Network Working Group
Internet Draft
Intended status: Standards Track
Expires: September 2007

M. Kelkar
T. Mistretta
P. Howard
Juniper Networks
V. Jain
Riverstone Networks

March 1, 2007

PPP over L2TP Tunnel Switching
draft-ietf-l2tpext-tunnel-switching-08.txt

Status of this Memo

Abstract

PPP [7] over L2TP Tunnel Switching, also called L2TP Multihop, is the process of forwarding PPP payload from an L2TP session to another L2TP session over a different tunnel. It facilitates moving the logical termination point of an L2TP session, based on layer 2 characteristics or administrative policies, to different L2tp

Endpoint. This document introduces the L2TP tunnel switching nomenclature and defines the behavior of standard AVPs in tunnel switching deployment. The scope of this document is limited to the discussion of switching PPP frames over L2TPv2 or L2TPv3 tunnels.

Table of Contents

Terminology

Tunnel Switching Aggregator (TSA): These are the devices that switch the layer 2 payload from a first L2TP session/tunnel on to second L2TP session/tunnel.

First Tunnel: The first L2tp Tunnel to be established at the TSA.

Second Tunnel: The second L2TP Tunnel to be established at the TSA.

First Session: The first L2tp Session to be established at the TSA.

Second Session: The second L2TP Session to be established at the TSA.

L2TP Control Connection Endpoint (LCCE): An L2TP node that exists at either end of an L2TP control connection. May also be referred to as an LAC or LNS, depending on whether tunneled frames are processed at the data link (LAC) or network layer (LNS).

L2TP, as defined in [1], is now referred to as "L2TPv2," while the extended version defined in [2] is referred to as "L2TPv3". The remainder of this document will refer simply to L2TP in general,

unless contrasting specific features of L2TPv2 or L2TPv3, which may differ in function.

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 [10].

1. Introduction

L2TP allows processing of PPP packets to be divorced from the termination of the layer 2 circuit. L2TP tunnel switching facilitates moving the termination point of a PPP session further on to another LCCE that is possibly unknown to the originating LCCE. It does so by re-tunneling the PPP session within another L2TP tunnel to a different LCCE. The knowledge of whether to switch a PPP session to another L2TP tunnel can be static or dynamic (for example, during PPP session establishment).

At the TSA, First and Second sessions are two discrete entities. The First session is established in the beginning and then TSA uses the negotiated parameters of the First session as a basis to negotiate the Second session. If the Second session fails to negotiate then it should be terminated. Same thing can be said for the tunnel.

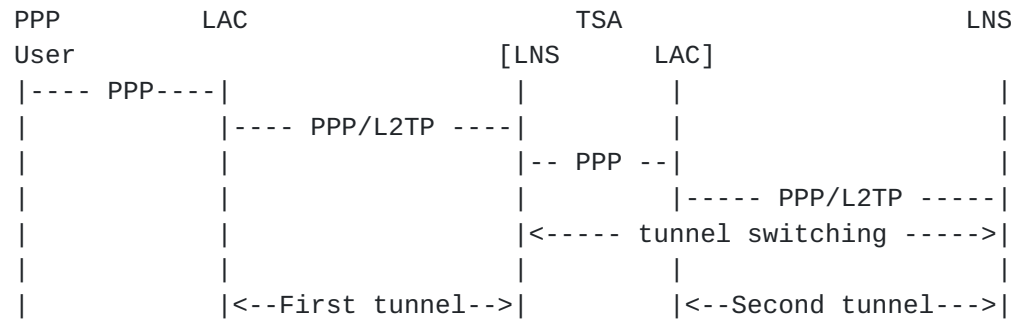


Figure 1: L2TP tunnel Switching for incoming calls

The figure above presents a typical tunnel switching scenario for incoming calls. The user opens a PPP session to a LAC, which tunnels the session to TSA as defined by L2TP protocol. The TSA, based on the local policies, determines if the First session should be further tunneled. If the TSA decides to tunnel the session further, then, for every such session it initiates another session onto another L2TP tunnel originating on the TSA terminating on a different LNS. Once the session is established, the data packets are switched from an incoming tunnel to a corresponding outgoing tunnel.

2. L2TPv2 to L2TPv3 switch

If First and Second tunnels use different versions of L2TP protocol at TSA i.e. if it involves a 'version switch', then it must adapt the data encapsulation change.

