

Network Working Group
Internet-Draft
Intended status: Informational
Expires: November 19, 2020

D. Farinacci
lispers.net
May 18, 2020

A Simple LISP NAT-Traversal Implementation
draft-farinacci-lisp-simple-nat-00

Abstract

This informational draft documents the lispers.net LISP NAT-Traversal implementation.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on November 19, 2020.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Definition of Terms	4
3.	Overview	6
4.	Protocol Messages	7
5.	xTR Map-Registering and Map-Server Proxy Map-Replying	10
6.	Packet Flow from ITR-behind-NAT to RTR	11
7.	Packet Flow from Remote ITR to RTR	11
8.	Packet Flow from RTR to ETR-behind-NAT	11
9.	Design Observations	13
10.	Security Considerations	14
11.	IANA Considerations	14
12.	Normative References	14
Appendix A.	Acknowledgments	15
	Author's Address	15

[1.](#) Introduction

This draft documents the LISP messages and protocol procedures for a simple mechanism for the NAT Traversal problem. A subset of message definitions and protocol procedures are taken from [\[I-D.ermagan-lisp-nat-traversal\]](#). This design was first implemented in the lispers.net LISP implementation dating back to January 2014.

The procedures described in this document are performed by LISP compliant [\[I-D.ietf-lisp-rfc6830bis\]](#) [\[I-D.ietf-lisp-rfc6833bis\]](#) xTRs that reside on the private side of one or more NAT devices that connect them to the public side of the network.

The solution is applicable to the following xTR deployments:

- o A physical ITR/ETR device that is directly connected or multiple hops away from a NAT device.
- o A LISP-MN acting as an ITR/ETR device on an cellular service where a mobile provider is providing a NAT function.
- o A logical ITR/ETR that resides in a VM that is behind a NAT device managed by a hypervisor or cloud provider.
- o A logical ITR/ETR that resides in a container where a NAT function is provided by the container service.
- o The above xTR deployments can operate through multiple levels of NATs.

- o The above deployments are also applicable to RTR and PxTR devices that may reside behind NAT devices.
- o The lispers.net lig [[RFC6835](#)] implementation uses the protocol messaging defined in this draft so any system behind a NAT (either running as a LISP xTR or not running LISP at all), can query the mapping system to obtain mappings for network maintenance and troubleshooting.

2. Definition of Terms

Routing Locator (RLOC): an RLOC address is a routable address on the public Internet. It is used by LISP to locate where EIDs are topologically located and appears in the outer header of LISP encapsulated packets. With respect to this design, an RLOC can be a private or public address. Private RLOCs can be registered to the LISP mapping system so they can be used by other LISP xTRs which reside in the same private network. Public RLOCs can be registered to the LISP mapping system and are used by LISP xTRs that are on the public side of the network.

Network Address Translator (NAT): is a router type device that isolates a private network from a public network. The addresses used on the private side of a network are known as private addresses and are not routable on the public side of the network. Therefore, a NAT device must translate private addresses to public addresses. In this document, xTRs that reside on the private side of the network use private RLOCs. These RLOCs must be translated to public addresses so they can be registered in the LISP mapping system.

Private RLOC: is the IP address of the interface of an xTR that faces outbound towards a NAT device. This address is typically translated to a public RLOC address before the packet appears on the public side of the network.

Ephemeral Port: is the UDP source port in a LISP data-plane or control-plane message. This address is typically translated by a NAT device when the packet goes from the private side of the NAT device to the public side of the network.

Global RLOC: is an address that has been translated by a NAT device. The Private RLOC is translated to a Global RLOC and is registered to the mapping system. This RLOC will be the source address in LISP encapsulated packets on the public side of the network.

Translated Port: is the Ephemeral Port that is translated by a NAT device. For an xTR outgoing packet, the source Ephemeral Port is translated to a source Translated Port seen by the public side of the network. For an incoming packet, the NAT device translates the destination Translated Port to the destination Ephemeral Port.

