L2TPEXT Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 10, 2015

M. Konstantynowicz, Ed.
G. Heron, Ed.
Cisco Systems
R. Schatzmayr
Deutsche Telekom AG
W. Henderickx
Alcatel-Lucent, Inc.
March 9, 2015

Keyed IPv6 Tunnel
draft-ietf-l2tpext-keyed-ipv6-tunnel-04

Abstract

This document describes a simple L2 Ethernet over IPv6 tunnel encapsulation with mandatory 64-bit cookie for connecting L2 Ethernet attachment circuits identified by IPv6 addresses. The encapsulation is based on L2TPv3 over IP.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119 [RFC2119].

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 http://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 September 10, 2015.

Copyright Notice

Table of Contents

1. Introduction

L2TPv3, as defined in [RFC3931], provides a dynamic mechanism for tunneling Layer 2 (L2) "circuits" across a packet-oriented data network (e.g., over IP), with multiple attachment circuits multiplexed over a single pair of IP address endpoints (i.e. a tunnel) using the L2TPv3 session ID as a circuit discriminator.

Implementing L2TPv3 over IPv6 [RFC2460] provides the opportunity to utilize unique IPv6 addresses to identify Ethernet attachment circuits directly, leveraging the key property that IPv6 offers, a vast number of unique IP addresses. In this case, processing of the L2TPv3 Session ID may be bypassed upon receipt as each tunnel has one and only one associated session. This local optimization does not hinder the ability to continue supporting the multiplexing of circuits via the Session ID on the same router for other L2TPv3 tunnels.

2. Static 1:1 Mapping Without a Control Plane

Static local configuration creates a one-to-one mapping between the access-side L2 attachment circuit and the IP address used in the network-side IPv6 encapsulation. The L2TPv3 Control Plane defined in RFC3931 is not used.

The IPv6 L2TPv3 tunnel encapsulating device uniquely identifies each Ethernet L2 attachment connection by a port ID or a combination of port ID and VLAN ID(s) on the access side, and by a local IPv6 address on the network side The local IPv6 address also identifies the tunnel endpoint. The local IPv6 addresses identifying L2TPv3 tunnels SHOULD NOT be assigned from connected IPv6 subnets facing towards remote tunnel endponts - since that approach would result in an IPv6 Neighbor Discovery cache entry per tunnel on the next hop router towards the remote tunnel endpoint. It is RECOMMENDED that local IPv6 addresses identifying L2TPv3 tunnels are assigned from dedicated subnets used only for such tunnel endpoints.

Certain deployment scenarios may require using a single IPv6 address (typically a globally routable unicast or anycast address assigned to a virtual interface) to identify a tunnel endpoint for multiple IPv6 L2TPv3 tunnels. For such cases the tunnel encapsulating device identifies each tunnel by a unique combination of local and remote IPv6 addresses.

As mentioned above Session ID processing is not required as each keyed IPv6 tunnel has one and only one associated session. However for compatibility with existing RFC3931 implementations, the packets need to be sent with Session ID. Routers implementing L2TPv3 according to RFC3931 can be configured with multiple L2TPv3 tunnels, with one session per tunnel, to interoperate with routers implementing the keyed IPv6 tunnel as specified by this document. Note that as Session ID processing is not enabled for keyed IPv6 tunnels that there can only be a single keyed IPv6 tunnel between two IPv6 addresses.

Note that a previous IETF draft [I.D.ietf-pppext-l2tphc] introduces the concept of an L2TP tunnel carrying a single session and hence not requiring session ID processing.

3. 64-bit Cookie

In line with RFC3931, the 64-bit cookie is used for an additional tunnel endpoint context check. All packets MUST carry the 64-bit L2TPv3 cookie field. The cookie MUST be 64 bits long in order to provide sufficient protection against spoofing and brute force blind insertion attacks. The cookie values SHOULD be randomly selected.

In the absence of the L2TPv3 Control Plane, the L2TPv3 encapsulating router MUST be provided with local configuration of the 64-bit cookie for each local and remote IPv6 endpoint. Note that cookies are asymmetric, so local and remote endpoints may send different cookie values, and in fact SHOULD do so. The value of the cookie MUST be able to be changed at any time in a manner that does not drop any legitimate tunneled packets - i.e. the receiver MUST be configurable to accept two discrete cookies for a single tunnel simultaneously. This enables the receiver to hold both the 'old' and 'new' cookie values during a change of cookie value. Cookie values SHOULD be changed periodically.

