Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: January 14, 2021

F. Maino, Ed.
Cisco
J. Lemon
Broadcom
P. Agarwal
Innovium
D. Lewis
M. Smith
Cisco

July 13, 2020

LISP Generic Protocol Extension
draft-ietf-lisp-gpe-18

Abstract

This document describes extensions to the Locator/ID Separation Protocol (LISP) Data-Plane, via changes to the LISP header, to support multi-protocol encapsulation and allow to introduce new protocol capabilities.

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 January 14, 2021.

Table of Contents

1. Introduction

The LISP Data-Plane is defined in [I-D.ietf-lisp-rfc6830bis]. It specifies an encapsulation format that carries IPv4 or IPv6 packets (henceforth jointly referred to as IP) in a LISP header and outer UDP/IP transport.

The LISP Data-Plane header does not specify the protocol being encapsulated and therefore is currently limited to encapsulating only IP packet payloads. Other protocols, most notably Virtual eXtensible Local Area Network (VXLAN) [RFC7348] (which defines a similar header format to LISP), are used to encapsulate Layer-2 (L2) protocols such as Ethernet.

This document defines an extension for the LISP header, as defined in [I-D.ietf-lisp-rfc6830bis], to indicate the inner protocol, enabling the encapsulation of Ethernet, IP or any other desired protocol all the while ensuring compatibility with existing LISP deployments.

A flag in the LISP header, called the P-bit, is used to signal the presence of the 8-bit Next Protocol field. The Next Protocol field, when present, uses 8 bits of the field that was allocated to the echo-noncing and map-versioning features in [I-D.ietf-lisp-rfc6830bis]. Those two features are no longer available when the P-bit is used. However, appropriate LISP-GPE (LISP Generic Protocol Extension) shim headers can be defined to specify capabilities that are equivalent to echo-noncing and/or map-versioning.

Since all of the reserved bits of the LISP Data-Plane header have been allocated, LISP-GPE can also be used to extend the LISP Data-Plane header by defining Next Protocol "shim" headers that implements new data plane functions not supported in the LISP header. For example, the use of the Group-Based Policy (GBP) header [I-D.lemon-vxlan-lisp-gpe-gbp] or of the In-situ Operations, Administration, and Maintenance (IOAM) header [I-D.brockners-ippm-ioam-vxlan-gpe] with LISP-GPE, can be considered an extension to add support in the Data-Plane for Group-Based Policy functionalities or IOAM metadata.

1.1. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. Definition of Terms

This document uses terms already defined in [I-D.ietf-lisp-rfc6830bis].

2. LISP Header Without Protocol Extensions

As described in Section 1, the LISP header has no protocol identifier that indicates the type of payload being carried. Because of this, LISP is limited to carrying IP payloads.

The LISP header [I-D.ietf-lisp-rfc6830bis] contains a series of flags (some defined, some reserved), a Nonce/Map-version field and an instance ID/Locator-status-bit field. The flags provide flexibility to define how the various fields are encoded. Notably, Flag bit 5 is the last reserved bit in the LISP header.

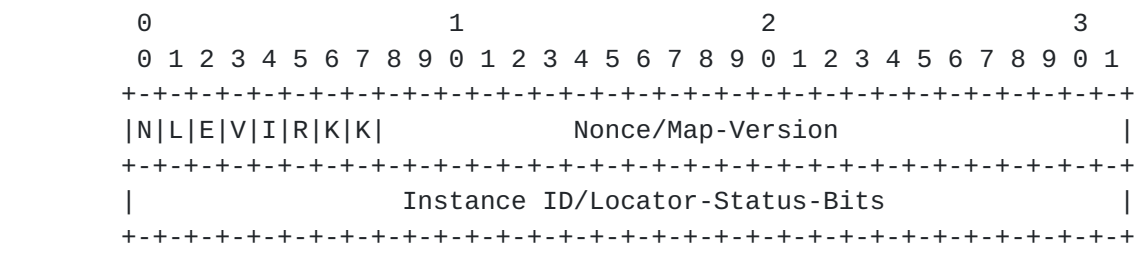

Figure 1: LISP Header

3. Generic Protocol Extension for LISP (LISP-GPE)

This document defines two changes to the LISP header in order to support multi-protocol encapsulation: the introduction of the P-bit and the definition of a Next Protocol field. This document specifies the protocol behavior when the P-bit is set to 1, no changes are introduced when the P-bit is set to 0. The LISP-GPE header is shown in Figure 2 and described below.