Re-encapsulating Tunnel Router (RTR): is a LISP network element that receives a LISP encapsulated packet, strips the outer header and prepends a new outer header. With respect to this NAT-Traversal design, an ITR (either behind a NAT device or on the public network) encapsulates a packet to the RTR's RLOC address. The RTR

strips this ITR prepended header and then prepends a its own new outer header and sends packet to the RLOC address of an ETR that registered the EID that appears as the destination address from the inner header.

NAT Info Cache: is a data structure managed by an RTR to track xTR hostname, Global RLOC and Translated Port information. The RTR uses this table so it knows what is the destination port to be used for LISP encapsulated packets that much travel through a NAT device.

Address Family Identifier (AFI): a term used to describe an address encoding in a packet [[AFI](#)] and [[RFC1700](#)]. All LISP control messages use AFI encoded addresses. The AFI value is 16-bits in length and precedes all LISP encoded addresses. In this document, the design calls for AFI encodings for IPv4 and IPv6 addresses as well as Distinguished-Name [[I-D.farinacci-lisp-name-encoding](#)] and LCAF [[RFC8060](#)] address formats.

3. Overview

The following sequence of actions describes at a high-level how the lispers.net implementation performs NAT-Traversal and is the basis for a simplified NAT-Traversal protocol design.

1. An xTR sends a Info-Request message to port 4342 to its configured Map-Servers so it can get a list of RTRs to be used for NAT-Traversal.
2. The Map-Servers return an Info-Reply message with the list of RTRs.
3. The xTR then sends an Info-Request message to port 4341 to each RTR.
4. Each RTR caches the translated RLOC address and port in a NAT Info Cache. At this point, the NAT device has created state to allow the RTR to send encapsulated packets from port 4341 to the translated port.
5. The RTR returns an Info-Reply message so the xTR can learn its translated Global RLOC address and Translated Port.
6. The xTR registers its EID-prefixes with an RLOC-set that contains all its global RLOCs as well as the list of RTRs it has learned from Info-Reply messages.
7. The Map-Servers are configured to proxy Map-Reply for these registered EID-prefixes.
8. When a remote ITR sends a Map-Request for an EID that matches one of these EID-prefixes, the Map-Server returns a partial RLOC-set which contain only the list of RTRs. The remote ITR encapsulates packets to the RTRs.
9. When one of the RTRs send a Map-Request for an EID that matches one of these EID-prefixes, the Map-Server returns a partial RLOC-set which contain only the global RLOCs so the RTR can encapsulate packets that will make it through the NAT device to the xTR.
10. The xTR behind a NAT device only stores default map-cache entries with an RLOC-set that contain the list of RTRs the Map-Server supplied it with. The xTR load-splits traffic across the RTRs based on the 5-tuple hash algorithm detailed in [\[I-D.ietf-lisp-rfc6830bis\]](#).

4. Protocol Messages

The lispers.net implementation uses the Info-Request and Info-Reply messages from [[I-D.ermagan-lisp-nat-traversal](#)] as well as the NAT-Traversal LISP Canonical Address Format (LCAF) from [[RFC8060](#)].

```

      0               1               2               3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|Type=7 |0|               Reserved                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               Nonce . . .                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               . . . Nonce                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               Key ID           | Authentication Data Length   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
~               Authentication Data                             ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               TTL                                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Reserved   | EID mask-len |           EID-prefix-AFI       |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               EID-prefix                                         |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               AFI = 0           | <Nothing Follows AFI=0>     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

LISP Info-Request Message Format

The lispers.net implementation will send an Info-Request message to each configured Map-Server. The message is sent to UDP destination port 4342 which is the control-plane port for LISP [[I-D.ietf-lisp-rfc6833bis](#)] from a UDP ephemeral source port. The source address is its Private RLOC. When the xTR is multi-homed to more than one NAT device, it sends the Info-Request on all interfaces facing NAT devices.

A randomized 64-bit nonce is selected for the message and no authentication is used. The EID-prefix AFI is 17 according to the encoding format in [[I-D.farinacci-lisp-name-encoding](#)] and the EID-prefix is the hostname of the xTR encoded as a string null terminated.

An Info-Request is sent out each outgoing interface, with the address of that interface as the Private RLOC, leading to a NAT device. The port pair in the UDP message is the same for each outgoing interface.

When the xTR receives an Info-Reply message from the Map-Server in response to this control-plane Info-Request, it caches a list of RTRs from the Info-Reply. If the list of RTRs are different from each Map-Server, the lists are merged. The xTR stores the merged list as the RLOC-set for 4 default map-cache entries. The map-cache entries have the following EID-prefixes:

