics. The main complexity of this scenario lies in the utilization of bundles of links. This means that both inbound and outbound interdomain traffic policies should be further refined, and coordinated with intradomain routing, to optimize the usage of the expensive intercontinental links. The optimization of the distribution of traffic in this scenario usually involves a mixture of manual BGP-based TE techniques, TE based on tweaking the IGP metrics, and/or multiprotocol label switching (MPLS)-TE tunneling. These kinds of settings usually aggravate even more the de-aggregation of IP prefixes as well as the dynamics of the global BGP routing table.

The third scenario consists of a variation on the previous one, where the intercontinental connectivity is supported by aggregation at the synchronous digital hierarchy (SDH) layer (Fig. 1c). The strength of this solution is that it simplifies the layer 3 topology, since multiple links that were formerly connected at layer 3 are now managed as a single trunk. This approach has the potential to reduce the de-aggregation of IP prefixes. The weakness, on the other hand, is that the trunks are frequently composed of heterogeneous circuits, so the loss of granularity at layer 3 makes it extremely hard to fine-tune the distribution of traffic within the trunks.

Overall, these scenarios show the complexity of traffic management for ISPs in LA, dominated by largely manual trial-and-error procedures.

LA Peering Infrastructures and De-aggregation Factor

Table 1 shows the number of Internet exchange points (IXPs) per regional Internet registry (RIR). It is worth highlighting that 12 of the 24 Latin American IXPs are located in Brazil, since the latter gathers a large amount of LA's interdomain traffic. This suggests that TE solutions for LA may differ depending on the geographical area, since the network infrastructure in Brazil is substantially different from the rest of LA. Another characteristic of LA is the lack of significant regional content providers. LA consumes traffic mainly from the United States and Europe, showing only a small exchange of traffic among countries in the region. This limits the interest of large providers in deploying IXPs in LA, and is why most of the transit links are terminated in U.S. NAPs. However, peer-to-peer applications are changing the region's traffic profile, increasing the amount of regional traffic as countries share languages and cultural habits.

As described in scenario #1 in Fig. 1a, the de-aggregation of IP prefixes occurs when a domain advertises CIDR blocks with longer prefixes than those allocated by its RIR. To quantify this, the Internet community has defined the de-aggregation factor (DF), which represents a measure of the current routing table size vs. its aggregated size, and is formally defined as

$$DF = \left(\frac{\text{Prefixes in the Global Routing Table}}{\text{Aggregatable Prefixes}} \right). \quad (1)$$

Up-to-date values for the global routing table size and the DF are shown in Table 1 (columns three and four, respectively). The distribution of the DF in LA is shown in Fig. 2. The figure shows the DF vs. the number of upstream autonomous systems (ASs) for all the ASs registered in the LACNIC region. We observe that the larger DFs come from ASs with a small number of upstream ASs — many of them even with a single upstream AS. This suggests that several of the ASs that are connected to a few transit providers, leak their internal partitioning of prefixes in their BGP advertisements.

The large DF in LA is a consequence of:

- The problem of managing traffic over heterogeneous and complex infrastructures like the ones described in Fig. 1
- The intrinsic limitations of BGP-based TE techniques [8]
- The overloading of IP address semantics [2]

The adverse effects of the permanent increase in the DF are, fundamentally, the processing capacity needed by the routers supporting the global routing table (e.g., stringent memory,

CPU, and power requirements) and the impact on the BGP convergence time, since the larger the routing table, the slower the convergence. Moreover, the practice of adjusting the distribution of interdomain traffic based on the de-aggregation of IP prefixes adversely affects the dynamics of the global routing table.

In the next section we overview a novel approach that can dramatically reduce the size and churn rate of the global routing table, and at the same time offer a promising perspective for dealing with the peculiarities and complexities of interdomain traffic management in LA.

Overview of LISP

In April 2009 the Internet Engineering Task Force (IETF) chartered the LISP Working Group, and experimental Requests for Comments (RFCs) are expected by 2010. LISP [3, 4] uses IP-over-IP tunnels deployed between border routers located at different domains. The IP addresses configured on the external interfaces of the border routers act as RLOC addresses for the end systems in the local domain. Since an AS usually has several border routers, the local EID addresses can be reached through multiple RLOC addresses. LISP separates the address space into two parts, where only addresses from the RLOC address space are assigned to the transit Internet. Therefore, only RLOC addresses are routable through the Internet; EID addresses are considered routable only within their local domain. To illustrate the basics of LISP we use Fig. 3 (extracted from [3]). For a comprehensive understanding of LISP, the reader is referred to [3, 4].