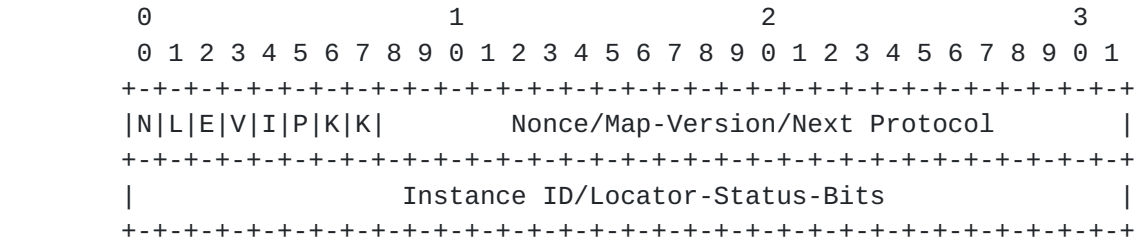

Figure 2: LISP-GPE Header

P-Bit: Flag bit 5 is defined as the Next Protocol bit. The P-bit is set to 1 to indicate the presence of the 8 bit Next Protocol field.

If the P-bit is clear (0) the LISP header is bit-by-bit equivalent to the definition in [I-D.ietf-lisp-rfc6830bis].

When the P-bit is set to 1, bits N, E, V, and bits 8-23 of the 'Nonce/Map-Version/Next Protocol' field MUST be set to zero on transmission and MUST be ignored on receipt. Features equivalent to those that were implemented with bits N,E and V in [I-D.ietf-lisp-rfc6830bis], such as echo-noncing and map-versioning, can be implemented by defining appropriate LISP-GPE shim headers.

When the P-bit is set to 1, the LISP-GPE header is encoded as:

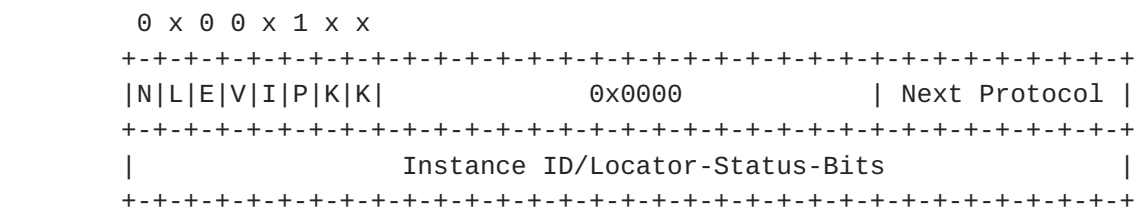

Figure 3: LISP-GPE with P-bit set to 1

Next Protocol: When the P-bit is set to 1, the lower 8 bits of the first 32-bit word are used to carry a Next Protocol. This Next Protocol field contains the protocol of the encapsulated payload packet.

This document defines the following Next Protocol values:

0x00 : Reserved

0x01 : IPv4

0x02 : IPv6

0x03 : Ethernet

0x04 : Network Service Header (NSH) [RFC8300]

0x05 to 0x7D: Unassigned

0x7E, 0x7F: Experimentation and testing

0x80 to 0xFD: Unassigned (shim headers)

0xFE, 0xFF: Experimentation and testing (shim headers)

The values are tracked in the IANA LISP-GPE Next Protocol Registry as described in Section 6.1.

Next protocol values 0x7E, 0x7F and 0xFE, 0xFF are assigned for experimentation and testing as per [RFC3692].

Next protocol values from Ox80 to 0xFD are assigned to protocols encoded as generic "shim" headers. All shim protocols MUST use the header structure in Figure 4, which includes a Next Protocol field. When shim headers are used with other protocols identified by next

protocol values from 0x00 to 0x7F, all the shim headers MUST come first.

Shim headers can be used to incrementally deploy new GPE features, keeping the processing of shim headers known to a given xTR implementation in the 'fast' path (typically an ASIC), while punting the processing of the remaining new GPE features to the 'slow' path.