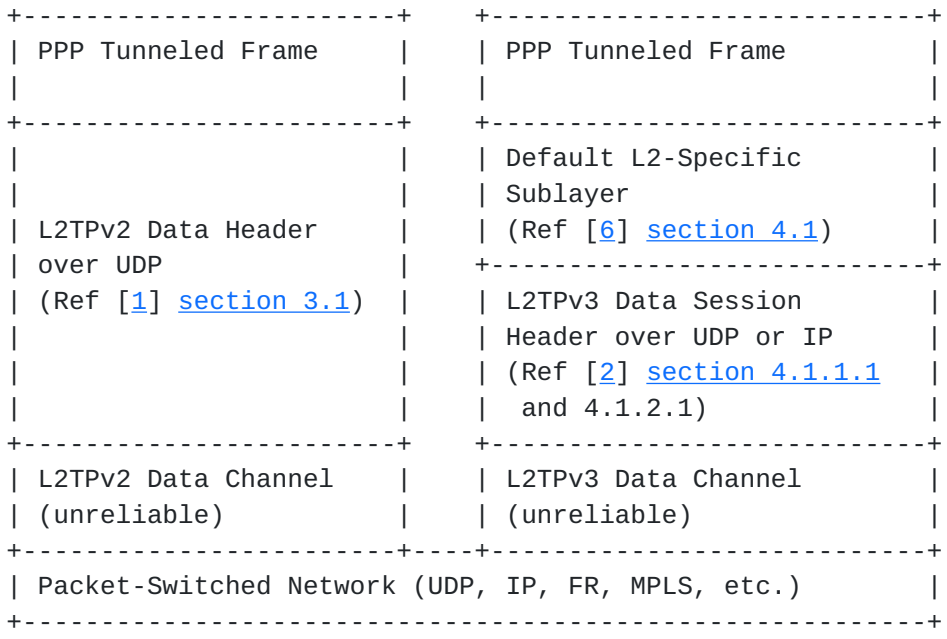


Figure 2: L2TPv2 to L2TPv3 data encapsulation switch

When PPP frames, which are encapsulated in the L2TPv2 header, are received at the TSA and are switched to Second tunnel using L2TPv3, then L2TPv2 headers are stripped and PPP frame is encapsulated with the L2TPv3 data header followed by the optional Default L2-Specific Sublayer and Offset Padding (Ref [6] section 4.2) fields, and forwarded over the session.

The version switch may involve a transport change i.e. L2TPv3-IP to L2TPv2-UDP. TSA MUST be able to adapt to such change.

3. AVP Behavior

An AVP negotiated by the First tunnel/session MUST be handled in four ways - it could be relayed, dropped, regenerated, or stacked. They are defined as follows.

- RELAYED AVP: (also known as pass-through AVP) AVP is forwarded transparently if it was negotiated by the First tunnel/session.

- DROPPED AVP: AVP is dropped if it was negotiated by the First tunnel/session.

- REGENERATED AVP: AVP negotiated by the First tunnel/session is ignored upon receipt. A new AVP is regenerated for the Second tunnel/session based on the local policy at the TSA. The local policy may or may not use the received AVP to regenerate the new value. The regenerated value MAY match with the received AVP value.

- STACKABLE AVP: Multiple instances of this AVP exist in the incoming message, each representing a hop in the tunnel switched path in order from first to last. When a TSA receives it, all the instances of AVP are copied as-is for the negotiation of the next hop. A locally generated AVP is appended to the outgoing message. If no value is appropriate then an AVP with a null value, as determined by the AVP definition, MUST be appended. However, If TSA couldn't copy all of incoming AVPs then it MUST not copy any one of them and drop all of the instances. If this is an AVP required to establish the tunnel or session and TSA cannot copy all of the stacked AVPS, then TSA MUST terminate the connection or session as appropriate.

3.1. IETF Vendor AVPs

This section defines the behavior of AVPs according to the guidelines in section 3. It describes the behavior of AVPs defined in [1], [2], [3], [4], [5], [6], and [8]. All the future AVP extensions MUST define AVP behavior for tunnel switching.

An optional AVP whose behavior is defined as RELAYED, MUST be RELAYED only if the AVP is negotiated by the First tunnel/session. Hence the behavior for such AVP is stated as 'RELAYED if negotiated by the first tunnel/session'.