LISP Data Plane

When local end host *S* with EID address 1.0.0.1 (Fig. 3) wants to communicate with end host *D* in a different domain whose EID address is 2.0.0.2, the following sequence of events occur in LISP.

The first step is the usual lookup of the destination address E_D in the DNS (E_D corresponds to 2.0.0.2 in the example in Fig. 3). Once E_D is obtained, the packets sourced from E_S traverse the domain and reach one of the local border routers. In LISP the latter are referred to as ingress tunnel routers (ITRs). Since only RLOC addresses are globally routable, when an ITR receives packets toward E_D, it queries the control plane to retrieve the E_D-to-RLOC mapping. After the E_D-to-RLOC mapping resolution, the ITR encapsulates and tunnels packets between the local RLOC address (ITR address 11.0.0.1 in the example) and the RLOC address retrieved from the mapping system, the egress tunnel router (ETR) address in LISP terminology (either 12.0.0.2 or 13.0.0.2 to E_D depending on the mapping). At the destination domain, the ETR decapsulates the packets received through the tunnel and forwards them to E_D — which, as mentioned above, is locally routable within the domain. From the first packet received, the ETR caches a new entry, solving in this way the reverse mapping for the packets to be tunneled back from E_D to E_S.

As shown at the bottom of Fig. 3, the mapping system can return multiple RLOC addresses for the same destination. Each of the entries returned has a *priority* and a *weight* attribute. The priority determines the order in which the ETRs must be selected, while the weight tells how to distribute the traffic among ETRs with the same priority. In the example the priorities

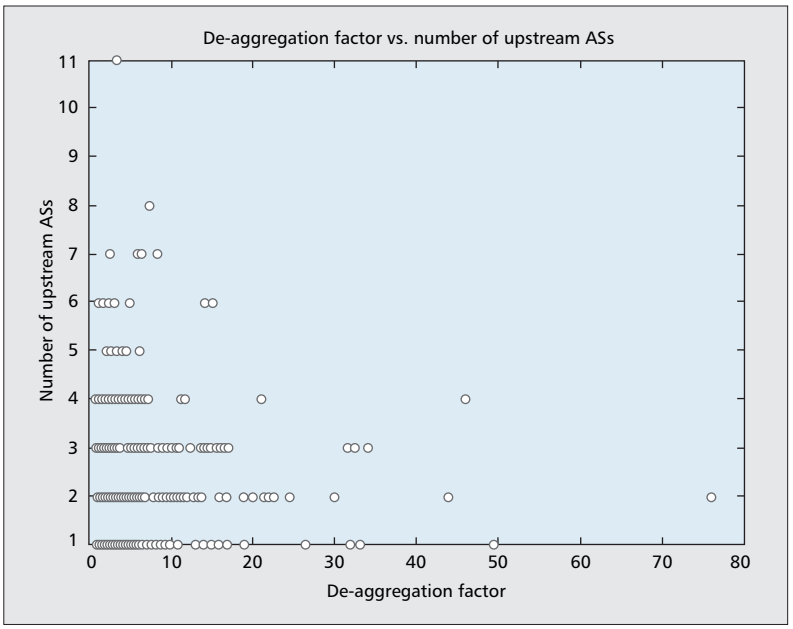


Figure 2. *Distribution of the de-aggregation factor as a function of the number of upstream providers in Latin America (data of April 2009, extracted from [1] and APNIC [7]).*