Shim protocols MUST have the first 32 bits defined as:

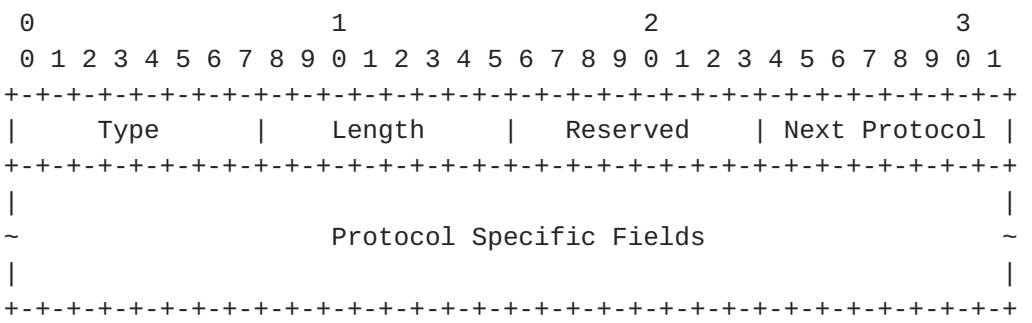

Figure 4: Shim Header

Where:

Type: This field identifies the different messages of this protocol.

Length: The length, in 4-octet units, of this protocol message not including the first 4 octets.

Reserved: The use of this field is reserved to the protocol defined in this message.

Next Protocol Field: The next protocol field contains the protocol of the encapsulated payload. The values are tracked in the IANA LISP-GPE Next Protocol Registry as described in Section 6.1.

4. Implementation and Deployment Considerations

4.1. Applicability Statement

LISP-GPE conforms, as an UDP-based encapsulation protocol, to the UDP usage guidelines as specified in [RFC8085]. The applicability of these guidelines are dependent on the underlay IP network and the nature of the encapsulated payload.

[RFC8085] outlines two applicability scenarios for UDP applications, 1) general Internet and 2) controlled environment. The controlled environment means a single administrative domain or adjacent set of

cooperating domains. A network in a controlled environment can be managed to operate under certain conditions whereas in general Internet this cannot be done. Hence requirements for a tunnel protocol operating under a controlled environment can be less restrictive than the requirements of general internet.

LISP-GPE scope of applicability is the same set of use cases covered by[I-D.ietf-lisp-rfc6830bis] for the LISP dataplane protocol. The common property of these use cases is a large set of cooperating entities seeking to communicate over the public Internet or other large underlay IP infrastructures, while keeping the addressing and topology of the cooperating entities separate from the underlay and Internet topology, routing, and addressing.

LISP-GPE is meant to be deployed in network environments operated by a single operator or adjacent set of cooperating network operators that fits with the definition of controlled environments in [RFC8085].

For the purpose of this document, a traffic-managed controlled environment (TMCE), outlined in [RFC8086], is defined as an IP network that is traffic-engineered and/or otherwise managed (e.g., via use of traffic rate limiters) to avoid congestion. Significant portions of text in this Section are based on [RFC8086].

It is the responsibility of the network operators to ensure that the guidelines/requirements in this section are followed as applicable to their LISP-GPE deployments

4.2. Congestion Control Functionality

LISP-GPE does not natively provide congestion control functionality and relies on the payload protocol traffic for congestion control. As such LISP-GPE MUST be used with congestion controlled traffic or within a network that is traffic managed to avoid congestion (TMCE). An operator of a traffic managed network (TMCE) may avoid congestion by careful provisioning of their networks, rate-limiting of user data traffic and traffic engineering according to path capacity.

Keeping in mind the reccomendation above, new encapsulated payloads, when registered with LISP-GPE, should be designed for explicit congestion signals to propagate consistently from lower layer protocols into IP. Then the IP internetwork layer can act as a portability layer to carry congestion notification from non-IP-aware congested nodes up to the transport layer (L4). Such new encapsulated payloads, when registered with LISP-GPE, MUST be accompanied by a set of guidelines derived from [RFC6040] and SHOULD

follow the guidelines defined in [I-D.ietf-tsvwg-ecn-encap-guidelines].