An optional AVP whose behavior is defined as REGENERATED, could be DROPPED from the negotiation of the Second tunnel/session at the TSA's discretion. Hence the behavior for such AVP is stated as 'REGENERATED or DROPPED'

In its default behavior TSA needs to be as transparent as possible. However, TSA shouldn't prevent local policies to override the default behavior and allow regeneration of the AVPs mentioned as 'REGENERATED'.

Message Type (All Messages) - MUST be REGENERATED

Result Code (CDN, StopCCN) - MUST be either RELAYED or REGENERATED based on recommendations discussed in section 5. In case of version

switch, if L2TPv3 Result Codes and Error Codes are RELAYED then they MUST be translated into general error (Result Code 2, Error Code 0).

Protocol Version (SCCRQ, SCCRP) - MUST be REGENERATED. This would allow TSA to switch sessions when the First and Second tunnels use different versions of the L2TP protocol.

Framing Capabilities (SCCRQ, SCCRP) - MUST be REGENERATED.

Bearer Capabilities (SCCRQ, SCCRP) - MUST be either REGENERATED or DROPPED.

Tie Breaker (SCCRQ) - MUST be either REGENERATED or DROPPED.

Firmware Revision (SCCRP, SCCRQ) - MUST be either REGENERATED or DROPPED.

Host Name (SCCRP, SCCRQ) - MUST be either REGENERATED or DROPPED.

Vendor Name (SCCRP, SCCRQ) - MUST be either REGENERATED or DROPPED.

Assigned tunnel ID (SCCRP, SCCRQ, StopCCN) - MUST be REGENERATED.

Receive Window Size (SCCRQ, SCCRP) - MUST be either REGENERATED or DROPPED.

Challenge (SCCRP, SCCRQ) - MUST be either REGENERATED or DROPPED.

Q.931 Cause Code (CDN) - MUST be either RELAYED if negotiated by the First session or DROPPED.

Challenge Response (SCCCN, SCCRP) - MUST be either REGENERATED or DROPPED.

Assigned Session ID (CDN, ICRP, ICRQ, OCRP, OCRQ) - MUST be REGENERATED.

Call Serial Number (ICRQ, OCRQ) - MUST be RELAYED. It would best serve the intended purpose of this AVP and facilitate easier debugging.

Minimum BPS (OCRQ) - MUST be RELAYED if negotiated by the First session. In case of version switch, TSA should relay it as a Tx Minimum Speed AVP (Ref [6])

Maximum BPS (OCRQ) - MUST be RELAYED if negotiated by the First session. In case of version switch, TSA should relay it as a Tx Maximum Speed AVP (Ref [6])

Bearer Type (ICRQ, OCRQ) - MUST be RELAYED if negotiated by the First session.

Framing Type (ICCN, OCCN, OCRQ) - MUST be RELAYED.

Called Number (ICRQ, OCRQ) - MUST be RELAYED if negotiated by the First session.

Calling Number (ICRQ) - MUST be RELAYED if negotiated by the First session.

Sub-Address (ICRQ, OCRQ) - MUST be RELAYED if negotiated by the First session.

Tx Connect Speed (ICCN, OCCN) - MUST be RELAYED. In case of version switch, TSA should relay it as a Tx Connect Speed AVP (Attribute Type 74).

Physical Channel ID (ICRQ, OCRP) - MUST be either RELAYED if negotiated by the First session, REGENERATED, or DROPPED.

Proxy LCP AVPs (ICCN) - All the Proxy LCP AVPs (Initial Received LCP CONFREQ, Last Sent LCP CONFREQ and Last Received LCP CONFREQ) MUST be either all RELAYED, all REGENERATED or all DROPPED. If an AVP is REGENERATED then it would mean the LCP was renegotiated; whereas, RELAYED conveys the fact that it was passed along and was not renegotiated.

Proxy Authentication AVPs (ICCN) - All the Proxy Authentication AVPs (Proxy Authen Type, Proxy Authen Name AVP, Proxy Authen Challenge Proxy Authen ID and Proxy Authen Response AVP) MUST be either all RELAYED, all REGENERATED, or all DROPPED. If an AVP is REGENERATED then it would mean the Authentication was renegotiated; whereas, RELAYED conveys the fact that it was passed along and was not renegotiated.

Call Errors (WEN) - MUST be RELAYED.

ACCM (SLI) - MUST be RELAYED.

Random Vector (All Messages) - MUST be either REGENERATED or DROPPED.

Private Group ID (ICCN) - MUST be RELAYED if negotiated by the First session.

