An Approach to a Fault Tolerance LISP Architecture

A. Martínez, W. Ramírez, M. Germán, R. Serral, E. Marín, M. Yannuzzi, X. Masip-Bruin

An Approach to a Fault Tolerance LISP Architecture[1]

A.Martínez, W.Ramírez, M.Germán, R.Serral, E.Marín
M.Yannuzzi, X.Masip-Bruin

Advanced Network Architectures Lab (CRAAX),
Universitat Politècnica de Catalunya (UPC)
Neàpolis Building, Rbla. Exposició, 59-69 – 08800 Vilanova i la Geltrú, Spain
{anny, wramirez, mgerman, rserral, eva, yannuzzi, xmasip}@ac.upc.edu

Abstract. Next Generation Internet points out the challenge of addressing "things" on both a network with (wired) and without (wireless) infrastructure. In this scenario, new efficient and scalable addressing and routing schemes must be sought, since currently proposed solutions can hardly manage current scalability issues on the current global Internet routing table due to for example multihoming practices. One of the most relevant proposals for an addressing scheme is the Locator Identifier Separation Protocol (LISP) that focuses its key advantage on the fact that it does not follow a disruptive approach. Nevertheless, LISP has some drawbacks especially in terms of reachability in the border routers. In face of this, in this paper we propose a protocol so-called LISP Redundancy Protocol (LRP), which provides an interesting approach for managing the reachability and reliability issues common on a LISP architecture, such as those motivated by an inter-domain link failure.

Keywords. Internet Routing Scalability, LISP, Fault Tolerance

1. Introduction

In the past years, early forms of ubiquitous communication have arisen and become more evident, as society expectations towards technology increases. These facts seem to prove that current Internet will naturally evolve to an Internet of Things (IoT) as a new dynamic communication scheme where objects, services, spaces and even people may be given a unique number, almost avoiding barriers for re-cognizing, locating, addressing, reaching, controlling and enjoying almost anything via the Internet, through a mix of heterogeneous wired and wireless network infrastructures.

[1] This work was partially funded by the Spanish Ministry of Science and Innovation under contract TEC2009-07041, and the CIRIT (Catalan Research Council) under contract 2009-SGR1508.

The current Internet semantic overloading of addresses, where addresses are simultaneously referred to identifiers and locators of a node, is an important constraint over mobility and scalability concerns. In a world where a huge volume of objects can be uniquely identified and also where objects capacity of mobility increases over time, decoupling of naming and location seems to be one of the first steps towards the evolution to a IoT. Several factors, such as the rise of multihoming sites, semantic overloading of IP addresses, among others, affect the scalability of the global Internet routing table, fueling its size and dynamics. Several proposals have emerged to solve these issues: SHIM6 [1], Six/One Router [2], HIP [3], Multipath TCP [4], GSE [5] and LISP [6] are just a few to name. The latter seems to be one of the strongest as already considered by the Internet Engineering Task Force (IETF).

The LISP concept is based on the idea of decoupling host identifier (Endpoint identifier, EID) and host localization (Routing locator, RLOC). The main benefit from using LISP is that while locators are AS-level distributed (allocated to the external interfaces of the border routers) the identifiers are only locally distributed, therefore reducing the overall routing load throughout the network. While from a deployment perspective LISP is a non-disruptive approach, the existing control plane proposals have some drawbacks especially in terms of reachability in the border routers. Since these border routers are responsible for carrying out the mapping between EIDs and RLOCs, these issues may hinder its possible deployment. In this paper, we conceptually propose the basics of a new protocol, so-called LISP Redundancy Protocol (LRP) which is designed for managing the reachability and reliability issues, such as those motivated by either an inter-domain link failure, a border router failure, an intra-domain link/node failure or an intra-domain TE action.

The rest of this paper is organized as follows. First, section 2 presents the current Internet architecture and its scalability problems; section 3 introduces LISP basics; section 4 presents the LRP as a potential solution for the control plane problems; section 5 discusses future work related issues. Finally, section 6 presents conclusions.