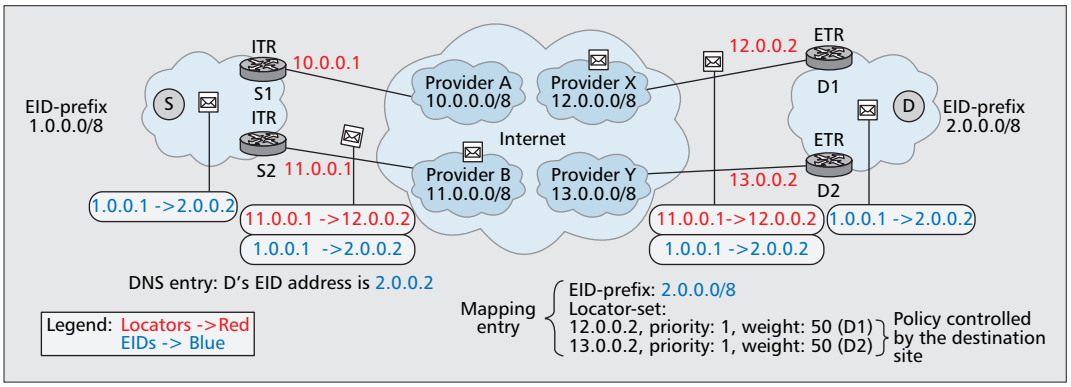
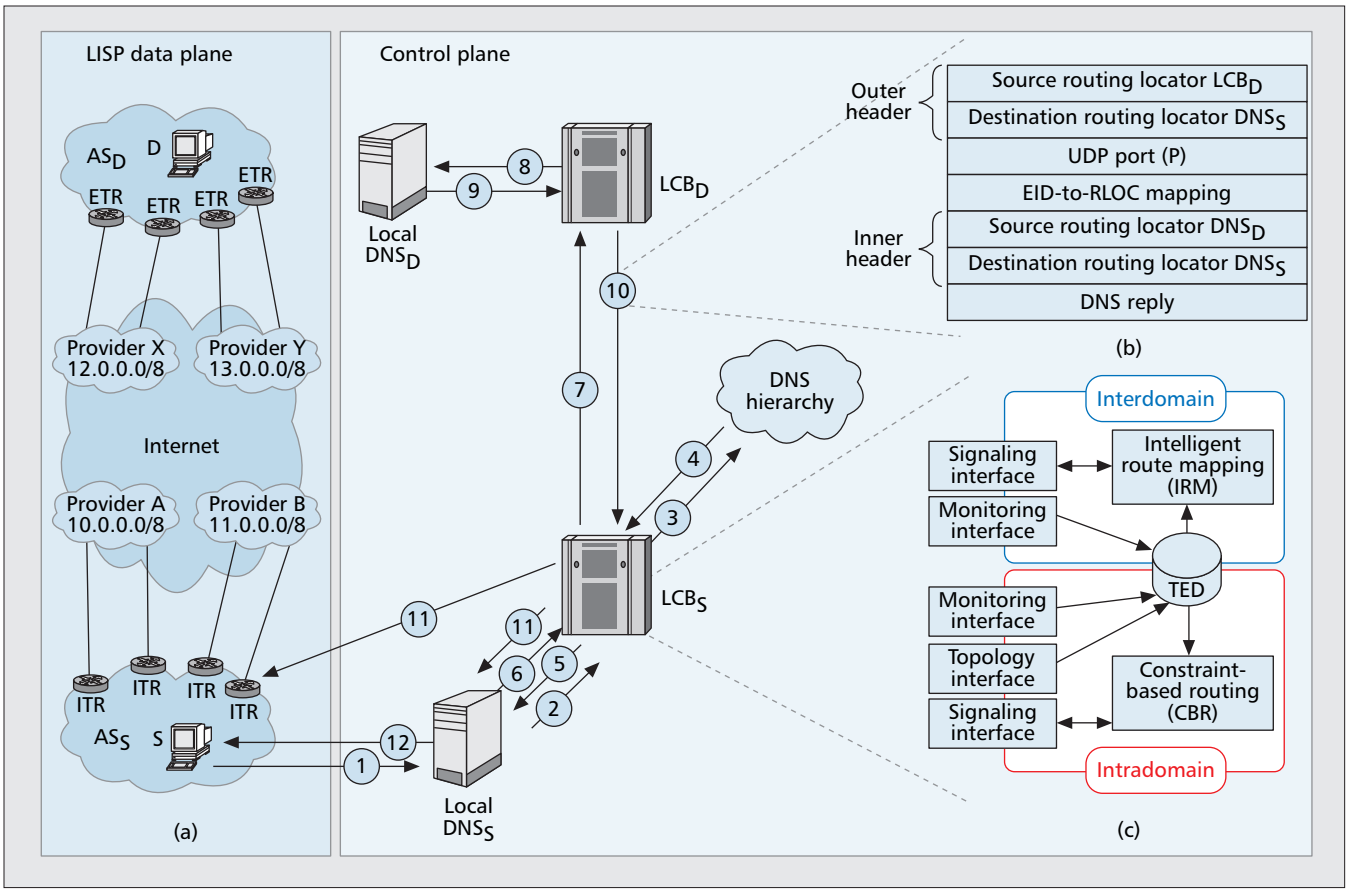


Figure 3. *The basics of LISP.*



■ **Figure 4.** *Proposed control plane architecture.*

and weights of ETRs D1 and D2 are equal, so the traffic from *S* to *D* will be balanced between D1 and D2.