4.3. UDP Checksum

For IP payloads, section 5.3 of [I-D.ietf-lisp-rfc6830bis] specifies how to handle UDP Checksums encouraging implementors to consider UDP checksum usage guidelines in section 3.4 of [RFC8085] when it is desirable to protect UDP and LISP headers against corruption.

In order to provide integrity of LISP-GPE headers, options and payload, for example to avoid mis-delivery of payload to different tenant systems in case of data corruption, outer UDP checksum SHOULD be used with LISP-GPE when transported over IPv4. The UDP checksum provides a statistical guarantee that a payload was not corrupted in transit. These integrity checks are not strong from a coding or cryptographic perspective and are not designed to detect physical-layer errors or malicious modification of the datagram (see Section 3.4 of [RFC8085]). In deployments where such a risk exists, an operator SHOULD use additional data integrity mechanisms such as offered by IPSec.

An operator MAY choose to disable UDP checksum and use zero checksum if LISP-GPE packet integrity is provided by other data integrity mechanisms such as IPsec or additional checksums or if one of the conditions in Section 4.3.1 a, b, c are met.

4.3.1. UDP Zero Checksum Handling with IPv6

By default, UDP checksum MUST be used when LISP-GPE is transported over IPv6. A tunnel endpoint MAY be configured for use with zero UDP checksum if additional requirements described in this section are met.

When LISP-GPE is used over IPv6, UDP checksum is used to protect IPv6 headers, UDP headers and LISP-GPE headers and payload from potential data corruption. As such by default LISP-GPE MUST use UDP checksum when transported over IPv6. An operator MAY choose to configure to operate with zero UDP checksum if operating in a traffic managed controlled environment as stated in Section 4.1 if one of the following conditions are met:

a. It is known that the packet corruption is exceptionally unlikely (perhaps based on knowledge of equipment types in their underlay network) and the operator is willing to take a risk of undetected packet corruption

b. It is judged through observational measurements (perhaps through historic or current traffic flows that use non zero checksum) that the level of packet corruption is tolerably low and where the operator is willing to take the risk of undetected corruption

c. LISP-GPE payload is carrying applications that are tolerant of misdelivered or corrupted packets (perhaps through higher layer checksum validation and/or reliability through retransmission)

In addition LISP-GPE tunnel implementations using Zero UDP checksum MUST meet the following requirements:

1. Use of UDP checksum over IPv6 MUST be the default configuration for all LISP-GPE tunnels

2. If LISP-GPE is used with zero UDP checksum over IPv6 then such xTR implementation MUST meet all the requirements specified in section 4 of [RFC6936] and requirements 1 as specified in section 5 of [RFC6936]

3. The ETR that decapsulates the packet SHOULD check the source and destination IPv6 addresses are valid for the LISP-GPE tunnel that is configured to receive Zero UDP checksum and discard other packets for which such check fails

4. The ITR that encapsulates the packet MAY use different IPv6 source addresses for each LISP-GPE tunnel that uses Zero UDP checksum mode in order to strengthen the decapsulator's check of the IPv6 source address (i.e the same IPv6 source address is not to be used with more than one IPv6 destination address, irrespective of whether that destination address is a unicast or multicast address). When this is not possible, it is RECOMMENDED to use each source address for as few LISP-GPE tunnels that use zero UDP checksum as is feasible

5. Measures SHOULD be taken to prevent LISP-GPE traffic over IPv6 with zero UDP checksum from escaping into the general Internet. Examples of such measures include employing packet filters at the PETR and/or keeping logical or physical separation of LISP network from networks carrying General Internet

The above requirements do not change either the requirements specified in [RFC2460] as modified by [RFC6935] or the requirements specified in [RFC6936].

The requirement to check the source IPv6 address in addition to the destination IPv6 address, plus the recommendation against reuse of source IPv6 addresses among LISP-GPE tunnels collectively provide

some mitigation for the absence of UDP checksum coverage of the IPv6 header. A traffic-managed controlled environment that satisfies at least one of three conditions listed at the beginning of this section provides additional assurance.