2. Background and Problem statement

The Internet architecture is organized into Autonomous Systems (ASes). The ASs interconnections generate a hierarchy between different Internet Service Providers (ISPs). In this hierarchical network structure, the Internet routing system is largely based in the Border Gateway Protocol (BGP) [8], a long lived path-vector protocol which is used to exchange reachability information between ASes. Being a policy-routing protocol, it provides operators with the freedom to express their enterprise requirements and policies, allowing the attachment of several attributes for each route or network prefix. However, it has been largely demonstrated in the literature that the currently deployed Internet architectural model suffers from some weaknesses, mainly on the terms of routing scalability issues that along with the specific problems on the Internet addressing scheme require the deployment of new solutions. Next subsections detail most relevant aspects limiting routing performance on the overall Internet. It is worth highlighting that any solution proposed for end-to-end addressing must consider a completely heterogeneous network scenario consisting of different wired and wireless network segments distributed

along the route, where some user (e.g., quality) and network (e.g., physical attributes) requirements must be met.

2.1 Internet Routing Scalability problems

Recent studies including the Internet Architecture Board (IAB) report [9], reveal that Internet routing is facing serious scalability problems, all involving both the size and dynamics of the global routing table in the Internet's Default Free Zone (DFZ).

The global routing table size in the DFZ has been growing at an alarming rate in recent years [10], till reaching now a total of 36.717 ASes that originate 355.262 IPv4 prefixes (see Figure 2) despite several limitations such as lack of IPv4 addresses, strict address allocation and routing announcement policies. Although IPv6 deployment would remove the problem of lack of IPv4 addresses, there is a strong concern that the deployment of IPv6 on a large scale could result in a significant growth of the routing table.

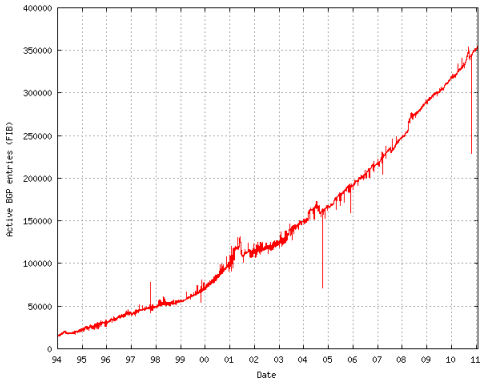

Fig. 1 Growth of the BGP Tables at DFZ Routers

The IAB report [9] identified the following factors as the main reasons behind the rapid growth of the global routing table in the DFZ:

- Multihoming.
- Traffic engineering.
- Non-aggregable address allocations

In [11] authors conclude that address fragmentation, caused by multi-homing and load balancing is the major reason of BGP table growth.

2.2 Dynamics of the BGP Control Plane Information

In [12], a systematic study of highly active prefixes is presented, concluding that a small fraction of advertised prefixes are responsible for a relevant amount of churn in BGP; furthermore, they found that some generators of BGP beacons, used for active monitoring of BGP updates, appear as highly active. Despite the big amount of related work, the dynamics of the BGP control plane information (i.e., the exchanging of updates messages due to the advertisement of new prefixes) remains unknown, but certain evidence exists of Long Range Dependence [13]. As BGP propagates changes to the best path, a single router may send multiple updates based on one triggering event, and further, cause induced updates at other locations; examples of such events are link failures, newly added networks, prefix deaggregation and policy changes, among others. Moreover, it is important to notice that, since the routing information is subject to successive filtering by internal ASes policies, any route view of the network is always partial, determined by the local point of observation. On the other hand, a relatively small number of ASes are responsible for a disproportionately large fraction of the update churn that is observed today. In turn, another problem motivated by the growing of the BGP updates is the BGP convergence time, since as the larger the topology complexity is the longest the convergence time, hence motivating the network to take longer to recover from failures.

2.3 Multihoming Sites

Another factor related to the growth of the routing table refers to the multihoming sites. For a network edge to be reachable by any service provider, the network-edge address-prefixes should be visible in the global routing tables. Meaning that no service provider can aggregate these address prefixes within their own address prefix, even if the network edges have addresses that belong to the provider-assigned address block. In addition, the network edges are increasingly obtaining provider-independent addresses from the Regional Internet Registries (RIRs), in order to avoid the renumbering every time a change of service provider happens. In summary, the topological information based on prefix-aggregation per provider is badly altered by multihoming, and in turn, leads to rapid growth of the global routing table.

2.4 Semantic Overloading

Another critical problem is the semantic overloading of IP addresses. This is because an IP address identifies a host (in fact its interface), and also serves to locate the host on the network. In this addressing scheme when a host changes of network provider, its IP address changes, therefore changing not only the network providing host access but also the host identifier. For upper layer applications that have IP addresses hard-coded for a host, this represents a severe mobility constraint. In short, the semantic overloading of IP addresses is the main problem when talking about renumbering a network.