Overall, LISP has three major advantages. First, it does not introduce major changes to the routing system, and therefore it might be feasible to implement and deploy in the near future. Second, it has the potential to significantly reduce the size of the global routing table [5]. Third, the mapping system brings a wide set of TE opportunities, which in principle, can reach a granularity of a /32 prefix without impacting on the size or dynamics of the global routing table. Nevertheless, it is important to notice that special care must be taken, since LISP might end up moving the scalability issues from the global routing table to the global mapping system.

LISP-Based TE Opportunities for LA

Let us consider now the application of LISP to the three reference scenarios described in Fig. 1.

Scenario #1 — LISP enables multihomed sites to completely avoid running BGP. The only IP prefixes that need to be advertised to the global routing system are those of the wide area network (WAN) interfaces of the border routers (i.e., the ITRs'/ETRs' external addresses). In this framework the load sharing and backup policies of ISP LA01 in Fig. 1a can be managed by intelligent EID-to-RLOC mapping functions, which may be dynamically tuned using the LISP weight and priority attributes, respectively. The different capacities of the links can be used during the mapping resolution process, by appropriately unbalancing the weight attribute of the different links (e.g., the weights advertised could keep the same proportion as the link capacities). The advantages of applying LISP in this scenario are evident, since LA ISPs are released from the burden of handcrafting their BGP configurations. In addition, these advantages can be achieved without de-aggregating prefixes, and therefore without adversely impacting on the size or the dynamics of the global routing table.

Scenario #2 — In this case the TE problem addressed is twofold, since not only must the interdomain mapping function be solved, but also the traffic should be load balanced over heterogeneous links among the upstream providers and the ISP's regional points of presence (POPs). We argue that tools like constraint-based routing (CBR) (for the intradomain part) and an IRM engine borrowing concepts from IRC techniques [6] (for the interdomain part) can work together to accomplish these objectives. At least two appealing solutions can be sketched to drive the intradomain traffic according to the ISP policies. One option is to set up MPLS label switched paths (LSPs) on the fly for different flows, in order to *stitch* the inbound interdomain traffic to the appropriate *pipe* downstream (i.e., to the corresponding ISP POP). This is the usual task of the PCE [9]. The second option is to introduce the idea of interior

ITR/ETRs, which can perform internal mapping from/onto border routers.

Scenario #3 — Similar to the previous scenario, a combination of CBR and IRM may apply here as well. Note that the utilization of SDH bundles suggests that cross-layer TE techniques might also be needed.

LISP Control Plane

The basic role of the control plane is to provide the mapping system. Among the proposals under discussion are LISP-ALT, LISP-CONS, and LISP-DHT [4, references therein]. At present, the more developed proposals for a LISP control plane present at least three problems. First, the initial packets sent from a source EID address (E_S) to a destination EID address (E_D) can be dropped at the ITR during the EID-to-RLOC mapping resolution. Although caching techniques are being proposed to store the mappings at ITRs, a hit might not necessarily be found, because the mapping either has aged out or simply was never requested before.

Second, LISP might considerably increase the latency to start up the communication between end systems. Considering the one-way delay (OWD), a TCP connection between two end systems is established roughly around

$$T_{\text{DNS}} + 2OWD(E_S,E_D) + OWD(E_D,E_S), \quad (2)$$

whereas with LISP (Fig. 3) it would roughly demand

$$T_{\text{DNS}} + T_{\text{resol}}^{\text{map}} + 2OWD(E_S,E_D) + OWD(E_D,E_S) \quad (3)$$

under the usual assumption that ITRs and ETRs can encapsulate/decapsulate at line rate.

Third, for each traffic flow from S to D, the egress ITR is also used as the local ETR for the packets sent from D to S. This is to avoid a two-way mapping resolution, which would increase even more the latency shown in Expression 3. Clearly, this introduces a limitation in terms of inbound TE, which is particularly important in the case of LA, given that outbound and inbound traffic management policies typically do not match.