4.4. DSCP, ECN, TTL, and 802.1Q

When encapsulating IP (including over Ethernet) packets [RFC2983] provides guidance for mapping DSCP between inner and outer IP headers. The Pipe model typically fits better Network virtualization. The DSCP value on the tunnel header is set based on a policy (which may be a fixed value, one based on the inner traffic class, or some other mechanism for grouping traffic). Some aspects of the Uniform model (which treats the inner and outer DSCP value as a single field by copying on ingress and egress) may also apply, such as the ability to remark the inner header on tunnel egress based on transit marking. However, the Uniform model is not conceptually consistent with network virtualization, which seeks to provide strong isolation between encapsulated traffic and the physical network.

[RFC6040] describes the mechanism for exposing ECN capabilities on IP tunnels and propagating congestion markers to the inner packets. This behavior MUST be followed for IP packets encapsulated in LISP-GPE.

Though Uniform or Pipe models could be used for TTL (or Hop Limit in case of IPv6) handling when tunneling IP packets, Pipe model is more aligned with network virtualization. [RFC2003] provides guidance on handling TTL between inner IP header and outer IP tunnels; this model is more aligned with the Pipe model and is recommended for use with LISP-GPE for network virtualization applications.

When a LISP-GPE router performs Ethernet encapsulation, the inner 802.1Q [IEEE.802.1Q_2014] 3-bit priority code point (PCP) field MAY be mapped from the encapsulated frame to the DSCP codepoint of the DS field defined in [RFC2474].

When a LISP-GPE router performs Ethernet encapsulation, the inner header 802.1Q [IEEE.802.1Q_2014] VLAN Identifier (VID) MAY be mapped to, or used to determine the LISP Instance IDentifier (IID) field.

5. Backward Compatibility

LISP-GPE uses the same UDP destination port (4341) allocated to LISP.

When encapsulating IP packets to a non LISP-GPE capable router the P-bit MUST be set to 0. That is, the encapsulation format defined in this document MUST NOT be sent to a router that has not indicated

that it supports this specification because such a router would ignore the P-bit (as described in [I-D.ietf-lisp-rfc6830bis]) and so would misinterpret the other LISP header fields possibly causing significant errors.

5.1. Detection of ETR Capabilities

The discovery of xTR capabilities to support LISP-GPE is out of the scope of this document. Given that the applicability domain of LISP-GPE is a traffic-managed controlled environment, ITR/ETR (xTR) configuration mechanisms may be used for this purpose.

6. IANA Considerations

6.1. LISP-GPE Next Protocol Registry

IANA is requested to set up a registry of LISP-GPE "Next Protocol". These are 8-bit values. Next Protocol values in the table below are defined in this document. New values are assigned under the Specification Required policy [RFC8126]. The protocols that are being assigned values do not themselves need to be IETF standards track protocols.

Next Protocol	Description	Reference
0x0	Reserved	This Document
0x1	IPv4	This Document
0x2	IPv6	This Document
0x3	Ethernet	This Document
0x4	NSH	This Document
0x05..0x7D	Unassigned	
0x7E..0x7F	Experimentation and testing	This Document
0x80..0xFD	Unassigned (shim headers)	
0x8E..0x8F	Experimentation and testing (shim headers)	This Document

7. Security Considerations

LISP-GPE security considerations are similar to the LISP security considerations and mitigation techniques documented in [RFC7835].

LISP-GPE, as many encapsulations that use optional extensions, is subject to on-path adversaries that by manipulating the P-Bit and the packet itself can remove part of the payload or claim to encapsulate any protocol payload type. Typical integrity protection mechanisms (such as IPsec) SHOULD be used in combination with LISP-GPE by those protocol extensions that want to protect from on-path attackers.

With LISP-GPE, issues such as data-plane spoofing, flooding, and traffic redirection may depend on the particular protocol payload encapsulated.

8. Acknowledgements and Contributors

A special thank you goes to Dino Farinacci for his guidance and detailed review. Thanks to Tom Herbert for the suggestion to assign codepoints for experimentations and testing.

This Working Group (WG) document originated as draft-lewis-lisp-gpe; the following are its coauthors and contributors along with their respective affiliations at the time of WG adoption. The editor of this document would like to thank and recognize them and their contributions. These coauthors and contributors provided invaluable concepts and content for this document's creation.