3. LISP overview

LISP uses IP-over-IP tunnels deployed between border routers located at different domains. The IP addresses allocated to the external interfaces of the border routers act as Routing Locator (RLOC) addresses for the end systems in the local domain. Since an AS usually groups several border routers, the local Endpoint Identifier (EID) addresses can be reached through multiple RLOC addresses. Hence, LISP separates the overall 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, that is, EID addresses are considered routable only within their local domain. In addition, a number of scaling benefits would be realized by separating the current IP address into two different spaces; among them are:

- Reduction of the routing table size in the Default Free Zone (DFZ)
- More cost-effective multihoming for sites that connect to different service providers
- Easy renumbering when clients change providers
- Traffic engineering capabilities
- Mobility without address changing

The basic idea is that an EID represents an end-host IP address, while RLOCs represent the IP addresses where end hosts are located. At border routers EIDs are mapped into RLOCs, following a map-and-encap scheme, a basic mechanism of a LISP architecture. The scaling benefits 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 LISP offers some key advantages. For instance, authors in [14] show that the size of the global routing table can be reduced by roughly two orders of magnitude with LISP. Next subsection details how LIST performs on both the data and control planes.

3.1 Data Plane

Data plane performance is described on the example shown in Figure 3[2]. When the local end host S with EID address 190.1.1.1 wants to communicate with end host D with EID address 200.1.1.2 in a different domain, the following sequence of events occur in LISP:

1) The first step is the usual lookup of the destination address ED in the DNS.
2) Once ED is obtained, the packets sourced from ESource traverse the domain and reach one of the local border routers. In LISP the latter are referred to as Ingress Tunnel Routers (ITRs).
3) Since only RLOC addresses are globally routable, when an ITR receives packets toward ED, it queries the control plane to retrieve the EDestination-to-RLOC mapping.
4) After the ED-to-RLOC mapping resolution, the ITR encapsulates and tunnels packets between the local RLOC address (ITR address 3.3.3.2 in the example) and

[2] This example is extracted from "The Locator Identifier Separation Protocol (LISP)", by David Meyer, published in Internet Protocol Journal, vol 11, nº:1

the RLOC address retrieved from the mapping system, the Egress Tunnel Router (ETR) address in LISP terminology (either 4.4.4.2 or 10.0.0.2 to ED depending on the mapping).

5) At the destination domain, the ETR decapsulates the packets received through the tunnel and forwards them to ED — 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 $E_{Destination}$ to E_{Source}.

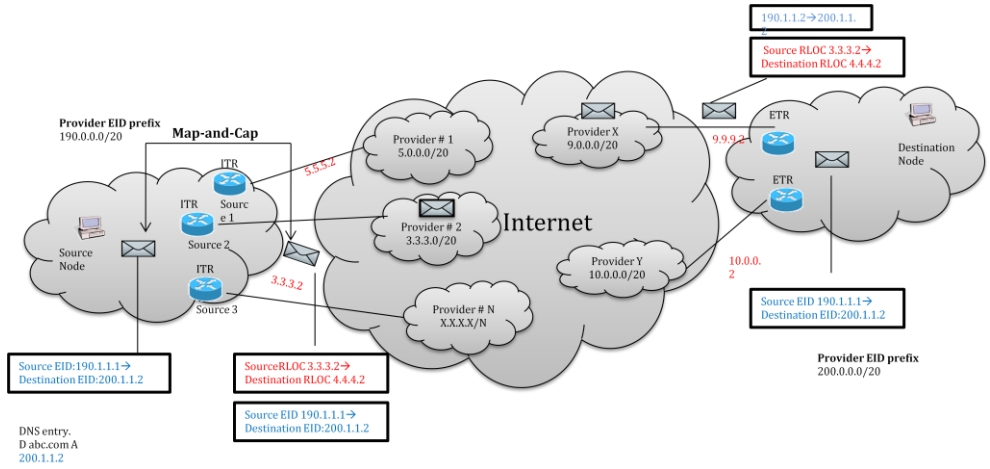

Fig. 2 The basics of LISP

3.2 Control Plane

Despite the benefits of using LISP described in section above, the proposals for the LISP control plane present some major challenges. 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. LISP does not specify a mandatory mapping system, and as a consequence, different proposals can be found in the recent literature, such as ALT [15], NERD [16] or Map Server [17].