To cope with these issues, alternative solutions are being discussed. These alternatives require either some major changes to the DNS system, the addition of some debatable features to border routers, or using the control plane to transport data while the mapping is being resolved.

Control Plane Proposal

Our goal is threefold. First, we aim to prevent the potential dropping of packets at the ITRs while the EID-to-RLOC mapping resolution is being computed. Second, we aim to obtain and configurie the corresponding mapping during the normal DNS resolution process for destination E_D, and we want to achieve this goal without introducing changes to the DNS system. More precisely, we seek

$$T_{\text{DNS}} + T_{\text{resol}}^{\text{map}} \approx T_{\text{DNS}}. \quad (4)$$

Third, we aim to have the TE flexibility to choose different local ITR and ETR LISP routers for any given flow sourced at the domain.

Architecture

To pursue these objectives, we propose the scheme and set of steps depicted in Fig. 4. We introduce an entity we call a LISP control box (LCB), which might be a standalone device or run as an instance of a PCE, implemented as described in [10] and tailored for this purpose.

Step 1 — S queries DNS_S to obtain the EID address of destination D. The LCB (LCB_S) obtains the EID address of S (E_S) by interprocess communication (IPC) with DNS_S (e.g., using sockets) and computes the local RLOC to be used for the reverse mapping (i.e., for the incoming traffic from E_D to E_S) based on TE constraints. The algorithms used to determine the ingress RLOC (RLOC_S) are inherently the same used today by IRC techniques [6]. The E_S-to-RLOC_S mapping computation is performed by the IRM module of LCB_S.

Steps 2–9 — The LCBs are in the data path of the DNS servers; therefore, they can transparently analyze the exchange of DNS messages. Steps 2–9 represent the usual flow of iterative DNS queries performed by DNS_S, and the corresponding replies received from the DNS servers in the hierarchy (root server, top-level domain server, etc.). It is important to observe that these steps require the use of iterative queries between DNS servers. This approach is in line with the current trend of avoiding recursion between DNS servers. Indeed, supporting iteration is mandatory, whereas recursion is strictly optional (by default BIND9 performs iterative queries).

Step 10 — When LCB_D detects that the reply issued from DNS_D carries the address E_D, it encapsulates the reply into a UDP message; clearly variants of this approach can work as well-with source address LCB_D, destination address DNS_S, and a special transport port P that will be listened to by LCB_S at the source domain S (Fig. 4b). The payload of the outer packet contains the EID-to-RLOC mapping for E_D. It is worth highlighting that the mapping selection performed at LCB_D is precomputed by an online IRC engine running in the background (as mentioned in step 1, this computation is performed by the IRM module shown in Fig. 4c), so the mapping is always known beforehand. This means that LCB_D can encapsulate the answer from DNS_D roughly at line rate.

Step 11 — LCB_S is in the data path of DNS_S, so when LCB_S detects a packet toward DNS_S using port number P, it intercepts and decapsulates the packet and forwards the usual DNS answer to DNS_S. From the outer packet LCB_S discovers the address of LCB_D, retrieves the mapping for E_D, and configures through the signaling interface of the LCB all the ITRs according to that mapping. The advantage of pushing the mapping to all ITRs is that AS_S can carry out local TE actions and move part of its internal traffic without caring whether a mapping will

To cope with these issues, alternative solutions are being discussed. These alternatives either require some major changes to the DNS system, the addition of some debatable features to border routers, or using the control plane to transport data while the mapping is being resolved.

To have a preliminary evaluation of the performance of our control plane, we have prototyped the LCB in a standard Linux PC. For simplicity, our prototype is built into BIND9, meaning that the results shown here were obtained by integrating the LCB functionality into the DNS.