Note that mandating a 64-bit cookie is a change from the optional variable-length cookie of RFC3931, and that this requirement constrains interoperability with existing RFC3931 implementations to those supporting a 64-bit cookie.

4. Encapsulation

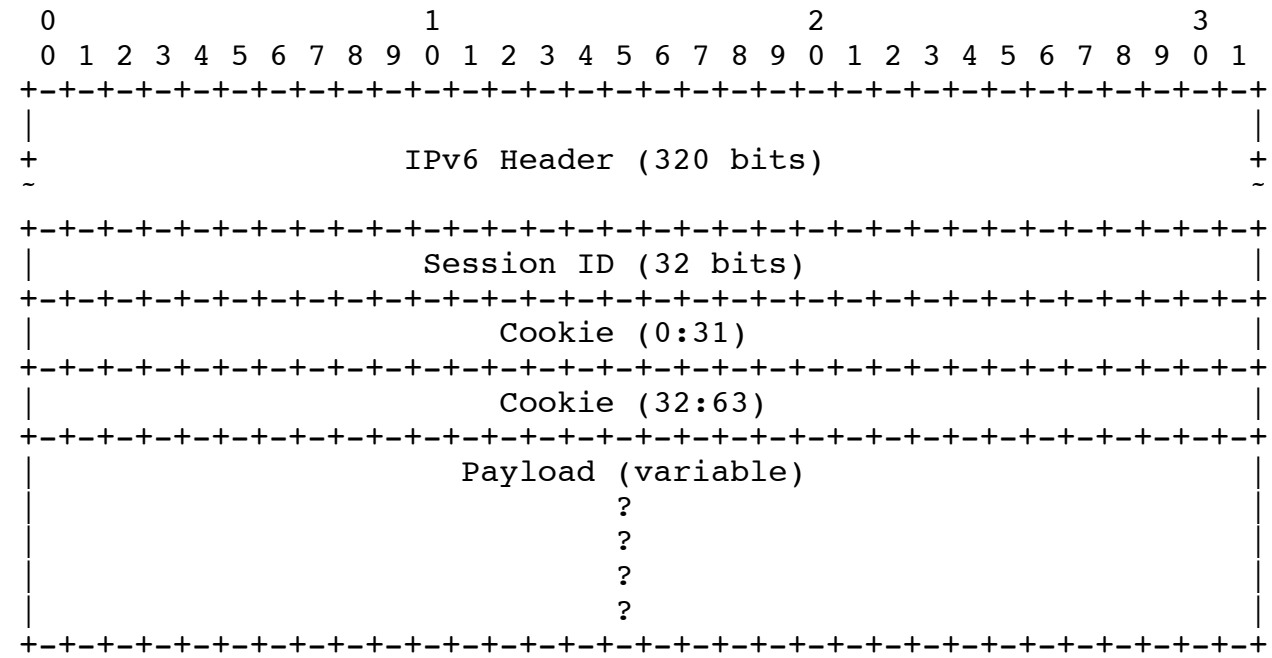
The ingress router encapsulates the entire Ethernet frame, without the preamble and frame check sequence (FCS) in L2TPv3 as per RFC4719 [RFC4719]. The L2TPv3 packet is encapsulated directly over IPv6 (i.e. no UDP header is carried).

The ingress router MAY retain the FCS as per section 4.7 of [RFC4720]. Support for retaining the FCS and for receiving packets with a retained FCS is OPTIONAL, and if present MUST be configurable. In the absence of the L2TPv3 control plane such configuration MUST be consistent for the two endpoints of any given tunnel - i.e. if one router is configured to retain the FCS then the other router MUST be configured to receive packets with the retained FCS. Any router configured to retain FCS for a tunnel MUST retain FCS for all frames sent over that tunnel. All routers implementing this specification MUST support the ability to send frames without retaining the FCS and to receive such frames.

Any service-delimiting IEEE 802.1Q [IEEE802.1Q] or IEEE 802.1ad [IEEE802.1ad] VLAN IDs - S-tag, C-tag or tuple (S-tag, C-tag) - are treated with local significance within the Ethernet L2 port and are MUST NOT be forwarded over the IPv6 L2TPv3 tunnel.

Note that the same approach may be used to transport protocols other than Ethernet.

The full encapsulation is as follows:



The combined IPv6 and Keyed IP Tunnel header contains the following fields:

o IPv6 Header. Note that:

  * The traffic class may be set by the ingress router to ensure correct PHB treatment by transit routers between the ingress and egress, and correct QoS disposition at the egress router.