Besides, in [18] we introduced a new control plane for LISP; the new control plane presents an improvement on three aspects respect to the existing solutions; (i) firstly, "First packets drop problem" when an ITR does not have a mapping for an EID-prefix; (ii) secondly, potential increase in the latency to start a communication due to the mapping resolution; and (iii) in order to avoid a two-way mapping resolution, the ITR is used as the local ETR for the packets sent from D to S. The latter introduces limitations in terms of inbound Traffic Engineering, especially, when outbound and inbound traffic policies do not match. Despite these improvements there are other issues relating to reachability and reliability that have not been resolved, such as those motivated by an inter-domain link failure.

4. Making the way to a fault tolerance LISP.

In [18] authors propose a new LISP control plane, aimed to overcome the issues derived by current mapping systems such as ALT, CONS or NERD. The main issue behind these approaches is related to the first packet drop problem, which refers to the mapping resolution for a prefixed EID for the first outgoing packet, this also derives in the potential increase in the latency to start a communication. This newly proposed control plane is based on the idea of retrieving EID-to-RLOC mappings within the DNS Resolution time. A major shortcoming of the solution proposed in [18] is that the mappings between EIDs and RLOCs are replicated in all of the edge routers within the same AS. Despite the fact that this option ensures improved reachability, unfortunately it may bring scalability problems since each router must store mapping information that rarely needs to use (i.e. the mapping table size), thus increasing the latency time to find a mapping. This new issue directly affects the memory component within the router, which is currently a bottleneck in the computer system compared with processing capacity. In order to minimize the mapping information managed by a router, hence minimizing the latency time, and in turn, ensuring the highest possible reachability, the LISP Redundancy Protocol (LRP) is proposed, which is inspired by the Cisco's Hot Standby Routing Protocol (HSRP) [19]. This architectural approach essentially permits to configure two routers for mapping purposes, so that before a failure on one of them, the other can supplant it (Master-Slave model). The contribution of the LRP is that it extends HSRP functionalities by creating different logical groups. In this scenario, border routers can be members of different groups, and the "key" difference between LRP and HSRP focuses on the fact that LRP enables a router to be Slave in a group and Master in another (see Figure 4). By implementing this feature all routers can be in forwarding mode, hence overcoming the limitation present in [18], namely to avoid the need of replicating the entire mapping on all the border routers. In case of either a link or border router failure, the last one will interchange his mapping with the border router that is now responsible of handling the traffic in this logical group.

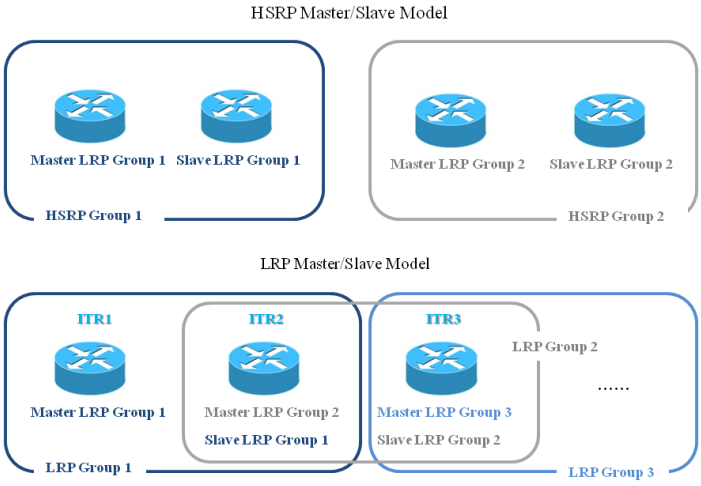

Fig. 3 Master/Slave Model HSRP vs. LRP

In summary, the main features offered by LRP are:

- The xTRs can be clustered into different LRP groups or pairs.
- The Mappings are pushed onto the LRP groups or pairs.
- All the xTRs in the group can carry traffic (active rather than standby).
- No need for data-probes when the xTR does not have a mapping.