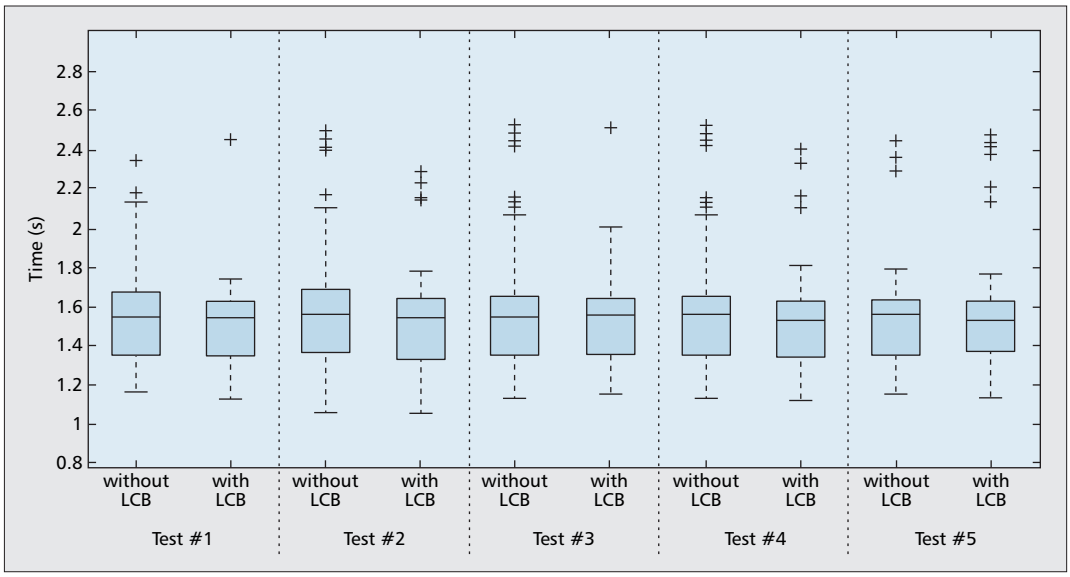


Figure 5. *Five tests showing the time distribution of a set of 1000 DNS lookups over the Internet. Each test corresponds to a round of 200 DNS lookups, 100 without LCBs, and 100 with LCBs.*

be in place in the relevant ITRs after TE optimization. The mapping information pushed to the ITRs consists of the tuple (E_S,E_D, $RLOC_S$, $RLOC_D$), supporting the utilization of two independent one-way tunnels depending on the reverse mapping computed by LCB_S during step 1. In other words, an ITR is capable of forwarding traffic to E_D, using as source address in the encapsulation an RLOC address that might be different from the addresses configured on its WAN interfaces (e.g., using source NAT) — although restricted to the pool of addresses assigned by the ISP to which the ITR is connected. Like in Fig. 4a, this approach helps AS_S exploiting the usual multiconnectivity to each of its providers.

Step 12 — DNS_S responds the usual DNS query to *S*.

After the usual DNS resolution process, E_S starts sending packets toward E_D, with the advantage that the mapping has already been configured at the ITRs, hence avoiding the potential dropping of packets. As we shall show in the next section, the overall process (steps 1–12) can be completed in approximately T_{DNS}, which we claim should be used as the upper bound for solving the mapping.

When the first data packet reaches the corresponding ETR in AS_D, the ETR:

1 Decapsulates the packet and forwards the inner packet to E_D
2 Obtains the reverse mapping (i.e., the E_S-to-$RLOC_S$ mapping)
3 Pushes this mapping to the rest of the ETRs and also updates the LCB_D mapping system via multicast or another mechanism

This action completes the two-way mapping resolution process. An interesting point is that our control plane allows each domain to achieve its TE policies congruently, since each domain has the freedom to independently decide its ingress and egress mappings.

In order to achieve the aforementioned goals, the IRM engine of the LCB needs to know the local topology and the availability of local resources. As shown in Fig. 4c, a promising approach is to provide a highly efficient coupling between the IRM engine of the LCB and the PCE — the basic components of a PCE are represented in the intradomain portion of Fig. 4c. Indeed, both can share and feed the TE database (TED), which is crucial to accomplish both intradomain and interdomain TE objectives concurrently. The LA scenarios shown in Figs. 1b and 1c can benefit considerably from this approach, since traffic can be coordinatedly balanced by the LCB and PCE, making possible the selection of a upstream/downstream forwarding path with high granularity (e.g., for prefixes or even end systems).

Preliminary Results

In order to have a preliminary evaluation of the performance of our control plane, we have prototyped the LCB in a standard Linux PC. For simplicity, our prototype is built into BIND9, meaning that the results shown here were obtained by integrating the LCB functionality into the DNS. More precisely, in Fig. 4 DNS_S and LCB_S are integrated in a single Linux PC, and the same applies to DNS_D and LCB_D. It is worth emphasizing that the control plane architecture proposed in this article is more general in scope; thus, more advanced implementations of the LCB prototype may not be embedded in BIND9 and may run as standalone devices.