Rx Connect Speed (ICCN, OCCN) - MUST be RELAYED if negotiated by the First session. In case of version switch, TSA should relay it as a Rx Connect Speed AVP (Attribute Type 75).

Sequencing Required (ICCN, OCCN) - MUST be either REGENERATED or DROPPED. In case of version switch, TSA should regenerate it as a Data Sequencing AVP (Attribute Type 70).

Rx Minimum BPS (OCRQ) (Ref [8]) - MUST be RELAYED if negotiated by the First session. In case of version switch, TSA should relay it as a Rx Minimum Speed AVP (Ref [6]).

Rx Maximum BPS (OCRQ) (Ref [8]) - MUST be RELAYED if negotiated by the First session. In case of version switch, TSA should relay it as a Rx Maximum Speed AVP (Ref [6]).

PPP Disconnect Cause AVP (CDN) (Ref [3])- MUST be either RELAYED if negotiated by the First tunnel/session or DROPPED if it's not supported.

Control Connection DS AVP (SCCRQ, SCCRP) (Ref [4]) - MUST be either RELAYED if negotiated by the First tunnel or DROPPED if it's not supported. The value of this AVP could be chosen based on 'PHB Code' used (or to be used) on the tunnels, which the TSA is going to be switching tunnels to. TSA need not use same PHB-to-DSCP mappings on a First tunnel and Second tunnel.

Session DS AVP (ICRQ, ICRP, OCRQ, OCRP) (Ref [4]) - MUST be either RELAYED if negotiated by the First session or DROPPED if it's not supported. The value of this AVP could be chosen based on 'PHB Code' used (or to be used) on the sessions, which the TSA is going to be switching sessions to. TSA need not use same PHB-to-DSCP mappings on a First session and Second session.

LCP Want Options (ICCN, OCCN) (Ref [5]) - MUST be either RELAYED if negotiated by the First session or DROPPED if it's not supported.

LCP Allow Options (ICCN, OCCN) (Ref [5]) - MUST be either RELAYED if negotiated by the First session or DROPPED if it's not supported.

Extended Vendor ID AVP (Version 3) (All Messages) - MUST be either REGENERATED or DROPPED.

Message Digest AVP (Version 3) (All Messages) - MUST be either REGENERATED or DROPPED.

Router Id AVP (Version 3) (SCCRQ, SCCRP) - MUST be either REGENERATED or DROPPED.

Assigned Control Connection Id AVP (Version 3) (SCCRQ, SCCRP, StopCCN) - MUST be either REGENERATED or DROPPED.

Pseudowire Capabilities List AVP (Version 3) (SCCRQ, SCCRP) - MUST be either REGENERATED or DROPPED.

Local Session Id AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN, CDN, WEN, SLI) - MUST be either REGENERATED or DROPPED.

Remote Session Id AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN, CDN, WEN, SLI) - MUST be either REGENERATED or DROPPED.

Assigned Cookie AVP (Version 3) (ICRQ, ICRP, OCRQ, OCRP) - MUST be either REGENERATED or DROPPED.

Remote End Id AVP (Version 3) (ICRQ, OCRQ) - MUST be either REGENERATED or DROPPED.

Session Tie Breaker AVP (Version 3) (ICRQ, OCRQ) - MUST be either REGENERATED or DROPPED.

Pseudowire Type AVP (Version 3) (ICRQ, OCRQ) - MUST be either REGENERATED or DROPPED.

L2-specific Sublayer AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN) - MUST be either REGENERATED or DROPPED.

Data Sequencing AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN) - Data sequencing AVP (Attribute Type 70) is a L2TPV3 AVP equivalent to the Sequencing required AVP (Attribute Type 39) in L2TPv2. In L2TPv3, any endpoint (LAC or LNS i.e. LCCE) can send the Data Sequencing AVP with the value 0 (no sequencing), 1 (sequencing only for non-ip packets) or 2 (sequencing for all the packets). In L2TPv2, only LAC can send the Sequencing Required AVP that requests the sequencing for all the packets. If data sequencing is enabled on the First session, then TSA should enable it on the Second session by sending the appropriate AVP (i.e. REGENERATED). If data sequencing is enabled on the First session, then TSA MAY choose (as decided by the local policy) not to enable the sequencing on Second session i.e. DROPPED. TSA SHOULD not enforce the sequencing but should send the data sequencing AVP on the Second session, if it's enabled on the

First session. There is no requirement to have all hops use the consistent sequencing configuration. As always TSA's local policy would take precedence over the default behavior of "REGENERATED or DROPPED"

Circuit Status AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN, SLI) - MUST be either REGENERATED or DROPPED.