In the following subsections we consider different scenarios that may originate reachability and/or reliability problems and require solutions to be managed by the current proposed protocol

4.1 Inter-domain link failure in an ITR

In the following we will discuss and describe the actions that are executed in order to solve the failure of an inter-domain link. In step 1 of Figure 5, the traffic is sent to the edge router (ITR1), which is responsible for encapsulating the traffic and send it through its international links to the destination. When the international link fails (step 2) the ITR1 automatically detects this event and in real time forwards all the incoming traffic to the other ITR belonging to its LRP group (step 3). ITR2 has the correspondent mapping since it shares the LRP Group with ITR1 and now is in charge of encapsulating and sending this traffic to its destination (step 4). On the other hand, by means of the internal routing protocol (running in the AS), the failure of the international link is notified to update the routing tables and hence the traffic is rerouted (step 5). In time, the LISP Control Box (LCB) would be responsible for reconfiguring the mapping of the different ITRs with the purpose of load balancing the outbound traffic (step 6). Finally, the traffic is rerouted according to internal routing policy (step 7). In conclusion, in this scenario our border architecture prevents packet loss and in particular the sending of data-probes.

4.2 ITR failure or unexpected shutdown

If one ITR fails or is shutdown, gigabits of data would be lost. To overcome this situation the LISP control plane must converge to the IGP running in the AS to send back first gigabits of data-probes and then after successful mapping, send data according to the normal procedure. The following describes the step followed by the LRP to prevent any loss of data and to avoid sending data-probes. In step 1 in Figure 6, the traffic is sent to edge router (ITR1), which is responsible for encapsulating the traffic and send it through its international link to destination. When ITR1 fails (step 2) the HSRP that runs between ITR2 and ITR1 converges, and automatically ITR2 assumes the role of Master of the LRP group and forwards all traffic that arrives (step 3) in real time. The convergence of HSRP (HSRP detected in about 3 seconds the router failure) is much faster than any IGP, such as Open Shortest Past First - OSPF. In turn, the internal routing protocol that is running notifies the failure of ITR1 (step 4) to update the routing tables and hence allowing the traffic to be rerouted. On the other hand, the LCB would be responsible for reconfiguring the mapping of the different ITR to balance the outbound traffic load (step 5). Finally, the

traffic is rerouted according to the IGP (step 6). In conclusion, in this scenario our border architecture minimizes the packet loss and in particular the sending of data-probes.

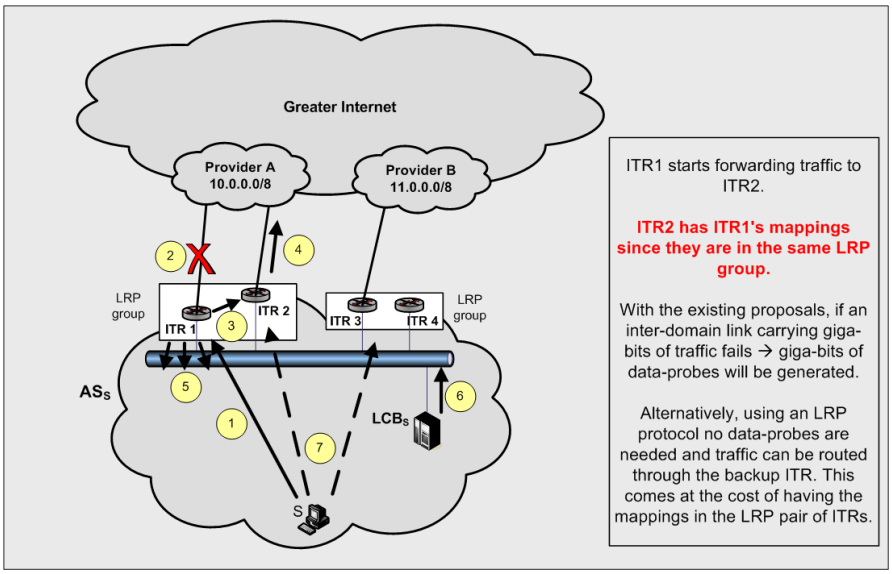

Fig. 4 LRP: Inter-domain link failure

4.3 ITR Mapping Miss

In this scenario, an action of Traffic Engineering or an internal failure (step 1 in Figure 7) makes the traffic to be rerouted. When the packet reaches a border router (step 2) that has no corresponding mapping, this router makes a broadcast to other LRP Groups of the packets that are arriving (step 3), and in turn, the LRP Group sends a Map-Request to the LCB (step 4) that is responsible for handling all mappings within an AS. A LCB (LISP Control Box) is an entity introduced in [18] responsible of all mappings within an AS which might be a standalone device or run as an instance of a PCE. The LRP Group that owns the required mapping, sends it via unicast to the LRP Group responsible for these packages (requester) (step 5), and encapsulates and forwards the traffic. While traffic is derived from the LRP Group mapping holder (step 6), the LCB sends to the LRP Group who sent the Map-Request the mapping necessary to encapsulate the packages (step 7). Finally, the LRP Group can now encapsulate packets and, therefore, makes the package forwarding through the LISP data plane (step 8). In conclusion, in this scenario our border architecture prevents packet loss and in particular the sending of data-probes.