```
IPv4 unicast:      0.0.0.0/0
IPv4 multicast:    (0.0.0.0/0, 224.0.0.0/4)
IPv6 unicast:      0::/0
IPv6 multicast:    (0::/0, ff00::/8)
```

Now that the xTR has a list of RTRs, it sends a data-plane Info-Request to each RTR to UDP destination port 4341 from a UDP ephemeral source port. The data-plane Info-Request is sent out each interface just like the control-plane Info-Request was sent for the multi-homed NAT device case.

When Map-Servers and RTRs return an Info-Reply message to xTRs behind NAT devices, the format of the Info-Reply message is the following:

```

      0              1              2              3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|Type=7 |1|              Reserved              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              Nonce . . .              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              . . . Nonce              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              Key ID              | Authentication Data Length |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
~              Authentication Data              ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              TTL              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|  Reserved  | EID mask-len |      EID-prefix-AFI      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              EID-prefix              |
+-->+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| |              AFI = 16387              | Rsvd1      | Flags      |
| +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| |      Type = 7      | Rsvd2      |      4 + n      |
| +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
N |      MS UDP Port Number      |      ETR UDP Port Number      |
A +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
T |      AFI = x      | Global ETR RLOC Address ... |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
L |      AFI = x      |      MS RLOC Address ...      |
C +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
A |      AFI = x      | Private ETR RLOC Address ... |
F +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| |      AFI = x      |      RTR RLOC Address 1 ...      |
| +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| |      AFI = x      |      RTR RLOC Address n ...      |
+-->+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

LISP Info-Reply Message Format

The information returned is the same information that was sent in the Info-Request message except the Info-Reply bit is set (the bit next to Type=7) and the NAT Traversal LCAF encoding is appended.

When a Map-Server returns the Info-Reply, the MS UDP Port Number and ETR UDP Port Number is set to 0. All Address fields are empty by

using AFI equal to 0. Except for the RTR RLOC address fields which the Map-Server is configured to return to xTRs behind NAT devices.

When an RTR returns the Info-Reply, the MS UDP Port Number is set to 0 and the ETR UDP Port Number is set to the UDP source port the RTR received from the Info-Request message. The Global ETR RLOC Address is set to the source address received by the RTR in the Info-Request message. All other address fields are empty by using AFI equal to 0.

5. xTR Map-Registering and Map-Server Proxy Map-Replying

EID-prefixes registered by an xTR behind a NAT include all the global RLOCs and RTR RLOCs it learns. The xTR can use the unicast priority to control ingress packet flow as described in [\[I-D.ietf-lisp-rfc6833bis\]](#). The RTR RLOCs must be registered with a unicast priority of 254 so the Map-Server can identify xTR global RLOCs from RTR RLOCs when proxy Map-Replying.

The Global RLOCs are encoded in a RLOC-record using the AFI-List LCAF encoding [\[RFC8060\]](#). There are two AFI encoded addresses in the list, one being AFI=1 which encodes the IPv4 translated NAT address and other being the Distinguished-Name AFI=17

[\[I-D.farinacci-lisp-name-encoding\]](#) which encodes the hostname of the xTR. When the xTR is multi-homed, the hostname is appended by a unique interface name. For example, for an xTR behind a NAT that has two interfaces facing the same or two different NAT devices, the Distinguished-Name for each RLOC-record could be "dino-xtr-eth0" and "dino-xtr-eth1" for an xTR configured to be named "dino-xtr".

Encoding a Distinguished-Name in an RLOC-record is important so an RTR can use the Global RLOC registered to the mapping system with the translated port stored in its NAT Info Cache. See [Section 8](#) for more details.