  * The flow label, as defined in [RFC6437] may be set by the ingress router to indicate a flow of packets from the client which may not be reordered by the network (if there is a requirement for finer grained ECMP load balancing than per-circuit load balancing).

  * The next header will be set to 0x73 to indicate that the next header is L2TPv3.

  * In the "Static 1:1" case the IPv6 source address may correspond to a port or port/VLAN being transported as an L2 circuit, or may correspond to a virtual interface terminating inside the router (e.g. if L2 circuits are being used within a multipoint VPN, or if an anycast address is being terminated on a set of data center virtual machines.)

  * As with the source address the IPv6 destination address may correspond to a port or port/VLAN being transported as an L2 circuit, or to a virtual interface.

- Session ID. In the "Static 1:1 mapping" case described in Section 2, the IPv6 address identifies an L2TPv3 session directly, thus at endpoints supporting one-stage resolution (IPv6 Address only) the Session ID SHOULD be ignored upon receipt. It is RECOMMENDED that the remote endpoint is configured to set the Session ID to all ones (0xFFFFFFFF) for easy identification in case of troubleshooting. For compatibility with other tunnel termination platforms supporting only two-stage resolution (IPv6 Address + Session ID), this specification recommends supporting explicit configuration of Session ID to any value other than zero (including all ones). The Session ID of zero MUST NOT be used, as it is reserved for use by L2TP control messages as specified in RFC3931. Note that the Session ID is unidirectional - the sent and received Session IDs at an endpoint may be different.

- Cookie. 64-bit cookie, configured and described as in Section 3. All packets for a destined L2 circuit (or L2TPv3 Session) MUST match one of the cookie values configured for that circuit. Any packets that do not contain a valid cookie value MUST be discarded (see RFC3931 for more details).

- Payload (variable length). As noted above the preamble and any service-delimiting tags MUST be stripped before encapsulation and the FCS MUST be stripped unless FCS retention is configured at both ingress and egress routers. Since a new FCS is added at each hop when the encapsulating IP packet is transmitted the payload is protected against bit errors.

5. Fragmentation and Reassembly

Using tunnel encapsulation, Ethernet L2 datagrams in IPv6 in this case, will reduce the effective MTU of the encapsulated traffic.

The recommended solution to deal with this problem is for the network operator to increase the MTU size of all the links between the devices acting as IPv6 L2TPv3 tunnel endpoints to accommodate both the IPv6 L2TPv3 encapsulation header and the Ethernet L2 datagram without fragmenting the IPv6 packet.

It is RECOMMENDED that routers implementing this specification implement IPv6 PMTU discovery as defined in [RFC1981] to confirm that the path over which packets are sent has sufficient MTU to transport a maximum length Ethernet frame plus encapsulation overhead.

Routers implementing this specification MAY implement L2TPv3 fragmentation (as defined in section 5 of [RFC4623]). In the absence of the L2TPv3 control plane, it is RECOMMENDED that fragmentation (if implemented) is locally configured on a per-tunnel basis. Fragmention configuration MUST be consistent between the two ends of a tunnel.

It is NOT RECOMMENDED for routers implementing this specification to enable IPv6 fragmentation (as defined in section 4.5 of RFC2460) for keyed IP tunnels. IP fragmentation issues for L2TPv3 are discussed in section 4.1.4 of RFC3931.

6. OAM Considerations

OAM is an important consideration when providing circuit-oriented services such as those described in this document, and all the more so in the absence of a dedicated tunnel control plane, as OAM becomes the only way to detect failures in the tunnel overlay.

Note that in the context of keyed IP tunnels, failures in the IPv6 underlay network can be detected using the usual methods such as through the routing protocol, potentially including the use of Bidirectional Forwarding Detection (BFD) [RFC5880] to rapidly detect link failures. Multi-Hop BFD MAY also be enabled between tunnel endpoints as per [RFC5883].

Since keyed IP tunnels always carry an Ethernet payload, and since OAM at the tunnel layer is unable to detect failures in the Ethernet service processing at the ingress or egress router, or on the Ethernet attachment circuit between the router and the Ethernet client, it is RECOMMENDED that Ethernet OAM as defined in [IEEE802.1ag] and/or [Y.1731] is enabled for keyed IP tunnels. More specifically the following Connectivity Fault Management (CFM) and/or Ethernet continuity check (ETH-CC) configurations are to be used in conjunction with keyed IPv6 tunnels:

- Connectivity verification between the tunnel endpoints across the tunnel - use an Up MEP located at the tunnel endpoint for transmitting the CFM PDUs towards, and receiving them from the direction of the tunnel.