- Darrel Lewis, Cisco Systems, Inc.
- Fabio Maino, Cisco Systems, Inc.
- Paul Quinn, Cisco Systems, Inc.
- Michael Smith, Cisco Systems, Inc.
- Navindra Yadav, Cisco Systems, Inc.
- Larry Kreeger
- Jennifer Lemon, Broadcom
- Puneet Agarwal, Innovium

9. References

9.1. Normative References

[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.

[IEEE.802.1Q_2014]
IEEE, "IEEE Standard for Local and metropolitan area networks--Bridges and Bridged Networks", IEEE 802.1Q-2014, DOI 10.1109/ieeestd.2014.6991462, December 2014, <http://ieeexplore.ieee.org/servlet/opac?punumber=6991460>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, DOI 10.17487/RFC6040, November 2010, <https://www.rfc-editor.org/info/rfc6040>.

9.2. Informative References

[I-D.brockners-ippm-ioam-vxlan-gpe]
Brockners, F., Bhandari, S., Govindan, V., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Kfir, A., Gafni, B., Lapukhov, P., and M. Spiegel, "VXLAN-GPE Encapsulation for In-situ OAM Data", draft-brockners-ippm-ioam-vxlan-gpe-03 (work in progress), November 2019.

[I-D.ietf-tsvwg-ecn-encap-guidelines]
Briscoe, B., Kaippallimalil, J., and P. Thaler, "Guidelines for Adding Congestion Notification to Protocols that Encapsulate IP", draft-ietf-tsvwg-ecn-encap-guidelines-13 (work in progress), May 2019.

[I-D.lemon-vxlan-lisp-gpe-gbp]
Lemon, J., Maino, F., Smith, M., and A. Isaac, "Group Policy Encoding with VXLAN-GPE and LISP-GPE", draft-lemon-vxlan-lisp-gpe-gbp-02 (work in progress), April 2019.

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <https://www.rfc-editor.org/info/rfc2003>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <https://www.rfc-editor.org/info/rfc2460>.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <https://www.rfc-editor.org/info/rfc2474>.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, DOI 10.17487/RFC2983, October 2000, <https://www.rfc-editor.org/info/rfc2983>.

[RFC3692] Narten, T., "Assigning Experimental and Testing Numbers Considered Useful", BCP 82, RFC 3692, DOI 10.17487/RFC3692, January 2004, <https://www.rfc-editor.org/info/rfc3692>.

[RFC6935] Eubanks, M., Chimento, P., and M. Westerlund, "IPv6 and UDP Checksums for Tunneled Packets", RFC 6935, DOI 10.17487/RFC6935, April 2013, <https://www.rfc-editor.org/info/rfc6935>.

[RFC6936] Fairhurst, G. and M. Westerlund, "Applicability Statement for the Use of IPv6 UDP Datagrams with Zero Checksums", RFC 6936, DOI 10.17487/RFC6936, April 2013, <https://www.rfc-editor.org/info/rfc6936>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

[RFC7835] Saucez, D., Iannone, L., and O. Bonaventure, "Locator/ID Separation Protocol (LISP) Threat Analysis", RFC 7835, DOI 10.17487/RFC7835, April 2016, <https://www.rfc-editor.org/info/rfc7835>.

[RFC8085] Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", BCP 145, RFC 8085, DOI 10.17487/RFC8085, March 2017, <https://www.rfc-editor.org/info/rfc8085>.

[RFC8086] Yong, L., Ed., Crabbe, E., Xu, X., and T. Herbert, "GRE-in-UDP Encapsulation", RFC 8086, DOI 10.17487/RFC8086, March 2017, <https://www.rfc-editor.org/info/rfc8086>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8300] Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed., "Network Service Header (NSH)", RFC 8300, DOI 10.17487/RFC8300, January 2018, <https://www.rfc-editor.org/info/rfc8300>.

Authors' Addresses

Fabio Maino (editor)
Cisco Systems
San Jose, CA 95134
USA

Email: fmaino@cisco.com

Jennifer Lemon
Broadcom
270 Innovation Drive
San Jose, CA 95134
USA

Email: jennifer.lemon@broadcom.com

Puneet Agarwal
Innovium
USA

Email: puneet@acm.org

Darrel Lewis
Cisco Systems
San Jose, CA 95134
USA

Email: darlewis@cisco.com

Michael Smith
Cisco Systems
San Jose, CA 95134
USA

Email: michsmit@cisco.com