When a remote ITR sends a Map-Request for a unicast or multicast EID registered by a xTR behind a NAT, the Map-Server returns a partial RLOC-set that contains all the RTRs (RLOC-records with unicast priority 254) in the proxied Map-Reply message.

When a RTR sends a Map-Request for a unicast or multicast EID registered by a xTR behind a NAT, the Map-Server returns a partial RLOC-set that contains all the Global RLOCs of the xTR behind the NAT in the proxied Map-Reply message.

6. Packet Flow from ITR-behind-NAT to RTR

All packets received by the ITR from the private side of the NAT will use one of the 4 default map-cache entries. There is a unicast and multicast IPv4 default EID-prefix and a unicast and multicast IPv6 default EID-prefix. The RLOC-set is the same for all 4 entries. The RLOC-set contains the globally reachable RLOCs of the RTRs. 5-tuple hashing is used to load-split traffic across the RTRs. RLOC-Probing is used to avoid encapsulating to unreachable RTRs.

7. Packet Flow from Remote ITR to RTR

A remote ITR will get a list of RTRs from the mapping system in a proxy Map-Reply when it sends a Map-Request for a unicast or multicast EID that is registered by an xTR behind a NAT device. The remote ITR will load split traffic across the RTRs from the RLOC-set. Those RTRs can get packets through the NAT devices destined for the xTR behind the NAT since an Info-Request/Info-Reply exchange had already happened between the xTR behind the NAT and the list of RTRs.

There can be a reachability situation where an RTR cannot reach the xTR behind a NAT but a remote ITR may 5-tuple hash to this RTR. Which means packets can travel from the remote ITR to the RTR but then get dropped on the path from the RTR to the xTR behind the NAT. To avoid this situation, the xTR behind the NAT RLOC-probes RTRs and when they become unreachable, they are not included in the xTR registrations.

8. Packet Flow from RTR to ETR-behind-NAT

The RTR will receive a list of Global RLOCs in a proxy Map-Reply from the mapping system for the xTR behind the NAT. The RTR 5-tuple load-splits packets across the RLOC-set of Global RLOCs that can travel through one or more NAT devices along the path to the ETR behind the NAT device.

When the RTR selects a Global RLOC to encapsulate to it must select the correct Translated Port for the UDP destination port in the encapsulation header. The RTR needs to use the same Translated Address and Translated Port pair a NAT device used to translate the Info-Request message otherwise the encapsulated packet will be dropped. The NAT Info Cache contains an entry for every hostname (and optionally appended interface name), translated address and port cached when processing Info-Request messages. The RTR obtains the correct Translated Port from the NAT Info Cache by using the Global RLOC and RLOC-record hostname from the registered RLOC-set.

The RTR can test reachability for xTRs behind NATs by encapsulating RLOC-Probe requests in data packets where the UDP source port is set to 4341 and the UDP destination port is set to the Translated Port. The outer header destination address is the Global RLOC for the xTR.

9. Design Observations

The following benefits and observations can be attributed to this design:

- o An ITR behind a NAT virtually runs no control-plane and a very simple data-plane. All it does is RLOC-probe the RTRs in the common RLOC-set for each default map-cache entry. And maintains a very small map-cache of 4 entries per instance-ID it supports.
- o An xTR behind a NAT can tell if another xTR is behind the same set of NAT devices and use Private RLOCs to reach each other on a short-cut path. It does this by comparing the Global RLOC returned from RTRs in Info-Reply messages.
- o An xTR behind a NAT is free to send a Map-Request to the mapping system for any EID to test to see if there is a direct path to the LISP site versus potentially using a sub-optimal path through an RTR. This happens when xTRs exist that are not behind NAT devices where their RLOCs are global RLOCs.
- o By sending Info-Requests to Map-Servers, an xTR behind a NAT can tell if they are reachable and if those Map-Servers also act as Map-Resolvers, the xTR behind the NAT can avoid sending Map-Requests to unreachable Map-Resolvers.
- o Enhanced data-plane security can be used via LISP-Crypto mechanisms detailed in [[RFC8061](#)] using this NAT-Traversal design so both unicast and multicast traffic are encrypted.
- o This design allows for the minimum amount of NAT device state since only RTRs are encapsulating to ETRs behind NAT devices. Therefore, the number of ITRs sending packets to EIDs behind NAT devices is aggregated out for scale. Scale is also achieved when xTRs behind NATs roam and RLOC-set changes need to update only RTR map-caches.
- o The protocol procedures in this document can be used when a LISP site has multiple xTRs each connected via separate NAT devices to the public network. Each xTR registers their Global RLOCs and RTRs with merge semantics to the mapping system so remote ITRs can load-split traffic across a merged RTR set as well as RTRs across each xTR behind different NAT devices.