Our initial goal is to verify that Eq. 4 holds when the LCBs are used. To this end, we performed a set of 1000 rounds of DNS lookups on the Internet. These rounds were clustered in five different tests (Fig. 5), each corresponding to a sequential set of 200 DNS lookups, half of them without the intervention of the proposed control plane and the other half utilizing our prototype. The five tests were carried out at different moments in time, where we measured the total time of the DNS lookups from the source E_S.

In our tests E_S and DNS_S/LCB_S were located at the University of the Republic premises in Montevideo, Uruguay, and E_D as well as DNS_D/LCB_D were located at the Technical University of Catalonia premises in Barcelona, Spain.

In this setting steps 2–9 in Fig. 4 are solved iteratively using the usual DNS hierarchy. As mentioned above, our goal is to assess if the control plane is able to obtain and configure the mapping during the usual DNS resolution process. Therefore, prior to each of the 1000 tests (i.e., both with and without LCBs), we flushed the DNS_S cache, enforcing in this way the lookup of E_D through the DNS system. This is only to ensure that during the trials, the validation of Eq. 4 is not biased by the DNS_S cache. It is important to observe that in an operational scenario, if E_D is cached in DNS_S, a mapping entry will already be configured at the ITRs in AS_S, so Eq. 4 will trivially hold; analysis of the potential interdependencies between the DNS and LISP caches and their aging policies is part of our future work.

After the steps 2–9 in Fig. 4, and upon detection of authoritative responses for A, AAAA, and MX record types in DNS_D, we obtain the EID-to-RLOC mapping for E_D (as mentioned before this is computed aforehand), and then proceed with steps 10–12.

The results for 1000 experiments are shown in Fig. 5. The horizontal mark inside each box is the median, the edges of the boxes are the 25th and 75th percentiles, and the whiskers cover the rest of the data gathered, excluding outliers. The outliers are represented as crosses and they are shown individually. Our preliminary results are promising, since Fig. 5 confirms that a simple prototype implementation of the control plane is able to accomplish the goal targeted in Eq. 4.

Discussion and Concluding Remarks

In this article we have presented the peculiarities of LA network infrastructures, and we have analyzed the TE opportunities that LISP may bring to the region, especially by developing customized intelligent route mapping techniques.

Our main contribution is the proposal of a control plane that tackles the challenges exposed in LISP. Among the strengths of this control plane are the following. First, the mappings can be configured during the normal DNS resolution process, and this can be implemented without introducing changes to the DNS system; the zone files, resource records, and so on remain unchanged. Second, by placing the LCBs in the data path of the DNS system, it is possible to transparently use the latter as a discovery mechanism of LCBs. Once a remote LCB is discovered, direct communication and even cooperation among LCBs might be exploited, adding therefore, extra capabilities to the control plane. Indeed, our control plane can support an overlay of LCBs, enabling on-the-fly refinement of the mappings toward (from) a set of popular destinations (sources) of a domain. Third, our control plane is technically supported by an architecture that blends two ongoing initiatives in the IETF, the PCE and LISP, which offers a promising perspective for LA, especially for the scenarios in Fig. 1.

Although the proposed control plane offers a promising approach, several aspects of the architecture need to be further explored. For example, both the DNS servers and the LCBs are end systems whose IP addresses are, in principle, EIDs (i.e., they are non-globally routable). This states the obvious problem that a *resolver of resolvers* is required. An option is to assign globally routable addresses to both of them, but arguments can be found against this too. In any case, the debate about the assignment of the address space to DNS servers is present in any locator-identifier separation scheme and must be appropriately addressed.

Issues such as how to solve mappings that, in principle, do not require a DNS resolution (e.g., ping 10.10.10.1), and the potential impact of these solutions on the overhead and dynamics of the DNS system need to be analyzed and evaluated. Other important issues to be explored are the security aspects of this control plane.

Issues such as how to solve mappings that, in principle, do not require a DNS resolution (e.g., ping 10.10.10.1), and the potential impact of these solutions on the overhead and dynamics of the DNS system, need to be analyzed and evaluated.