Preferred Language AVP (Version 3) (SCCRQ, SCCRP) - MUST be either REGENERATED or DROPPED.

Tx Connect Speed AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN) - MUST be either RELAYED, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Tx Connect Speed AVP (Attribute Type 24). If value is greater than 4 octets, it SHOULD be dropped.

Rx Connect Speed AVP (Version 3) (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN) - MUST be either RELAYED if negotiated by the First session, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Rx Connect Speed AVP (Attribute Type 38). If value is greater than 4 octets, it SHOULD be dropped.

Offset Size (Version 3) (ICRQ, ICRP, ORCQ, OCRP) (Ref [6]) - MUST be either REGENERATED or DROPPED.

Tx Minimum Speed AVP (Version 3) (OCRQ) (Ref [6]) - MUST be either RELAYED, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Minimum BPS AVP (Attribute Type 16). If value is greater than 4 octets, it SHOULD be dropped.

Tx Maximum Speed AVP (Version 3) (OCRQ) (Ref [6]) - MUST be either RELAYED, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Maximum BPS AVP (Attribute Type 17). If value is greater than 4 octets, it SHOULD be dropped.

Rx Minimum Speed AVP (Version 3) (OCRQ) (Ref [6]) - MUST be either RELAYED, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Rx Minimum BPS AVP (Attribute Type 40). If value is greater than 4 octets, it SHOULD be dropped.

Rx Maximum Speed AVP (Version 3) (OCRQ) (Ref [6]) - MUST be either RELAYED, REGENERATED or DROPPED. In case of version switch, TSA should relay it as a Rx Maximum BPS AVP (Attribute Type 41). If value is greater than 4 octets, it SHOULD be dropped.

Failover Capability AVP (SCCRQ, SCCRP) (Ref [9]) - MUST be REGENERATED based on the TSA's capabilities

Tunnel Recovery AVP (SCCRQ) (Ref [9]) - MUST be REGENERATED based on failover negotiations with the peer on an individual tunnel.

Suggested Control Sequence AVP (SCCRP) (Ref [9]) - MUST be REGENERATED based on failover negotiations with the peer on an individual tunnel.

Failover Session State AVP (FSQ, FSR) (Ref [9]) - MUST be REGENERATED based on failover negotiations with the peer on an individual tunnel.

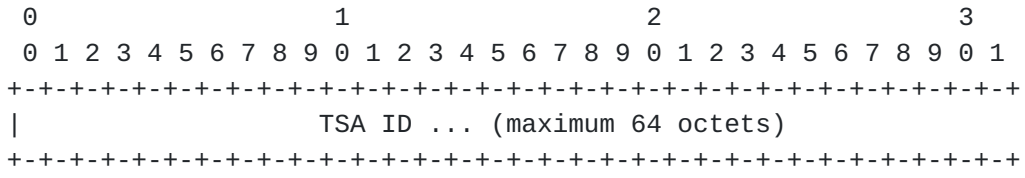
TSA ID AVPs (defined in this document) - MUST be STACKED. The local TSA ID AVP is stacked to the incoming set of TSA ID AVPs.

4. Loop Detection

The Tunnel Switching Aggregator (TSA) ID AVP, Attribute Type 93, could be used to detect if a session is looping in an L2TP tunnel switched network. This AVP MUST be STACKED.

If this AVP was received in an incoming control packet (ICRQ, OCRQ) then the TSA MUST check to see if it's own TSA ID (a configured value) is present in the stack of incoming TSA ID AVPs. Upon finding a match, the TSA MUST respond with a CDN carrying a Result Code indicating 'Loop Detected' 26, and optionally a description indicating the loop condition. A match comparison MUST only be performed if TSA has configured non-null TSA ID.

The Attribute Value field for this AVP has the following format:



TSA ID is a configured value (human readable string) with a maximum length of 64 octets. It is administratively controlled to ensure its uniqueness among all the inter-connected LACs, LNSs and TSAs. If no value is configured then the AVP value MUST be of length 0.

This AVP MUST be either hidden (the H-bit can be either 0 or 1). The M-bit for this AVP MUST be set to 0.