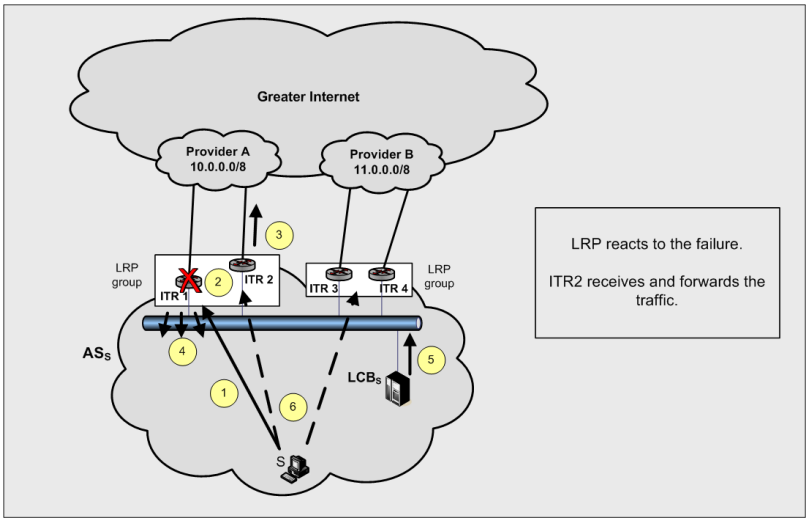

Fig. 5 ITR Failure

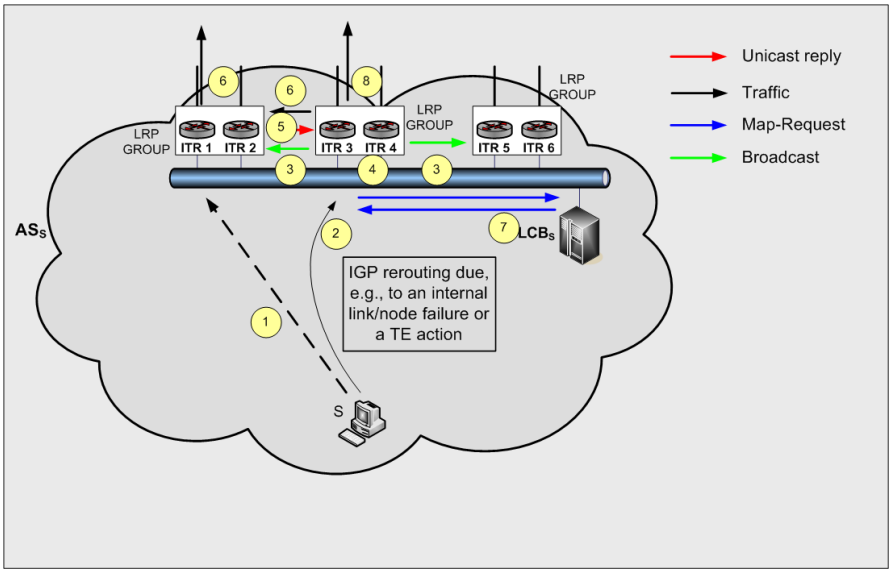

Fig. 6 ITR has no mapping from EID-to-RLOC

5. Future Work.

In our network scenario we assume that LRP nodes are all part of a broadcast domain. In today's ISP core networks this is not always the case, with the introduction of new technologies like MPLS, where edge routers reach each other using LSP (Label switched Path) across an MPLS core. In this network scenario a single broadcast domain does not exist anymore.

Another similar problem can occur when BGP confederations are introduced. To overcome these challenges, extensions to LRP have to been made and configurations in edge routers have to be carefully taken into account.

We expect to present simulations of our proposal with measures of improvement of a response to a EID-RLOC mapping in a future article.