- Connectivity verification from the tunnel endpoint across the local attachment circuit - use a Down MEP located at the tunnel endpoint for transmitting the CFM PDUs towards, and receiving them from the direction of the local attachment circuit.

- o Intermediate connectivity verification - use a MIP located at the tunnel endpoint to generate CFM PDUs in response to received CFM PDUs.

In addition Pseudowire Virtual Circuit Connectivity Verification (VCCV) [RFC5085] MAY be used. Additionally BFD MAY be enabled over the VCCV channel [RFC5885].

7. IANA Considerations

None.

8. Security Considerations

Packet spoofing for any type of Virtual Private Network (VPN) tunneling protocol is of particular concern as insertion of carefully constructed rogue packets into the VPN transit network could result in a violation of VPN traffic separation, leaking data into a customer VPN. This is complicated by the fact that it may be particularly difficult for the operator of the VPN to even be aware that it has become a point of transit into or between customer VPNs.

Keyed IPv6 encapsulation provides traffic separation for its VPNs via use of separate 128-bit IPv6 addresses to identify the endpoints. The mandatory use of the 64 bit L2TPv3 cookie provides an additional check to ensure that an arriving packet is intended for the identified tunnel.

In the presence of a blind packet spoofing attack, the 64-bit L2TPv3 cookie provides security against inadvertent leaking of frames into a customer VPN, as documented in section 8.2 of RFC3931.

For protection against brute-force, blind, insertion attacks, the 64-bit cookie MUST be used with all tunnels.

Note that the cookie provides no protection against a sophisticated man-in-the-middle attacker who can sniff and correlate captured data between nodes for use in a coordinated attack.

The L2TPv3 64-bit cookie must not be regarded as a substitute for security such as that provided by IPsec when operating over an open or untrusted network where packets may be sniffed, decoded, and correlated for use in a coordinated attack.

9. Contributing Authors

Peter Weinberger
Cisco Systems

Email: peweinbe@cisco.com

Michael Lipman
Cisco Systems

Email: mlipman@cisco.com

Mark Townsley
Cisco Systems

Email: townsley@cisco.com

10. Acknowledgements

The authors would like to thank Carlos Pignataro, Stewart Bryant, Karsten Thomann, Qi Sun and Ian Farrer for their insightful suggestions and review.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4719] Aggarwal, R., Townsley, M., and M. Dos Santos, "Transport of Ethernet Frames over Layer 2 Tunneling Protocol Version 3 (L2TPv3)", RFC 4719, November 2006.

11.2. Informative References

[I.D.ietf-pppext-l2tphc] Valencia, A., "L2TP Header Compression", December 1997.

[IEEE802.1Q]
IEEE, "802.1Q-2014 - IEEE Standard for Local and metropolitan area networks - Bridges and Bridged Networks", 2014.

[IEEE802.1ad]
IEEE, "802.1ad-2005 - IEEE Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Area Networks - Amendment 4: Provider Bridges", 2005.

[IEEE802.1ag]
IEEE, "IEEE Standard for Local and metropolitan area networks - Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Managements", 2007.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

[RFC4623] Malis, A. and M. Townsley, "Pseudowire Emulation Edge-to-Edge (PWE3) Fragmentation and Reassembly", RFC 4623, August 2006.

[RFC4720] Malis, A., Allan, D., and N. Del Regno, "Pseudowire Emulation Edge-to-Edge (PWE3) Frame Check Sequence Retention", RFC 4720, November 2006.

[RFC5085] Nadeau, T. and C. Pignataro, "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, December 2007.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

[RFC5883] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for Multihop Paths", RFC 5883, June 2010.

[RFC5885] Nadeau, T. and C. Pignataro, "Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)", RFC 5885, June 2010.

[RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, November 2011.

[Y.1731] ITU, "ITU-T Recommendation G.8013/Y.1731 - OAM functions and mechanisms for Ethernet based networks", 2011.

Authors' Addresses

Maciek Konstantynowicz (editor)
Cisco Systems

Email: maciek@cisco.com

Giles Heron (editor)
Cisco Systems

Email: giheron@cisco.com

Rainer Schatzmayr
Deutsche Telekom AG

Email: rainer.schatzmayr@telekom.de

Wim Henderickx
Alcatel-Lucent, Inc.

Email: wim.henderickx@alcatel-lucent.com