5. CDN Messages and L2TP tunnel Switching

To identify error conditions explicitly in the multi-TSA environment, new error codes are defined. Existing error codes are not used because they might trigger an unwarranted behavior depending upon why it was generated. Error codes are defined as follows:

Next hop unreachable (Result Code 2, Error Code 10) - TSA MUST disconnect the First tunnel/session with this Error Code, if next hop is unreachable and no other alternative paths are available as determined by the local policy.

Next hop busy (Result Code 2, Error Code 11) - TSA MUST disconnect the First tunnel/session with this Error Code, if next hop disconnects the Second tunnel/session with an error code 'TSA Busy' or other indications from next hop indicate that it is too busy to take more tunnels/sessions and no other alternative paths are available as determined by the local policy

TSA busy (Result Code 2, Error Code 12) - TSA MUST disconnect the first tunnel/session with this Error Code, if it is congested or temporarily running out of resources.

In the case of multiple levels of TSAs, error code SHOULD be propagated back until it reaches either the original LCCE or an intermediate TSA, which has an alternate path. On the receipt of error code, local policy on the LCCE or the intermediate TSA should handle the fallback and use it for the congestion recovery design.

6. IANA Considerations

6.1. Control Message Attribute Value Pairs (AVPs)

This number space is managed by IANA as per section 2.1 of [11].

A summary of the new AVPs follows:

Control Message Attribute Value Pairs

Attribute Type	Description
93	Tunnel Switching Aggregator ID AVP

6.2. Result Code AVP Values

New L2TP Result Codes appear in section 4 and 5, which need assignment by IANA as described in section 2.3 of [11].

```
Result Code AVP (Attribute Type 1) Values
-----------------------------------------
Defined Result Code values for the CDN message are:
26 - Loop Detected

General Error Codes
10 - Next hop unreachable
11 - Next hop busy
12 - TSA busy
```

7. Security Considerations

TSA ID AVP could reveal the set of nodes that a given L2TP session is traversing in the network.

If the AVPs described in this document are of concern then AVP hiding, described in [1] MAY be used to help mitigate the security threat; though it is not widely regarded as cryptographically secure, [12] describes a more robust method for securing L2TP in general, and should be used to encrypt all L2TP messages.

8. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this

specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

9. Copyright Statement

10. Acknowledgments

Authors gratefully acknowledge the valuable contributions of: Mark W. Townsley, Reinaldo Penno, Ly Loi and Marc Eaton-Brown. We would like to thank Carlos Pignataro for a thorough review.

11. References

11.1. Normative References

[1] RFC 2661, W. Townsley, A. Valencia, A. Rubens, G. Pall, G. Zorn, B. Palter, "Layer 2 Tunnel Protocol (L2TP)", August 1999.

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

[3] RFC 3145, Verma, et. al. "L2TP Disconnect Cause Information", July 2001.

[4] RFC 3308, P. Calhoun, W. Luo, D. McPherson, K. Peirce, "Layer Two Tunneling Protocol (L2TP) Differentiated Services Extension", November 2002.

[5] RFC 3437, W. Palter, W. Townsley, "Layer-Two Tunneling Protocol Extensions for PPP Link Control Protocol Negotiation", December 2002.

[6] C. Pignataro, Ed, "PPP Tunneling Using Layer Two Tunneling Protocol Version 3" work in progress, draft-ietf-l2tpext-l2tp-ppp-05.txt, November 2006.

[7] RFC 1661, W. Simpson, "The Point-to-Point Protocol (PPP)", STD 51, July 1994.

[8] RFC 3301, Y. T'Joens, B. Sales, P. Crivellari, "Layer Two Tunneling Protocol (L2TP): ATM access network extensions", June 2002

[9] V. Jain, Ed, "Fail Over extensions for L2TP failover" work in progress, draft-ietf-l2tpext-failover-12.txt, February 2007.

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

11.2. Informative References

[11] BCP0068, Townsley, W., "Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers Authority (IANA) Considerations Update" RFC3438, BCP0068, December 2002

[12] RFC 3193, B. Patel, B. Aboba, W. Dixon, G. Zorn, S. Booth, "Securing L2TP using IPsec", November 2001

Author's Addresses

Mahesh Kelkar
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
Email: mkelkar@juniper.net

Tom Mistretta
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
Email: tmistretta@juniper.net

Paul Howard
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
Email: phoward@juniper.net

Vipin Jain
Riverstone Networks
5200 Great America Parkway
Santa Clara, CA 95054
Email: vipinietf@yahoo.com