Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 2, 2009

C. Filsfils
P. Mohapatra
C. Pignataro
Cisco Systems
July 1, 2008

Load Balancing for Mesh Softwires
draft-pmohapat-softwire-lb-00

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on January 2, 2009.

Abstract

Payloads carried over a Softwire mesh service as defined by BGP Encapsulation Subsequent Address Family Identifier (SAFI) information exchange often carry a number of identifiable, distinct flows. It can in some circumstances be desirable to distribute these flows over the equal cost multiple paths (ECMPs) that exist in the packet switched network. Currently, the payload of a packet entering the Softwire can only be interpreted by the ingress and egress routers. Thus the load balancing decision of a core router is only based on the encapsulating header, presenting much less entropy than available in the payload or the encapsulated header since the Softwire encapsulation acts in a tunneling fashion. This document describes a

method for achieving comparable load balancing efficiency in a network carrying Softwire mesh service over Layer Two Tunneling Protocol - Version 3 (L2TPv3) over IP or Generic Routing Encapsulation (GRE) encapsulation to what would be achieved without such encapsulation.

Table of Contents

1. Introduction 3
1.1. Requirements Language 3
2. Load Balancing Block sub-TLV 3
2.1. Applicability to Tunnel Types 4
2.2. Encapsulation Considerations 5
3. IANA Considerations 5
4. Security Considerations 5
5. Acknowledgements 5
6. Normative References 5
Authors' Addresses 6
Intellectual Property and Copyright Statements 7

1. Introduction

Consider the case of a router R1 which encapsulates a packet P into a Softwire bound to router R3. R2 is a router on the shortest path from R1 to R3. R2's shortest path to R3 involves equal cost multiple paths (ECMPs). The goal is for R2 to be able to choose which path to use on the basis of the full entropy of packet P.

This is achieved by carrying in the encapsulation header a signature of the inner header, hence enhancing the entropy of the flows as seen by the core routers. The signature is carried as part of one of the fields of the encapsulation header. To aid with better description in the document, we define the generic term "load balancing field" to mean such a value that is specific to an encapsulation type. For example, for L2TPv3-over-IP [RFC3931] encapsulation, the load balancing field is the Session Identifier (Session ID). For GRE [RFC2784] encapsulation, the key field [RFC2890], if present, represents the load balancing field. This mechanism assumes that core routers base their load-balancing decisions on a flow definition that includes the load balancing field. This is an obvious and generic functionality as, for example, for L2TPv3-over-IP tunnels, the Session ID is at the same well-known constant offset as the TCP/UDP ports in the encapsulating header.

The "Encapsulation SAFI" [I-D.ietf-softwire-encaps-safi] is extended such that a contiguous block of the load balancing field is bound to the Softwire advertised by a BGP next-hop. On a per-inner flow basis, the ingress PE selects one value of the load balancing field from the block to preserve per-flow ordering, and at the same time to enhance the entropy across flows.

1.1. 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 RFC 2119 [RFC2119].

2. Load Balancing Block sub-TLV

This document defines a new sub-TLV for use with the Tunnel Encapsulation Attribute defined in [I-D.ietf-softwire-encaps-safi]. The new sub-TLV is referred to as the "Load Balancing Block sub-TLV" and MAY be included in any Encapsulation SAFI UPDATE message where load balancing is desired.

The sub-TLV type of the Load Balancing Block sub-TLV is 5. The sub-TLV length is 2 octets. The value represents the length of the block

in bits and it MUST NOT exceed the size of the load balancing field. This format is very similar to the variable-length subnet masking (VLSM) used in IP addresses to allow arbitrary length prefixes. The block is determined by extracting the initial sequence of 'block size' bits from the load balancing field.

As an example, Assume that there is a Softwire set up between R1 and R3 with L2TPv3-over-IP tunnel type. Assume that R3 encodes the Session ID with value 0x1234ABCD in the encapsulation sub-TLV. It also includes the load balancing block sub-TLV and encodes the value 24. This should be interpreted as follows:

- If an ingress router does not understand Load Balancing block sub-TLV, it continues to use the Session ID 0x1234ABCD and encapsulates all packets with that Session ID,

- If an ingress router understands Load Balancing Block sub-TLV, it picks the first 24 bits out of the Session ID (0x1234AB) to be used as the block and fills in the lower-order 8 bits with a per-flow identifier (e.g. it can be determined based on the inner packet's source, destination addresses and TCP/UDP ports). This selection preserves per-flow ordering of packets.

This requirement and solution applies equally to GRE where the key plays the same role as the Session ID in L2TPv3.

Needless to say, if an egress router does not support load balancing block sub-TLV, the Softwire continues to operate with a single load balancing field that all ingress routers encapsulate with.

2.1. Applicability to Tunnel Types

The load balancing block sub-TLV is applicable to Tunnel types that define a load balancing field. This document defines load balancing fields for tunnel types 1 (L2TPv3 over IP) and 2 (GRE) as follows:

- L2TPv3 over IP - Session ID. Special care needs to be taken to always create a non-zero Session ID. When an egress router includes a load balancing sub-TLV, it MUST encode the Session ID field of the Encapsulation sub-TLV in a way that ensures that the most significant bits of the Session ID after extracting the block are non-zero.

- GRE - GRE key

Future tunnel types that desire to use the load balancing sub-TLV MUST define a load balancing field that is part of the encapsulating header.

2.2. Encapsulation Considerations

Fields included in the encapsulation header besides the load balancing field are not affected by the load balancing block sub-TLV. All other encapsulation fields are shared between variations of the load balancing field. For example, for L2TPv3-over-IP tunnel type, if the optional cookie is included in the Encapsulation sub-TLV by the egress router during Softwire signaling, it applies to all the "Session ID" values derived at the ingress router after applying the load balancing block as described in this document.

3. IANA Considerations

IANA is requested to assign the type of 5 for the Load Balancing Block sub-TLV, in the tunnel sub-TLV types of the Tunnel Encapsulation attribute registry (number space created as part of the publication of [I-D.ietf-softwire-encaps-safi]):

Sub-TLV name	Type
Load Balancing Block	5

4. Security Considerations

There are no additional security risks introduced by this design.

5. Acknowledgements

The authors would like to thank Stewart Bryant and Mark Townsley for their review and comments.

6. Normative References

[I-D.ietf-softwire-encaps-safi] Mohapatra, P. and E. Rosen, "BGP Encapsulation SAFI and BGP Tunnel Encapsulation Attribute", draft-ietf-softwire-encaps-safi-03 (work in progress), June 2008.

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

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784,

March 2000.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, September 2000.

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

Authors' Addresses

Clarence Filsfils
Cisco Systems
Brussels,
Belgium

Email: cfilsfil@cisco.com

Pradosh Mohapatra
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: pmohapat@cisco.com

Carlos Pignataro
Cisco Systems
7200 Kit Creek Road, PO Box 14987
Research Triangle Park, NC 27709
USA

Email: cpignata@cisco.com

Full Copyright Statement

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

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.