References

[1] CIDR Report; http://www.cidr-report.org/as2.0/
[2] D. Meyer, L. Zhang, and K. Fall, "Report from the IAB Workshop on Routing and Addressing," IETF RFC 4984, Sept. 2007.
[3] D. Meyer, "The Locator Identifier Separation Protocol (LISP)," *Cisco Internet Protocol J.*, Mar. 2008, pp. 23–36.
[4] D. Farinacci *et al.*, "Locator/ID Separation Protocol (LISP)," work in progress, May. 2009; IETF draft-ietf-lisp-01.txt.
[5] B. Quoitin *et al.*, "Evaluating the Benefits of the Locator/Identifier Separation," *Proc. ACM SIGCOMM MobiArch*, Kyoto, Japan, Aug. 2007.
[6] M. Yannuzzi *et al.*, "Improving the Performance of Route Control Middleboxes in a Competitive Environment," *IEEE Network*, vol. 22, no. 5, Sept./Oct. 2008, pp. 56–64.
[7] BGP Routing Table Analysis; http://thyme.apnic.net/
[8] R. Gagliano, "IPv4 De-aggregation in LACNIC Region-LACNIC XI," May 2008; http://www.lacnic.net
[9] A. Farrel, J. P. Vasseur, and J. Ash, "A Path Computation Element (PCE)-Based Architecture," IETF RFC 4655, Aug. 2006.
[10] E. Grampín and J. Serrat, "Cooperation of Control and Management Plane for Provisioning in MPLS Networks," *9th IFIP/IEEE Int'l. Symp. Integrated Net. Mgmt.*, Nice, France, May 2005.

Biographies

MARCELO YANNUZZI (yannuzzi@ac.upc.edu) received a degree in electrical engineering from the University of the Republic (UdelaR), Uruguay, in 2001, and D.E.A. (M.Sc.) and Ph.D. degrees in computer science from the Department of Computer Architecture, Technical University of Catalonia (UPC), Spain, in 2005 and 2007, respectively. He is with the Advanced Network Architectures Laboratory (CRAAX) at UPC, where he is an associate professor. He held previous positions with the Physics Department of the School of Engineering, UdelaR (1997–2003), and the Electrical Engineering Department of the same university (2003–2006). He worked in the industry for 10 years at the national telco in Uruguay (1993–2003).

EDUARDO GRAMPIN (grampin@fing.edu.uy) holds a degree in electrical engineering from UdelaR, and he got his Ph.D. from UPC in 2005. Currently, he is an associate professor at the Computer Science Department of the School of Engineering, UdelaR. Between 1993 and 2006 he worked at the National Telecommunications Administration (ANTEL) in Uruguay. He participated in the deployment of commercial Internet services in Uruguay in the mid–1990s,

and has worked on different aspects of planning and management of IP networks in ANTEL, mainly linked to projects toward a convergent architecture based on IP/MPLS, with an integrated management system.

ROQUE GAGLIANO (rgaglian@fing.edu.uy) received his electrical engineering degree from UdelaR in 2001, and his M.S. in electrical engineering from the University of Kansas in 2005. He worked as an area engineer in ANTEL between 1999–2003 and 2006–2008. He held previous positions with the Physics Department, UdelaR (1997–2002), Sprint Nextel Corp, Kansas City, Missouri (2005–2006), and South-East and ITTC Engineering, University of Kansas (2004–2005). He is now with LACNIC in Uruguay, and since 2002 he is also an assistant professor in the Electrical Engineering Department, UdelaR.

ALBERTO CASTRO (acastro@ac.upc.edu) got his engineering degree in computer science from UdelaR in 2007. Since 2008 he is a Ph.D. student working at CRAAX, UPC. He is an assistant professor in the Computer Science Department of the School of Engineering, UdelaR. His research interests are in IP/MPLS interdomain routing, peer-to-peer networks, network performance, and interdomain traffic engineering.

MARTIN GERMAN (mgerman@ac.upc.edu) received his computer science degree from UdelaR in 2007. He is a Ph.D. student at CRAAX, UPC. He is also an assistant professor in the Computer Science Department of the School of Engineering, UdelaR. His research interests are in the design of scalable routing paradigms for the Internet, novel routing/switching paradigms, network performance, and traffic engineering.

XAVI MASIP-BRUIN (xmasip@ac.upc.edu) received M.S. and Ph.D. degrees from UPC, both in telecommunications engineering, in 1997 and 2003, respectively. He is with CRAAX at UPC, where he is currently an associate professor in computer science. Since 2000 he has participated in many international and Spanish research projects. His publications include around 60 papers in national and international refereed journals and conferences. His current research interests lie in broadband communications, QoS management and provision, and traffic engineering.