6. Conclusions

In this paper we present the concepts of a new protocol aimed to overcome the scalibility issues that surround the new control plane approach presented in [18], in order to improve reachability and reliability for an Autonomous System. LISP Redundancy Protocol is built based on the HSRP protocol with the addition of new features. In turn, this new proposed architecture deals with two problems. Firstly, the sending of data-probes which involves sending a lot of information without knowing where is to be sent. Secondly, it avoids packet losses in the presence of a failure in the network, what is achieved in all scenarios except when the failure occurs at the edge router of the network (in this case, LRP only minimizes packet loss).

Currently, we are working on an implementation of the LRP Protocol, in order, to obtain measurements on the convergence time in the network, considered as a primary metric for determining the scalability and efficiency of routing schemes.

References

1. E. Nordmark and M. Bagnulo. Shim6: Level 3 multihoming shim protocol for ipv6, RFC 5533 (Draft Standard), June 2009.
2. C. Vogt, "Six/One Router: A Scalable and Backwards Compatible Solution for Provider-Independent Addressing," in MobiArch '08: Proceedings of the 3rd International Workshop on Mobility in the Evolving Internet Architecture, Seattle, WA, USA, Aug. 22, 2008, pp. 13–18.
3. X. Yang and X. sheng Ji, "Host Identity Protocolrealizing the Separation of the Location and Host Identity," in Information and Automation, 2008. ICIA 2008. International Conference on, Changsha, Jun. 20–23, 2008, pp. 749–752.
4. M. Handley, D. Wischik, and M. Bagnulo, "Multipath Transport, Resource Pooling, and implications for Routing," Aug. 1, 2008, RRG. [Online]. Available: http://www.ietf.org/proceedings/72/slides/RRG-2.pdf
5. D. Massey, L. Wang, B. Zhang and L. Zhang, "A Proposal for Scalable Inter-net Routing & Addressing", Internet Draft, draft-wang-ietf-efit-00.txt, February 2007.
6. D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "Locator/ID Separation Pro-tocol (LISP)," Internet Draft, draft-ietf-lisp-09.txt, October 2010.
7. J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an Autonomous System (AS). RFC 1930 (Best Current Practice), March 1996.

8. Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271 (Draft Standard), January 2006.
9. D. Meyer, L. Zhang, and K. Fall, “Report from the IAB Workshop on Routing and Addressing,” IETF, RFC 4984 (Draft Standard), September 2007.
10. G. Houston, “The growth of the BGP table - 1994 to present.” http://bgp.potaroo.net, last visit: January 2011.
11. Tian Bu, Lixin Gao and D. Towsley, “On characterizing BGP routing table growth,” IEEE Global Telecommunications Conference 2002 (GLOBECOM ’02), Taipei, Taiwan, R.O.C., November 2002.
12. R. Oliveira, R. Izhak-Ratzin, Beichuan Zhang and Lixia Zhang, “Measure-ment of highly active prefixes in BGP,” IEEE Global Telecommunications Confe-rence 2005 (GLOBECOM ’05), St. Louis, Missouri, USA, November 2005.
13. Flavel, M. Roughan, N. Bean, and O. Maennel, ”Modeling BGP Table Fluctuations,” 20th International Teletraffic Congress, Ottawa, Canada, June 2007.
14. Quoitin et al., “Evaluating the Benefits of the Locator/Identifier Separa-tion,” Proc. ACM SIGCOMM MobiArch, Kyoto, Japan, Aug. 2007.
15. Farinacci, V. Fuller, and D. Meyer. “Lisp alternative topology (lisp+alt)”. Internet Draft, draft-ietf-lisp-alt-05.txt, October 2010.
16. Lear. “Nerd: A not-so-novel eid to rloc database”. Internet Draft, draft-lear-lisp-nerd-08.txt, March 2010.
17. V. Fuller, D. Farinacci. “LISP Map Server”. Internet Draft, draft-ietf-lisp-ms-06.txt, October 2010.
18. M. Yannuzzi, X. Masip-Bruin, E. Grampin, R. Gagliano, A. Castro, and M. German, “Managing interdomain traffic in Latin America: a new perspective based on LISP [Topics in Network and Service Management],” IEEE Commun. Mag., vol. 47, no. 7, pp. 40–48, Jul. 2009.
19. http://www.cisco.com/en/US/tech/tk648/tk362/tk321/tsd_technology_support_sub-protocol_home.html