10. Security Considerations

There are no security considerations at this time.

However, the lispers.net implementation can be enhanced easily to allow the same authentication xTRs use with Map-Register messages for Info-Request messages when they send to Map-Servers.

For authentication of Info-Requests to RTRs, more work is required to maintain key management associations between xTR behind NATs and RTRs. It is not trivial to make this happen with a dynamic list of RTRs that can change as the xTR behind a NAT roams to other parts of the network and desire shorter paths to RTRs.

11. IANA Considerations

The code-point values in this specification are already allocated in [AFI] or [RFC8060].

12. Normative References

- [AFI] "Address Family Identifier (AFIs)", ADDRESS FAMILY NUMBERS <http://www.iana.org/numbers.html>, February 2007.
- [I-D.ermagan-lisp-nat-traversal]
Ermagan, V., Farinacci, D., Lewis, D., Maino, F., Portoles-Comeras, M., Skriver, J., and C. White, "NAT traversal for LISP", [draft-ermagan-lisp-nat-traversal-16](#) (work in progress), June 2019.
- [I-D.farinacci-lisp-name-encoding]
Farinacci, D., "LISP Distinguished Name Encoding", [draft-farinacci-lisp-name-encoding-09](#) (work in progress), March 2020.
- [I-D.ietf-lisp-rfc6830bis]
Farinacci, D., Fuller, V., Meyer, D., Lewis, D., and A. Cabellos-Aparicio, "The Locator/ID Separation Protocol (LISP)", [draft-ietf-lisp-rfc6830bis-32](#) (work in progress), March 2020.
- [I-D.ietf-lisp-rfc6833bis]
Farinacci, D., Maino, F., Fuller, V., and A. Cabellos-Aparicio, "Locator/ID Separation Protocol (LISP) Control-Plane", [draft-ietf-lisp-rfc6833bis-27](#) (work in progress), January 2020.

- [RFC1700] Reynolds, J. and J. Postel, "Assigned Numbers", [RFC 1700](#), DOI 10.17487/RFC1700, October 1994, <<https://www.rfc-editor.org/info/rfc1700>>.
- [RFC6835] Farinacci, D. and D. Meyer, "The Locator/ID Separation Protocol Internet Groper (LIG)", [RFC 6835](#), DOI 10.17487/RFC6835, January 2013, <<https://www.rfc-editor.org/info/rfc6835>>.
- [RFC8060] Farinacci, D., Meyer, D., and J. Snijders, "LISP Canonical Address Format (LCAF)", [RFC 8060](#), DOI 10.17487/RFC8060, February 2017, <<https://www.rfc-editor.org/info/rfc8060>>.
- [RFC8061] Farinacci, D. and B. Weis, "Locator/ID Separation Protocol (LISP) Data-Plane Confidentiality", [RFC 8061](#), DOI 10.17487/RFC8061, February 2017, <<https://www.rfc-editor.org/info/rfc8061>>.

[Appendix A](#). Acknowledgments

The author would like to thank the authors of the LISP NAT-Traversal specification [[I-D.ermagan-lisp-nat-traversal](#)] for their initial ideas and prototyping to allow a simpler form of NAT-Traversal to be designed.

Author's Address

Dino Farinacci
lispers.net
San Jose, CA
USA

Email: farinacci@gmail.com

